Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3611~3622
e-ISSN: 2548-6861

Classification of Depression Indication Based on Facial Expression Using MobileNetV2 and Support Vector

Muhammad Eswin Bakkar [1], Christy Atika Sari [2]*, Eko Hari Rachmawanto [3]
Informatics Engineering, Universitas Dian Nuswantoro, Semarang, Indonesia
eswinbkkr@gmail.com [1], christy.atika.sari@dsn.dinus.ac.id [2], eko.hari@dsn.dinus.ac.id [3]

Article Info

Article history:

Received 2026-07-01
Revised 2026-07-28
Accepted 2026-08-10

Keyword:

Depression,
Facial Expression Recognition,
MobileNetV2,
Support Vector Machine,
Deep Learning.

ABSTRACT

Depression is a mental health disorder that can affect an individual's emotional condition, behavior, and quality of life. Facial expressions can represent a person's emotional state and therefore have the potential to be utilized as a visual source of information in the development of artificial intelligence-based classification systems. This study aims to develop a depression indication classification model based on facial expressions using MobileNetV2 as a feature extractor and Support Vector Machine (SVM) as a classifier. The FER2013 dataset was used and grouped into two classes, namely depression indication and non-depression indication based on predefined facial expression categories used in this study. After the labeling process, a total of 19,275 facial images were obtained, with 3,855 images used as testing data. The proposed method consists of image preprocessing, feature extraction using MobileNetV2, classification using SVM, threshold optimization, and model evaluation. Experimental results show that the proposed model achieved an accuracy of 79.69% with an AUC value of 88.62%. Threshold optimization produced an optimal threshold value of 0.44 and improved the accuracy to 80.34%. The precision, recall, and F1-score values indicate relatively balanced performance across both classes. The results demonstrate that the combination of MobileNetV2 and SVM can provide good classification performance on the FER2013 dataset grouped into depression indication and non-depression indication classes.

I. INTRODUCTION

Depression is one of the most common mental health disorders and has a significant impact on an individual's emotional well-being, behavior, productivity, and overall quality of life [1]. Early identification of depressive symptoms is important because it may support timely intervention and appropriate clinical decision-making [2]. Recent advances in artificial intelligence (AI), particularly machine learning and deep learning, have encouraged the development of computational approaches for analyzing various types of data to assist mental health assessment and decision support [1], [2]. However, depression is a complex psychiatric disorder that cannot be diagnosed solely based on facial expressions or other single behavioral modalities. Therefore, AI-based facial analysis should be regarded only as a complementary tool for identifying emotional patterns that may be associated with depressive symptoms rather than as a substitute for clinical diagnosis [3].

Emotion recognition technology has experienced rapid development over the past decade and has been widely adopted in various domains, including healthcare, education, human–computer interaction, and psychological assessment. Among various affective signals, facial expressions remain one of the most informative and non-invasive sources for recognizing human emotional states [4].

Furthermore, machine learning algorithms have become one of the dominant computational approaches for automatic human emotion recognition because they are capable of learning complex emotional representations from large-scale facial image datasets while providing robust performance across various recognition tasks [5].

Various sources of information have been explored for automatic analysis of mental health conditions, including text, speech, physiological signals, and visual information [3]. Among these modalities, facial expressions have attracted considerable attention because they provide nonverbal cues that reflect an individual's emotional state without requiring invasive measurement procedures [3]. Facial expressions are therefore considered a valuable source of information for affective computing and human behavior analysis, as emotional states can be represented through changes in facial appearance [6]. Nevertheless, facial expressions alone cannot provide sufficient evidence for diagnosing depression because they represent only observable emotional responses that may be influenced by various psychological and environmental factors. Consequently, in this study, facial expressions are utilized as visual information for an exploratory classification task rather than as direct indicators of clinical depression [3].

Facial Expression Recognition (FER) focuses on automatically identifying human emotional expressions through the analysis of facial visual characteristics [6]. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved FER performance by enabling automatic extraction of discriminative visual features without requiring manual feature engineering [6]. Furthermore, transfer learning has become an effective approach for improving model performance by utilizing knowledge learned from large-scale datasets, allowing robust feature representations to be obtained even when the target dataset is relatively limited [7]. These advantages make CNN-based transfer learning approaches widely adopted in facial expression recognition and other image classification tasks. Recent survey studies have reported substantial progress in facial expression recognition through the development of larger benchmark datasets, improved deep learning architectures, and more comprehensive evaluation methodologies, resulting in more robust and reliable facial emotion recognition systems under diverse imaging conditions [8].

Among various CNN architectures, MobileNetV2 has emerged as an effective feature extractor due to its lightweight architecture, computational efficiency, and ability to learn discriminative visual representations [9]. By employing depthwise separable convolution and inverted residual blocks, MobileNetV2 significantly reduces computational complexity while maintaining competitive classification performance [9]. In addition, the availability of pretrained weights through transfer learning enables the model to extract robust facial features without requiring extensive training from scratch, making it particularly suitable for image classification tasks with limited computational resources and moderate-sized datasets [7]. Therefore, MobileNetV2 was adopted in this study as the feature extraction backbone to generate representative facial feature vectors for the subsequent classification stage.

Although end-to-end deep learning architectures have demonstrated excellent performance in various image classification tasks, hybrid approaches that combine deep feature extraction with conventional machine learning classifiers have also shown competitive results, particularly when working with limited or moderate-sized datasets [10]. Among conventional classifiers, Support Vector Machine (SVM) is widely recognized for its strong generalization capability and effectiveness in handling high-dimensional feature spaces by constructing an optimal decision boundary [11]. Several previous studies have reported that CNN-extracted features classified using SVM can achieve better or comparable performance to conventional Softmax classifiers while requiring fewer trainable parameters [10]. Furthermore, recent studies have demonstrated the effectiveness of the MobileNetV2–SVM hybrid framework in medical image classification, indicating that combining efficient deep feature extraction with a robust classifier can improve classification performance while maintaining computational efficiency [12], [13]. Therefore, this study adopts the MobileNetV2–SVM hybrid approach to exploit the complementary strengths of both methods for facial expression-based classification.

Recent studies have demonstrated the potential of facial analysis for supporting depression-related assessment using various deep learning approaches and clinically validated datasets. Existing methods have employed two-dimensional and three-dimensional facial expressions, temporal facial landmarks, facial expression videos, multimodal feature fusion, and attention mechanisms to distinguish individuals with depressive symptoms from healthy controls [14], [15], [16], [17]. In parallel, substantial progress has also been achieved in Facial Expression Recognition through the development of advanced transfer learning strategies and landmark-based facial representations, enabling more robust emotion recognition under various imaging conditions [18], [19]. These findings indicate that facial visual information contains meaningful emotional characteristics that may support depression-related analysis; however, differences in datasets, input modalities, and model architectures make direct comparison among existing approaches challenging.

Despite these advancements, several research gaps remain. Most previous studies on depression-related facial analysis have relied on clinically annotated datasets, facial videos, multimodal information, or end-to-end deep learning architectures, making them difficult to compare with image-based facial expression recognition datasets [14], [15], [16]. Meanwhile, although hybrid CNN–SVM models have demonstrated promising performance in various image classification tasks, their application to an exploratory facial expression grouping scenario using the FER2013 dataset has received relatively limited attention [10], [12]. Furthermore, previous studies have rarely investigated the influence of decision threshold optimization on the classification performance of MobileNetV2–SVM hybrid models in this context [9]. These limitations provide an opportunity to investigate an efficient hybrid classification framework while carefully interpreting the results within an exploratory facial expression-based scenario rather than as clinically validated depression detection.

Based on the identified research gaps, this study proposes a hybrid classification framework that combines MobileNetV2 as a deep feature extractor and Support Vector Machine (SVM) as the classifier to classify facial expressions grouped into depression indication and non-depression indication categories under an experimental setting. In addition, decision threshold optimization is performed to determine an appropriate classification threshold and to analyze its influence on model performance [10], [11]. The primary contribution of this study lies in evaluating the effectiveness of the MobileNetV2–SVM hybrid architecture for facial expression-based classification using the FER2013 dataset and investigating the impact of threshold optimization on binary classification performance. It is important to emphasize that the proposed framework is intended for exploratory research on facial expression analysis and should not be interpreted as a clinically validated system for depression diagnosis. Therefore, further validation using clinically annotated depression datasets and collaboration with mental health professionals are necessary before practical implementation [9].

II. METHOD

A. Research Stages

The first stage involved selecting the facial expression categories from the FER2013 dataset. These categories were subsequently reorganized into two experimental classes, namely depression indication and non-depression indication. Since FER2013 was originally developed for facial expression recognition rather than clinical depression assessment, the grouping adopted in this study was designed as an experimental label mapping to investigate whether facial expression patterns could be classified into predefined categories associated with negative emotional cues reported in depression-related literature rather than representing clinically validated depression status [20]. This approach was motivated by previous studies showing that facial expressions may provide observable behavioral cues associated with depressive symptoms, although facial expressions alone are insufficient for establishing a clinical diagnosis of depression [21], [22]. Therefore, the resulting labels should be interpreted exclusively as experimental categories for machine learning evaluation and not as clinical depression labels [20]. After the label formation process, image preprocessing was performed to adapt the facial images to the input requirements of the MobileNetV2 architecture. A detailed description of the label formation procedure is presented in Section II-B.

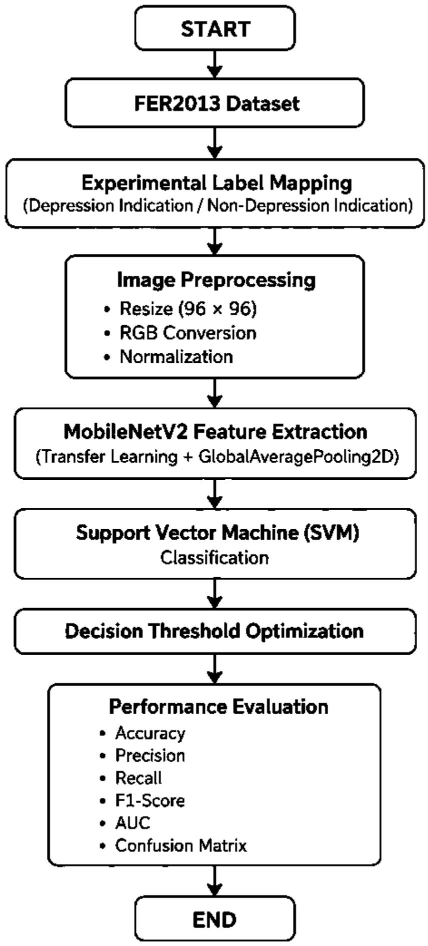

Figure 1. Research Stages

Based on Figure 1, the proposed research consists of three main stages, namely data preparation, feature extraction and classification, and model evaluation, to develop a facial expression-based classification model capable of distinguishing between the depression indication and non-depression indication classes.

1. The first stage involved selecting facial expression categories from the FER2013 dataset and reorganizing them into two experimental classes through an experimental label mapping procedure. Since FER2013 was originally developed for facial expression recognition rather than clinically validated depression assessment, the generated labels were intended solely for experimental machine learning evaluation and should not be interpreted as clinical depression labels. Subsequently, image preprocessing was performed to prepare the facial images according to the input requirements of the MobileNetV2 architecture [20], [21].
2. In the feature extraction stage, facial images were processed using MobileNetV2 pretrained on the ImageNet dataset through transfer learning to produce relevant feature representations. The obtained feature representations were then condensed using

GlobalAveragePooling2D to produce a more concise feature vector ready for use in the classification process.

3. The extracted feature vectors were subsequently classified using Support Vector Machine (SVM). The resulting prediction probabilities were further utilized during the threshold optimization stage by evaluating different threshold values to determine the optimal decision threshold. The model performance was then evaluated using metrics of accuracy, precision, recall, F1-score, and Area Under Curve (AUC) to measure the model's ability to distinguish between depression indication and non-depression indication classes.

B. Dataset and Label Formation

The proposed study utilized the FER2013 dataset, a publicly available benchmark dataset originally developed for facial expression recognition tasks. The dataset consists of grayscale facial images with a resolution of 48 × 48 pixels, representing seven basic facial expression categories, namely angry, disgust, fear, happy, sad, surprise, and neutral. Owing to its large number of facial images collected under diverse facial appearances and expression variations, FER2013 has been widely employed to evaluate facial expression recognition models and deep learning-based feature extraction methods [6]. It should be noted, however, that FER2013 was not designed as a clinically annotated depression dataset and therefore does not contain validated depression diagnoses [20].

Since the objective of this study was to investigate the feasibility of facial expression-based classification associated with depression-related emotional characteristics, the original FER2013 emotion labels were reorganized into two experimental categories through an experimental label mapping strategy. This mapping was introduced solely to construct a binary classification problem for machine learning evaluation and should not be interpreted as a clinically validated depression labeling scheme [20]. Consequently, the generated classes represent experimental categories referred to as depression indication and non-depression indication, rather than confirmed clinical depression status.

In this study, the original FER2013 emotion categories were grouped into two experimental classes. The sad and fear expression categories were assigned to the depression indication class, whereas the happy and surprise categories were assigned to the non-depression indication class. The remaining expression categories (angry, disgust, and neutral) were excluded from the analysis because this study focused on facial expressions representing contrasting emotional characteristics for experimental binary classification. This grouping was established solely as an experimental label mapping strategy and should not be interpreted as clinically validated depression labelling [20], [21].

TABLE I
EXPERIMENTAL LABELLING

Original FER2013 Expression	Experimental Class
Sad	Depression Indication
Fear	Depression Indication
Happy	Non-Depression Indication
Surprise	Non-Depression Indication

Table I summarizes the experimental label mapping adopted in this study to construct the binary classification problem.

Figure 2. Representative Facial Expression Samples From The Fer2013 Dataset Used In The Experimental Label Mapping

The experimental label mapping adopted in this study was motivated by psychological findings indicating that facial expressions can reflect observable emotional and behavioral cues associated with depressive symptoms. Individuals experiencing depression commonly exhibit persistent negative affect, reduced positive affect, and alterations in facial expressiveness, although facial expressions alone are insufficient to establish a clinical diagnosis of depression [21], [22]. However, these facial expressions should be interpreted as behavioral cues rather than definitive clinical indicators of depression. Therefore, the proposed binary grouping should be interpreted as an exploratory representation of facial expression patterns potentially associated with depression-related emotional characteristics rather than as evidence of clinically diagnosed depression [20].

Following the experimental label mapping process, the resulting binary dataset was divided into training and testing subsets using an 80:20 ratio. The dataset partitioning was performed using stratified sampling with a fixed random state (random_state = 42) to preserve the original class distribution across both subsets and ensure reproducibility of the experimental results. Subsequently, the training subset was utilized for model learning, whereas the testing subset was reserved exclusively for performance evaluation to provide an unbiased assessment of the proposed classification model.

Because FER2013 was originally developed for facial expression recognition rather than depression assessment, the proposed experimental label mapping represents an exploratory approach for investigating facial expression-based classification. Consequently, the developed model should not be considered a clinical depression screening or diagnostic tool. Further validation using clinically annotated depression datasets and collaboration with mental health professionals are required before such a model can be considered for practical mental health applications [20].

C. Data Preprocessing

The preprocessing stage was performed to transform the facial images into a format compatible with the MobileNetV2 architecture. The original FER2013 images with a resolution of 48 × 48 pixels were resized to 96 × 96 pixels to satisfy the input requirements of the pretrained MobileNetV2 model while preserving the essential facial characteristics required for feature extraction. Subsequently, the grayscale images were converted into three-channel RGB images, resulting in image dimensions of 96 × 96 × 3, because MobileNetV2 was originally pretrained on the ImageNet dataset using RGB images.

Afterward, image normalization was applied using the preprocess_input() function provided by MobileNetV2. This preprocessing function transforms pixel values according to the normalization scheme employed during ImageNet pretraining, thereby ensuring compatibility between the input images and the pretrained network and facilitating effective feature extraction.

D. Feature Extraction Using MobileNetV2

Feature extraction was performed using the MobileNetV2 architecture through a transfer learning approach. MobileNetV2 was selected because it provides computationally efficient deep feature extraction while maintaining robust feature representation performance for image classification tasks. Recent studies have shown that improving deep convolutional architectures can further enhance the discriminative capability of facial feature representations, highlighting the importance of effective feature extraction for facial emotion recognition tasks [23].

Instead of training a convolutional neural network from scratch, this study adopted transfer learning by utilizing ImageNet-pretrained weights with the include_top = False configuration. This configuration removes the original classification layer, allowing MobileNetV2 to function exclusively as a deep feature extractor. Furthermore, all pretrained layers were frozen (trainable = False) during training to preserve the learned visual representations while reducing computational cost and minimizing the risk of overfitting. Transfer learning has been widely adopted to improve feature extraction performance, particularly when the available training data are limited [24].

Rather than employing an end-to-end convolutional neural network classifier, this study adopted a hybrid MobileNetV2–SVM architecture, in which MobileNetV2 was responsible for extracting high-level facial features, whereas Support Vector Machine (SVM) performed the final binary classification. This hybrid architecture separates feature extraction from classification, allowing the discriminative deep features learned by MobileNetV2 to be combined with the decision boundary generated by SVM without requiring fine-tuning of the entire convolutional network. Such a strategy is appropriate for datasets of moderate size and provides flexibility in integrating deep feature extraction with conventional machine learning classifiers. Previous studies have demonstrated the effectiveness of MobileNetV2–SVM hybrid architectures in image classification tasks and highlighted the advantages of combining deep feature extraction with conventional machine learning classifiers [12], [13].

The feature maps generated by MobileNetV2 were subsequently processed using a GlobalAveragePooling2D layer to transform the multidimensional feature maps into compact one-dimensional feature vectors before classification. This operation reduces feature dimensionality while preserving the most representative information extracted by the convolutional layers. The resulting feature vectors were subsequently used as the input to the SVM classifier for binary classification.

E. Classification Using Support Vector Machine

Classification was performed using the Support Vector Machine (SVM) algorithm. SVM was selected because it has demonstrated strong capability in handling high-dimensional feature vectors generated by deep convolutional neural networks and is particularly suitable for binary classification problems. In the proposed hybrid architecture, the feature vectors extracted by MobileNetV2 were utilized as the input to the SVM classifier to distinguish between the depression indication and non-depression indication classes. Previous studies have reported that combining deep feature extraction with SVM has shown competitive performance in various image classification tasks by integrating the representation learning capability of convolutional neural networks with the discriminative decision boundary of SVM [11], [12], [13].

The classification process employed the Radial Basis Function (RBF) kernel because it is capable of constructing nonlinear decision boundaries in high-dimensional feature spaces, enabling the classifier to capture complex relationships among the extracted facial features. Such characteristics make the RBF kernel suitable for image classification tasks where the feature distribution is not linearly separable [11].

The SVM classifier was configured using C = 10, gamma = scale, and probability = True. The parameter C controls the trade-off between maximizing the decision margin and minimizing classification errors, whereas gamma determines the influence of individual training samples on the resulting decision boundary. The gamma = scale setting automatically adjusts the kernel coefficient according to the distribution of the extracted feature vectors, providing an adaptive configuration for the classification model.

The probability = True configuration enables the SVM classifier to generate prediction probabilities instead of only class labels. These probability estimates were subsequently utilized during the threshold optimization stage to determine the optimal decision threshold for binary classification. The detailed threshold optimization procedure, including the determination of the optimal threshold value, is presented in Section II-F.

F. Threshold Optimization

The SVM classifier generated prediction probabilities for each testing sample because the probability = True configuration was enabled during model training. Instead of directly applying the default decision threshold of 0.50, a threshold optimization procedure was performed to identify the decision boundary that produced the highest classification performance for the proposed binary classification problem. Threshold optimization is commonly employed to adjust the decision boundary according to the distribution of prediction probabilities, thereby influencing the final classification performance [25].

The optimization procedure evaluated threshold values ranging from 0.30 to 0.79 with an interval of 0.01. For each threshold value, the prediction probabilities generated by the SVM classifier were converted into binary class labels, and the corresponding classification accuracy was subsequently calculated using the testing dataset.

The optimal threshold was determined based on the highest classification accuracy obtained during the evaluation process. Although the primary criterion for threshold selection was classification accuracy, modifying the decision threshold also influences the distribution of predicted positive and negative classes, which consequently affects other evaluation metrics such as precision, recall, and F1-score [25].

The optimization results indicated that a threshold value of 0.44 yielded the highest classification accuracy among all evaluated threshold values. Therefore, this threshold was adopted as the final decision boundary for distinguishing between the depression indication and non-depression indication classes during the final model evaluation.

G. Model Evaluation

Model performance was evaluated using accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC). These evaluation metrics were selected to provide a comprehensive assessment of the proposed classification model from different perspectives [11], [25].

Accuracy measures the proportion of correctly classified samples among all testing samples according to (1). Precision measures the proportion of correctly predicted depression indication samples among all samples predicted as the depression indication class according to (2). Recall measures the ability of the model to correctly identify all samples belonging to the depression indication class according to (3). The F1-score, calculated as the harmonic mean of precision and recall according to (4), provides a balanced evaluation of classification performance when considering both false positive and false negative predictions.

The Area Under the Receiver Operating Characteristic Curve (AUC) was employed to evaluate the discriminative capability of the proposed model across different decision thresholds. A higher AUC value indicates better capability of the classifier to distinguish between the depression indication and non-depression indication classes [11].

In addition to these quantitative metrics, model performance was also analyzed using a Confusion Matrix and the Receiver Operating Characteristic (ROC) Curve. The confusion matrix provides detailed information regarding the numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), thereby facilitating error analysis of the classification results. Meanwhile, the ROC curve illustrates the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) over different decision thresholds, providing a visual representation of the classifier's discriminative performance.

All evaluation metrics reported in this study were calculated using the optimal decision threshold of 0.44 obtained through the threshold optimization procedure described in Section II-F.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Precision = TP / (TP + FP) \quad (2)$$

$$Recall = TP / (TP + FN) \quad (3)$$

$$F1 - score = (2\,x\,Precision\,x\,Recall) / (Precision + Recall) \quad (4)$$

III. RESULT AND DISCUSSION

A. Classification Results Using MobileNetV2 and SVM

The proposed hybrid MobileNetV2–SVM model was evaluated using the FER2013 dataset following the experimental label mapping, feature extraction, classification, and threshold optimization procedures described in Section II. The evaluation was conducted to investigate the capability of the proposed model in classifying facial expression patterns into the experimentally defined depression indication and non-depression indication categories. The overall performance of the proposed model is summarized in Table II.

TABLE II
PERFORMANCE EVALUATION OF THE PROPOSED MOBILENET V2–SVM MODEL

Metric	Value
Accuracy	79,69%
AUC	0.8862
Optimal Threshold	0,44
Accuracy after Threshold Optimization	80.34%

As presented in Table II, the proposed model achieved an initial classification accuracy of 79.69% with an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.8862. The obtained AUC indicates that the proposed hybrid architecture consistently discriminates between the two experimentally defined classes across different decision thresholds. This result suggests that the deep feature representations extracted by MobileNetV2 provide sufficient discriminative information for the SVM classifier to distinguish facial expression patterns associated with the experimental categories adopted in this study [11], [12].

The obtained performance can be attributed to the complementary roles of MobileNetV2 and SVM within the proposed hybrid architecture. MobileNetV2, implemented through transfer learning using pretrained ImageNet weights, extracted high-level facial representations from the input images, whereas SVM classified the resulting feature vectors into the predefined experimental categories. Previous studies have demonstrated that combining deep feature extraction with conventional machine learning classifiers can improve classification performance by integrating robust feature representation learning with the discriminative capability of SVM, particularly for image classification tasks involving high-dimensional feature vectors [12], [13], [24]. These findings are consistent with recent advances in facial emotion recognition, as recent surveys have highlighted continuous improvements in deep learning models and feature representation techniques that contribute to more robust emotion recognition performance [8].

To further improve classification performance, threshold optimization was applied to the prediction probabilities generated by the SVM classifier. Decision thresholds ranging from 0.30 to 0.79 were systematically evaluated to identify the threshold that produced the highest classification accuracy. The optimization process identified 0.44 as the optimal decision threshold, increasing the classification accuracy from 79.69% to 80.34%. Although the improvement was relatively modest (0.65%), it demonstrates that adjusting the decision threshold enabled more effective utilization of the prediction probabilities generated by the classifier without modifying either the MobileNetV2 feature extractor or the SVM classification model. This finding is consistent with previous studies reporting that threshold optimization can improve classification performance by adapting the decision boundary to the probability distribution of the prediction results [25].

These findings indicate that the proposed hybrid architecture is capable of learning discriminative facial expression representations for the experimentally constructed classification task despite the absence of clinically annotated depression labels in the FER2013 dataset. Considering that FER2013 was originally developed for facial expression recognition rather than depression assessment, the reported performance should be interpreted as the capability of the proposed model to classify experimentally constructed facial expression categories instead of identifying clinically diagnosed depression. Therefore, further validation using clinically annotated depression datasets remains necessary before the proposed approach can be considered for clinical decision support or mental health screening applications [1], [20].

B. Precision, Recall, and F1-Score Analysis

Model performance was further evaluated using precision, recall, and F1-score to provide a more comprehensive assessment of the proposed classifier beyond overall accuracy. These evaluation metrics were calculated using the optimal decision threshold of 0.44, as determined through the threshold optimization procedure described in Section II-F. The classification results for each experimental class are presented in Table III.

As presented in Table III, the proposed model achieved relatively consistent precision, recall, and F1-score values across the two experimental classes. The Non-Depression Indication class obtained a slightly higher precision (0.82) than the Depression Indication class (0.77), indicating that predictions assigned to the non-depression indication class were more frequently correct. Meanwhile, the comparable recall values (0.80 and 0.79) indicate that the proposed model maintained a similar capability in identifying samples belonging to both experimental classes.

TABLE III
PERFORMANCE METRICS FOR EACH EXPERIMENTAL CLASS

Experimental Class	Precision	Recall	F1-Score	Support
Non-Depresion Indication (Class 0)	0,82	0,80	0,81	2070
Depresion Indication (Class 1)	0,77	0,79	0,78	1785
Accuracy	-	-	0,80	3855
Macro Average	0,80	0,80	0,80	3855
Weighted Average	0,80	0,80	0,80	3855

The slightly lower precision observed for the Depression Indication class suggests that a small proportion of Non-Depression Indication samples were predicted as belonging to the depression indication class. Nevertheless, the similar recall values between both classes indicate that the classifier did not substantially favor either experimental class during prediction. This behavior reflects the typical trade-off between precision and recall commonly observed in binary classification problems and demonstrates that the proposed model maintained relatively consistent classification performance across the two experimental categories.

The relatively consistent performance may be associated with the complementary roles of MobileNetV2 and SVM in the proposed hybrid architecture. MobileNetV2 extracted discriminative deep facial representations through transfer learning, whereas SVM constructed the decision boundary based on the extracted feature vectors. Previous studies have reported that combining deep feature extraction with conventional machine learning classifiers can improve classification stability by integrating robust feature representations with the discriminative capability of SVM [11], [12], [13].

The macro-average and weighted-average values both reached 0.80 for precision, recall, and F1-score. The identical values indicate that the overall evaluation was not substantially influenced by the slight difference in class proportions, suggesting that the proposed model maintained consistent performance across both experimental classes. Consequently, the reported evaluation metrics provide a reliable representation of the overall classification performance rather than being dominated by one class.

It should be emphasized that these evaluation metrics assess the capability of the proposed model to classify the experimentally defined depression indication and non-depression indication categories established through the experimental label mapping procedure. Since FER2013 does not contain clinically validated depression annotations, the reported precision, recall, and F1-score should not be interpreted as indicators of clinical depression diagnosis. Instead, these results demonstrate the feasibility of the proposed MobileNetV2–SVM architecture for classifying experimentally constructed facial expression categories, while further validation using clinically annotated depression datasets remains necessary before considering practical clinical applications [1], [20].

C. Confusion Matrix Analysis

The confusion matrix was employed to provide a detailed analysis of the classification results by illustrating the numbers of correctly and incorrectly classified samples for each experimental class. Based on the classification results using the optimal decision threshold of 0.44, the proposed model correctly classified 1,660 samples belonging to the Non-Depression Indication class (True Negative) and 1,412 samples belonging to the Depression Indication class (True Positive). Meanwhile, 410 non-depression indication samples were misclassified as depression indication (False Positive), whereas 373 depression indication samples were misclassified as non-depression indication (False Negative). The resulting confusion matrix is presented in Figure 3.

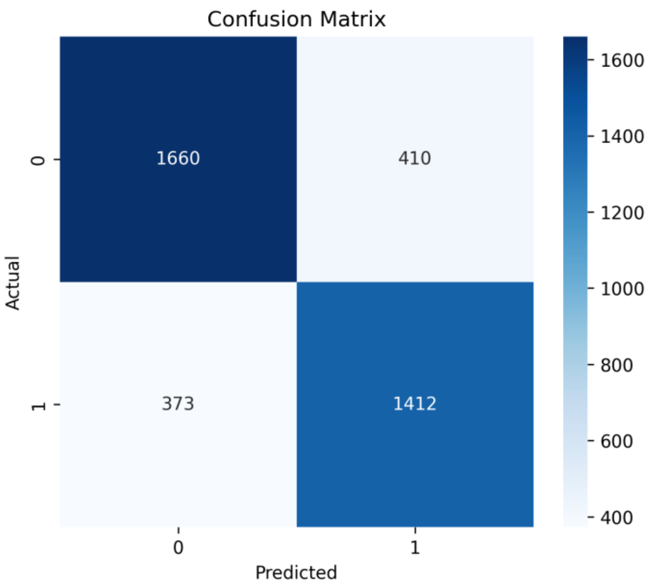

Figure 3. Confusion Matrix of the Proposed MobileNetV2-SVM Model

The confusion matrix demonstrates that the proposed classifier successfully identified the majority of samples in both experimental classes while maintaining relatively comparable numbers of false positives and false negatives. The difference between the two error types was relatively small (410 and 373, respectively), indicating that the classifier did not exhibit a strong tendency to favor either experimental class during prediction. This observation is consistent with the relatively similar precision and recall values presented in Section III-B, suggesting that the proposed model maintained consistent classification performance across both experimental classes.

The observed misclassifications may be associated with the characteristics of the FER2013 dataset and the experimental label mapping adopted in this study. Since FER2013 was originally developed for facial expression recognition rather than clinically annotated depression assessment, facial expressions assigned to different experimental classes may still share similar visual characteristics. Furthermore, the dataset consists of grayscale facial images with a relatively low spatial resolution (48 × 48 pixels), accompanied by variations in facial pose, illumination, facial occlusion, and expression intensity. These characteristics may reduce the visual distinction between some facial expression samples, making them more difficult for the proposed MobileNetV2–SVM model to classify correctly [6], [9], [10].

The relatively balanced distribution of false positives and false negatives suggests that the proposed hybrid architecture maintained consistent classification behavior without excessively favoring either experimental class. This observation is likely associated with the complementary roles of MobileNetV2 and SVM, where MobileNetV2 extracted discriminative deep facial representations through transfer learning, whereas SVM constructed the classification boundary based on the extracted feature vectors. Previous studies have likewise reported that combining deep feature extraction with conventional machine learning classifiers can improve classification robustness by integrating informative feature representations with the discriminative capability of [10], [12], [13].

It should be emphasized that the misclassifications presented in Figure 3 should not be interpreted as failures in detecting clinically diagnosed depression. Instead, they reflect the limitations of classifying experimentally constructed facial expression categories derived through the experimental label mapping procedure. Therefore, although the proposed MobileNetV2–SVM model demonstrates promising capability for distinguishing the defined experimental classes, further validation using clinically annotated depression datasets remains necessary before considering practical clinical applications [1], [20].

D. ROC Curve and AUC Analysis

The Receiver Operating Characteristic (ROC) curve was employed to evaluate the discriminative capability of the proposed MobileNetV2–SVM model across different decision thresholds. Unlike accuracy, which measures classification performance at a single decision threshold, the ROC curve illustrates the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) over the entire range of possible threshold values. Consequently, ROC analysis provides a more comprehensive assessment of the classifier's ability to distinguish between the two experimentally defined classes.

As illustrated in Figure 4, the ROC curve remains substantially above the diagonal reference line, indicating that the proposed classifier achieved considerably better discrimination than random guessing. The curve also increases steeply at relatively low false positive rates, suggesting that the classifier correctly identified a large proportion of Depression Indication samples while maintaining a relatively small number of false positive predictions. This observation is consistent with the confusion matrix analysis presented in Section III-C, where the numbers of false positives and false negatives remained relatively balanced.

The proposed model achieved an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.8862 (88.62%), indicating good discriminative capability for distinguishing between the Depression Indication and Non-Depression Indication experimental classes. This relatively high AUC suggests that the deep feature representations extracted by MobileNetV2 provided informative facial characteristics that enabled the SVM classifier to effectively separate the two experimental classes across various decision thresholds. Previous studies have likewise demonstrated that combining deep feature extraction with conventional machine learning classifiers can improve classification robustness in facial image analysis and other image classification tasks [10], [12], [13].

The ROC analysis also complements the threshold optimization procedure described in Section II-F. While the ROC curve evaluates classifier performance over the entire range of possible decision thresholds, the threshold optimization procedure identifies the operating point that provides the best classification performance for the proposed experimental task. In this study, the optimal threshold of 0.44 was selected based on the highest classification accuracy, whereas the AUC value demonstrates that the proposed classifier maintained good discriminative capability even beyond that single operating threshold.

It should be emphasized that the reported AUC value reflects the ability of the proposed model to distinguish between the experimentally defined Depression Indication and Non-Depression Indication classes established through the experimental label mapping procedure. Since FER2013 does not contain clinically validated depression annotations, the ROC and AUC results should not be interpreted as evidence of clinical depression detection. Instead, these findings demonstrate the feasibility of the proposed MobileNetV2–SVM architecture for classifying experimentally constructed facial expression categories, while further validation using clinically annotated depression datasets remains necessary before practical clinical applications can be considered [1], [20].

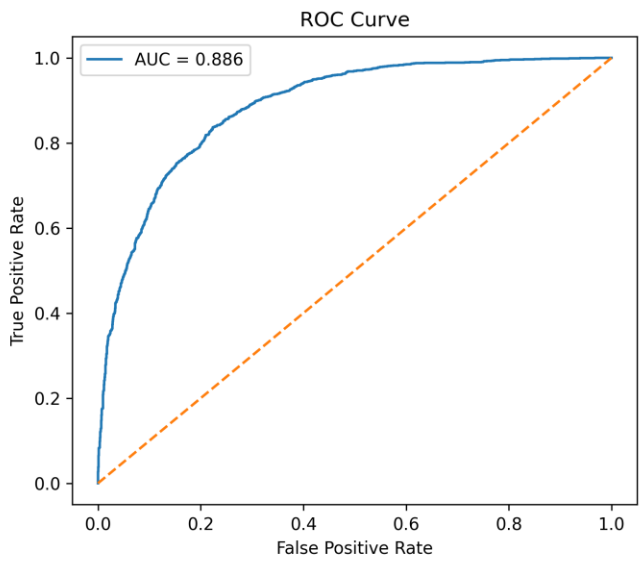

Figure 4. Receiver Operating Characteristic (ROC) Curve of the Proposed MobileNetV2–SVM Model

E. Discussion of Threshold Optimization and Comparison with Previous Studies

Threshold optimization was performed to determine the operating point that yielded the highest classification accuracy for the proposed MobileNetV2–SVM model. Instead of adopting the default decision threshold of 0.50, threshold values ranging from 0.30 to 0.79 with an interval of 0.01 were systematically evaluated, as described in Section II-F. The experimental results indicated that the optimal threshold was 0.44, resulting in the highest classification accuracy of 80.34%, compared with 79.69% obtained before threshold optimization. Although the improvement was relatively small, these findings indicate that selecting an appropriate decision threshold can provide a more suitable operating point for the proposed binary classification task.

The threshold optimization results are consistent with the evaluation presented in the previous subsections. The confusion matrix demonstrated relatively balanced false positive and false negative predictions, while the ROC analysis yielded an AUC of 0.8862, indicating good discriminative capability across different decision thresholds. Together, these findings suggest that the probability estimates generated by the SVM classifier were sufficiently discriminative to enable effective threshold adjustment, allowing the proposed model to achieve its best classification performance at a threshold of 0.44.

To further evaluate the proposed approach, the experimental results were compared with several previous studies related to facial expression-based depression analysis, as summarized in Table IV. However, direct comparison of classification performance should be interpreted cautiously because the studies employed different datasets, data modalities, annotation strategies, and experimental objectives. Most previous studies utilized clinically diagnosed depression datasets or facial expression videos, whereas the present study employed the FER2013 facial expression dataset that was experimentally reorganized into Depression Indication and Non-Depression Indication classes through the experimental label mapping procedure. Consequently, the comparison presented in Table IV primarily highlights methodological differences rather than providing a direct benchmark of classification performance [3], [20].

TABLE IV
COMPARISON OF THE PROPOSED METHOD WITH PREVIOUS STUDIES

Study	Method	Dataset	Modality	Dataset Size	Main Findings
[21]	Facial expression analysis	Clinical depression dataset	Static facial expressions	Older adult participants	Facial expressions were significantly associated with depressive symptoms.
[15]	LSTM with Attentional Intermediate Feature Fusion and Label Smoothing	Clinical interview videos	Video	474 videos	Accuracy 91,67%
[16]	Dual-Stream Multiple Instance Learning	Facial expression videos	Video	150 videos	Accuracy 74,7%
This Study	MobileNetV2 + Support Vector Machine (SVM)	FER2013 (Experimental Label Mapping)	Static facial images	19,275 images	Accuracy 80,34% (Optimal Threshold = 0.44)

Although the proposed model achieved an accuracy of 80.34%, this result should not be directly compared with studies reporting higher or lower classification performance because the underlying classification tasks differ substantially. Previous studies investigated clinically diagnosed depression using facial images or videos, whereas the present study evaluated experimentally constructed facial expression categories derived from FER2013. Furthermore, differences in dataset characteristics, image modality, sample size, annotation strategy, and model architecture also contribute to variations in classification performance. Therefore, the comparison presented in Table IV should be interpreted as a comparison of methodological approaches rather than as a direct comparison of classification accuracy.

Despite these differences, the proposed MobileNetV2–SVM architecture demonstrated promising classification performance for the experimentally defined facial expression classification task. The transfer learning capability of MobileNetV2 enabled the extraction of informative facial feature representations, while the SVM classifier effectively separated the extracted feature vectors using a nonlinear decision boundary. These findings are consistent with previous studies reporting that hybrid CNN–machine learning approaches provide an effective strategy for image classification, particularly when computational efficiency and limited training data are important considerations [10], [12], [13], [24].

Nevertheless, several limitations should be considered when interpreting the findings of this study. First, the FER2013 dataset was originally developed for facial expression recognition rather than clinical depression assessment. Therefore, the Depression Indication and Non-Depression Indication classes were established through an experimental label mapping strategy instead of clinically validated depression diagnoses. Second, the proposed model utilized static facial expression images, which do not capture temporal facial dynamics that may contain additional behavioral information associated with depressive conditions. Consequently, the proposed model should be interpreted as an exploratory facial expression classification model rather than a clinical depression diagnosis system. Future studies should validate the proposed approach using clinically annotated depression datasets, investigate fine-tuning strategies for MobileNetV2, and evaluate larger facial expression datasets to further examine the robustness and generalizability of the proposed framework [1], [2], [20].

IV. CONCLUSION

This study developed a facial expression-based Depression Indication classification model using MobileNetV2 as a feature extractor and Support Vector Machine (SVM) as a classifier. The FER2013 dataset was experimentally reorganized into Depression Indication and Non-Depression Indication classes based on facial expression grouping. Experimental results showed that the proposed model achieved an accuracy of 79.69% and an AUC of 0.8862. Through threshold optimization, the optimal decision threshold of 0.44 improved the classification accuracy to 80.34%. In addition, the precision, recall, and F1-score values demonstrated relatively balanced classification performance across both classes, indicating that the proposed model did not exhibit a dominant prediction bias toward either class. The experimental results demonstrated that the combination of MobileNetV2 for deep feature extraction, GlobalAveragePooling2D for feature representation, and SVM for classification provided an effective hybrid architecture for the experimental facial expression classification task investigated in this study. Furthermore, threshold optimization contributed to improved classification performance by identifying a more suitable decision boundary than the default threshold. Nevertheless, the findings of this study should be interpreted within the scope of the experimental design because the FER2013 dataset was originally developed for facial expression recognition rather than clinical depression assessment. Therefore, the proposed model should be considered as a facial expression-based Depression Indication classification model rather than a clinical depression diagnosis system. Future research should validate the proposed approach using clinically annotated depression datasets, investigate fine-tuning strategies for MobileNetV2, and compare different feature extraction and classification methods to further improve the robustness and generalizability of the proposed framework.

REFFERENCES

[1] M. Squires *et al.*, “Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment,” Dec. 01, 2023, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1186/s40708-023-00188-6.

[2] N. K. Iyortsuun, S. H. Kim, M. Jhon, H. J. Yang, and S. Pant, “A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis,” Feb. 01, 2023, *MDPI*. doi: 10.3390/healthcare11030285.

[3] M.-M. Han *et al.*, “Automatic recognition of depression based on audio and video: A review,” *World J. Psychiatry*, vol. 14, no. 2, pp. 225–233, Feb. 2024, doi: 10.5498/wjp.v14.i2.225.

[4] R. Guo, H. Guo, L. Wang, M. Chen, D. Yang, and B. Li, “Development and application of emotion recognition technology — a systematic literature review,” Dec. 01, 2024, *BioMed Central Ltd*. doi: 10.1186/s40359-024-01581-4.

[5] E. M. G. Younis, S. Mohsen, E. H. Houssein, and O. A. S. Ibrahim, “Machine learning for human emotion recognition: a comprehensive review,” Jun. 01, 2024, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1007/s00521-024-09426-2.

[6] E. G. Dada, D. O. Oyewola, S. B. Joseph, O. Emebo, and O. O. Oluwagbemi, “Facial Emotion Recognition and Classification Using the Convolutional Neural Network-10 (CNN-10),” *Applied Computational Intelligence and Soft Computing*, vol. 2023, 2023, doi: 10.1155/2023/2457898.

[7] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, “Facial emotion recognition using transfer learning in the deep CNN,” *Electronics (Switzerland)*, vol. 10, no. 9, May 2021, doi: 10.3390/electronics10091036.

[8] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, “Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets,” *Information (Switzerland)*, vol. 15, no. 3, Mar. 2024, doi: 10.3390/info15030135.

[9] Q. Zhu, H. Zhuang, M. Zhao, S. Xu, and R. Meng, “A study on expression recognition based on improved mobilenetV2 network,” *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-58736-x.

[10] H. M. Shahzad, S. M. Bhatti, A. Jaffar, S. Akram, M. Alhajlah, and A. Mahmood, “Hybrid Facial Emotion Recognition Using CNN-Based Features,” *Applied Sciences (Switzerland)*, vol. 13, no. 9, May 2023, doi: 10.3390/app13095572.

[11] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, “An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review,” *Information (Switzerland)*, vol. 15, no. 4, Apr. 2024, doi: 10.3390/info15040235.

[12] M. J. Adamu *et al.*, “Efficient and Accurate Brain Tumor Classification Using Hybrid MobileNetV2–Support Vector Machine for Magnetic Resonance Imaging Diagnostics in Neoplasms,” *Brain Sci.*, vol. 14, no. 12, Dec. 2024, doi: 10.3390/brainsci14121178.

[13] M. Azeez Joodi, M. Hadi Saleh, and D. Jasim Kadhim, “A New Proposed Hybrid Learning Approach with Features for Extraction of Image Classification,” *Journal of Robotics*, vol. 2023, 2023, doi: 10.1155/2023/9961421.

[14] W. Guo, H. Yang, Z. Liu, Y. Xu, and B. Hu, “Deep Neural Networks for Depression Recognition Based on 2D and 3D Facial Expressions Under Emotional Stimulus Tasks,” *Front. Neurosci.*, vol. 15, Apr. 2021, doi: 10.3389/fnins.2021.609760.

[15] Y. Mahayossanunt, N. Nupairoj, S. Hemrungrojn, and P. Vateekul, “Explainable Depression Detection Based on Facial Expression Using LSTM on Attentional Intermediate Feature Fusion with Label Smoothing,” *Sensors*, vol. 23, no. 23, Dec. 2023, doi: 10.3390/s23239402.

[16] Z. Shangguan, Z. Liu, G. Li, Q. Chen, Z. Ding, and B. Hu, “Dual-Stream Multiple Instance Learning for Depression Detection With Facial Expression Videos,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 554–563, 2023, doi: 10.1109/TNSRE.2022.3204757.

[17] R. Wang *et al.*, “FacialPulse: An Efficient RNN-based Depression Detection via Temporal Facial Landmarks,” in *MM 2024 - Proceedings of the 32nd ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, Oct. 2024, pp. 311–320. doi: 10.1145/3664647.3681546.

[18] H. Arabian, T. Abdulbaki Alshirbaji, J. G. Chase, and K. Moeller, “Emotion Recognition beyond Pixels: Leveraging Facial Point Landmark Meshes,” *Applied Sciences (Switzerland)*, vol. 14, no. 8, Apr. 2024, doi: 10.3390/app14083358.

[19] S. Alsubai, A. Alqahtani, A. Alanazi, M. Sha, and A. Gumaei, “Facial emotion recognition using deep quantum and advanced transfer learning mechanism,” *Front. Comput. Neurosci.*, vol. 18, 2024, doi: 10.3389/fncom.2024.1435956.

[20] K. Mao, Y. Wu, and J. Chen, “A systematic review on automated clinical depression diagnosis,” Dec. 01, 2023, *Nature Publishing Group*. doi: 10.1038/s44184-023-00040-z.

[21] H. Kim *et al.*, “Facial Expressions Track Depressive Symptoms in Old Age,” *Sensors*, vol. 23, no. 16, Aug. 2023, doi: 10.3390/s23167080.

[22] M. Monferrer, A. S. García, J. J. Ricarte, M. J. Montes, A. Fernández-Caballero, and P. Fernández-Sotos, “Facial emotion recognition in patients with depression compared to healthy controls when using human avatars,” *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-31277-5.

[23] R. A. Elsheikh, M. A. Mohamed, A. M. Abou-Taleb, and M. M. Ata, “Improved facial emotion recognition model based on a novel deep convolutional structure,” *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-79167-8.

[24] A. W. Salehi *et al.*, “A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope,” Apr. 01, 2023, *MDPI*. doi: 10.3390/su15075930.

[25] Y. Terhorst, L. B. Sander, D. D. Ebert, and H. Baumeister, “Optimizing the predictive power of depression screenings using machine learning,” *Digit. Health*, vol. 9, Jan. 2023, doi: 10.1177/20552076231194939.