Journal of Information Systems and Informatics

**Vol. 7, No. 2, March 2025** e-ISSN: **2656-4882** p-ISSN: **2656-5935**

DOI: 10.51519/journalisi.v7i2.1071 Published By DRPM-UBD

# EfficientNet B0 Feature Extraction with L2-SVM Classification for Robust Facial Expression Recognition

**Ahmad Taufiq Akbar[1], Shoffan Saifullah[2,3], Hari Prapcoyo[4], Heru Rustamadji[5], Nur Heri Cahyana[6]**

[1,2,4,5,6]Department of Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Yogyakarta, Indonesia
[3]Faculty of Computer Science, AGH University of Krakow, Krakow, Poland
Email: [1]ahmadtaufiq.akbar@upnyk.ac.id, [2]shoffans@upnyk.ac.id, [3]saifulla@agh.edu.pl, [4]hari.prapcoyo@upnyk.ac.id, [5]herucr@upnyk.ac.id, [6]nur.hericahyana@upnyk.ac.id

## Abstract

Facial expression recognition (FER) remains a challenging task due to subtle visual distinctions among emotion classes and the limitations of small, controlled datasets. Traditional deep learning methods often require extensive training and data augmentation to generalize effectively. This paper proposes a hybrid FER framework combining EfficientNet B0 as a deep feature extractor with an L2-regularized Support Vector Machine (L2-SVM) classifier. Notably, no data augmentation is used, emphasizing the method's effectiveness under minimal preprocessing. Experimental results on the JAFFE and CK+ datasets show superior performance, with up to 100% accuracy on various hold-out splits and 99.8% under 5-fold cross-validation. Precision, recall, and F1-score exceeded 95% across all emotion classes. Confusion matrix analysis reveals perfect classification for distinctive emotions like Happiness and Surprise, while subtle misclassifications occur in ambiguous categories such as Fear and Sadness. The findings validate the model's efficiency, generalization, and potential for real-time FER tasks on resource-constrained platforms.

**Keywords**: Facial Expression Recognition, EfficientNet, SVM, Deep Features, Emotion Classification.

## 1. INTRODUCTION

Facial expression recognition (FER) has emerged as a pivotal research area in computer vision, human-computer interaction, and affective computing [1], [2]. Emotions, as fundamental components of human behavior, are visually conveyed through facial expressions, providing significant cues to an individual's emotional state and intentions [3], [4]. FER is vital in real-world applications such as healthcare monitoring, virtual assistants, entertainment, security, online education, gaming, and marketing [5]. For instance, FER systems can enhance user experience by adapting interfaces based on emotional states, support therapeutic interventions by tracking patient emotions, and improve customer sentiment analysis in commercial settings [2], [6], [7].

Despite its wide applications, FER remains challenging due to several complexities inherent to human facial expressions [8]. Variations in illumination, head pose, occlusion, facial appearance, and subtle differences between expressions contribute to performance limitations in real-world scenarios [1], [2]. These challenges are especially significant when working with limited or imbalanced datasets, leading to model overfitting and poor generalization [8]. Additionally, low-resolution images further exacerbate the difficulty of learning meaningful features, especially when deep learning models are trained from scratch [9].

Over the past decade, numerous studies have attempted to address these issues. Traditional methods such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Gabor filters have been widely applied to extract handcrafted features, which are then classified using conventional machine learning algorithms like Support Vector Machines (SVM) or k-Nearest Neighbors (k-NN) [10]–[12]. However, these handcrafted approaches often lack the representational power needed for complex patterns present in facial expressions. Deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized FER by automatically learning hierarchical representations of features. Several researchers have explored various CNN architectures such as VGGNet, ResNet, and custom-designed shallow CNNs to improve classification accuracy on datasets like JAFFE and CK+ [13]–[15]. However, most of these models require large-scale data and extensive fine-tuning, making them unsuitable for lightweight applications. Many rely on aggressive data augmentation to simulate diversity, which may introduce distortions not representative of actual conditions.

EfficientNet B0, in contrast, delivers high accuracy with fewer parameters through compound scaling of depth, width, and resolution. Compared to older CNNs like VGG or ResNet, it provides a better balance of efficiency and representational power, making it ideal for small FER datasets and real-time deployment scenarios. Despite these advances, there are notable gaps in the literature. Few studies have systematically explored the effectiveness of **intermediate pretrained layers in lightweight CNNs** for FER. Furthermore, the integration of EfficientNet with simple yet robust classifiers like L2-regularized SVM, particularly without data augmentation, remains underexplored.

To address these gaps, this paper proposes a novel hybrid approach using the pretrained EfficientNet B0 model for deep feature extraction specifically utilizing the 181st batch normalization layer combined with L2-SVM for classification. By decoupling feature learning and classification, the approach improves accuracy while maintaining computational simplicity and avoiding augmentation. The key contributions of this paper are as follows.

1) We propose a hybrid FER model using EfficientNet B0 feature extraction and L2-SVM classification, without data augmentation or fine-tuning.
2) We identify and validate the optimal batch normalization layer (181st) of EfficientNet B0 for FER tasks.
3) We evaluate the model under multiple hold-out and cross-validation setups to demonstrate performance robustness.
4) We compare our results with state-of-the-art methods and show that our model achieves competitive or superior accuracy with lower computational cost.

The remainder of this article is structured as follows: Section 2 details the methodology, including dataset descriptions, preprocessing, feature extraction, and classification pipeline. Section 3 presents and compares the experimental results with related studies. Finally, Section 4 concludes the paper by summarizing key findings and outlining directions for future research, such as expanding the model evaluation to more diverse and complex real-world FER datasets.

## 2. RELATED WORKS

Facial Expression Recognition (FER) has been extensively studied using a range of techniques, from traditional handcrafted approaches to deep learning-based methods. Early FER systems typically relied on handcrafted features such as Local Binary Patterns (LBP) [10], Histogram of Oriented Gradients (HOG) [11], and Gabor filters [12]. These features capture texture, edge orientation, and frequency-based representations, which are then classified using machine learning algorithms like Support Vector Machines (SVM) or k-Nearest Neighbors (k-NN). Although these methods are computationally efficient, they often struggle to generalize well due to their limited feature expressiveness in complex or ambiguous emotional states.

The advent of deep learning—particularly Convolutional Neural Networks (CNNs)—marked a major advancement in FER. Deep networks learn hierarchical features directly from pixel data, improving accuracy and robustness. Architectures such as VGGNet, ResNet, and custom CNNs have been widely applied to datasets like JAFFE and CK+, offering state-of-the-art performance [13]–[15]. For example, Akhand et al. [14] fine-tuned a VGG16 model to achieve 100% accuracy on JAFFE using a 90:10 split, while Li et al. [13] achieved 95.7% using a CNN-SVM hybrid. Despite this success, CNN-based methods typically require large training datasets and benefit significantly from data augmentation to prevent overfitting—something not always feasible with small, controlled FER datasets.

To bridge this gap, hybrid models have emerged that combine deep feature extraction with classical machine learning classifiers. For instance, pretrained CNNs such as CaffeNet and VGG are used to extract high-level features, which are then classified using SVM. This decouples the learning of representations from classification and reduces the need for end-to-end fine-tuning. However, most hybrid approaches still rely on standard deep architectures with high computational costs and often apply data augmentation or manual preprocessing steps like face alignment, which may not be ideal for constrained or real-time environments.

EfficientNet B0, introduced as a highly scalable and lightweight CNN, presents a promising alternative for FER in low-resource scenarios. Its compound scaling strategy achieves state-of-the-art performance with significantly fewer parameters compared to VGG or ResNet [16]. Despite its advantages, few studies have explored EfficientNet as a fixed feature extractor for small-scale FER tasks, nor have they combined it with L2-regularized SVM in a purely augmentation-free setting. This study fills that gap by proposing a hybrid FER model that utilizes EfficientNet B0's intermediate features and classifies them using L2-SVM—without any data augmentation or fine-tuning. The following section describes the proposed methodology in detail, including dataset preparation, feature extraction, and classification steps.

## 3. METHODS

This section outlines the methodology adopted in this research, encompassing dataset acquisition, preprocessing, feature extraction using the EfficientNet B0 model, and classification with the L2-SVM algorithm. The framework is designed to leverage a lightweight deep learning model for feature extraction and a classical machine learning classifier to optimize recognition performance, especially on small-scale datasets.

### 3.1. Dataset Description

This study utilizes two benchmark datasets commonly used in facial expression recognition (FER): the Japanese Female Facial Expression (JAFFE) dataset [17]–[19] and the Extended Cohn-Kanade (CK+) dataset [20]. Both datasets contain grayscale facial images labeled with one of seven universal emotion categories: anger, disgust, fear, happiness, neutral, sadness, and surprise.

### 1) JAFFE Dataset

The JAFFE dataset consists of 213 grayscale images collected from 10 Japanese female subjects. Each subject displays all seven expressions under controlled conditions. The images have a consistent resolution of 256×256 pixels, with faces well-aligned, frontal, and centered in the frame. This uniformity in pose and lighting makes the dataset widely adopted for FER studies. However, the limited sample size presents a significant challenge for deep learning models, which often require large datasets to generalize well and avoid overfitting.

### 2) CK+ Dataset

The CK+ dataset includes 981 grayscale images from 123 subjects, capturing both posed and spontaneous expressions. It offers greater variability than JAFFE, introducing slight differences in expression intensity and head pose. The original resolution of the CK+ images is 48×48 pixels, which is substantially lower than JAFFE. While the reduced resolution introduces challenges for feature extraction, the variability contributes to a more realistic testing scenario that reflects everyday human expressions.

### 3) Class Distribution and Folder Structure

Both datasets are organized into subfolders representing the seven emotion classes, enabling straightforward label extraction during the preprocessing stage. In the JAFFE dataset, each emotion class contains approximately 30 to 32 images, while the CK+ dataset provides a more balanced distribution across its larger sample size.

Figure 1 presents sample images from both datasets, showcasing the visual diversity across all seven emotion classes. The top row displays examples from the JAFFE dataset, which demonstrate high clarity and alignment. The bottom row contains CK+ images, highlighting variations in expression style and pose that make this dataset more complex and realistic.

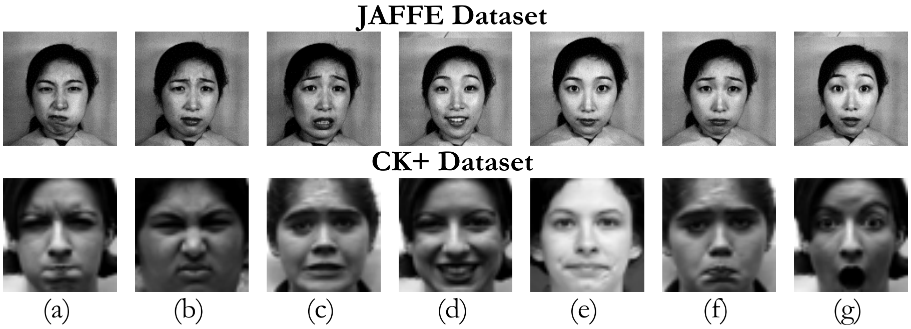

**Figure 1.** Sample images from the JAFFE and CK+ datasets showing the seven facial expression categories: (a) anger, (b) disgust, (c) fear, (d) happiness, (e) neutral, (f) sadness, and (g) surprise.

### 3.2. Preprocessing

Preprocessing is pivotal in preparing input data for robust facial expression recognition (FER) [21], mainly when working with small and diverse datasets like JAFFE and CK+. In this study, the preprocessing pipeline is deliberately simplified to isolate the contribution of deep feature extraction via the EfficientNet B0 model without the confounding influence of data augmentation or handcrafted preprocessing techniques [22].

The first step involves resizing all input images to 224×224 pixels, matching the required input size of EfficientNet B0 (originally trained on ImageNet). For the JAFFE dataset (256×256 pixels), a two-stage resizing process is used: images are first downscaled to 128×128 to reduce memory usage, and then upscaled to 224×224 using bilinear interpolation. For the CK+ dataset (48×48 pixels), images are directly upsampled to 224×224. This standardization step reduces variability introduced by differing image resolutions. The resizing operation can be mathematically described as in Equation 1.

$$I_{resized}\left(x^{'},y^{'}\right)=I\left(\frac{x^{'}}{s_x},\frac{y^{'}}{s_y}\right) \quad (1)$$

where $I_{resized}(x',y')$ represents the resized image, $I(x,y)$ is the original image, and $s_x, s_y$ are the scaling factors along the horizontal and vertical dimentions, respectively.

Following resizing, all pixel intensities are normalized to the range [0, 1], improving the convergence and stability of convolution operations in the EfficientNet B0 architecture. The normalization is defined as in Equation 2.

$$I_{norm}\left(x,y\right)=\frac{I\left(x,y\right)}{255} \quad (2)$$

where $I(x,y)$ is the original grayscale intensity at position $(x,y)$, and $I_{norm}(x,y)$ is the normalized pixel value.

No data augmentation is applied in this study—such as flipping, rotation, brightness adjustment, or cropping. While these techniques are commonly used to improve generalization in FER models, they can introduce distortions that misrepresent the controlled conditions of JAFFE and CK+. Similarly, no handcrafted preprocessing (e.g., Local Binary Patterns, Histogram of Oriented Gradients, Gabor filters, or facial landmark alignment) is employed. These traditional techniques are deliberately excluded to avoid overlapping or interfering with the hierarchical feature learning of the deep CNN.

This minimal preprocessing strategy aims to highlight the true performance of EfficientNet B0 as a deep feature extractor. By focusing exclusively on raw-to-deep representation learning—without engineered features or artificial expansion—the study seeks to assess whether a pretrained EfficientNet B0, combined with L2-SVM, can generalize effectively even under limited data conditions.

### 3.3. Evaluation Metrics

To quantitatively assess the performance of the proposed method, four standard classification metrics were adopted [33]: Accuracy, Precision, Recall, and F1-score [34]. These metrics are derived from the confusion matrix values, which include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) per class [35], [36].

1) Accuracy reflects the overall correctness of the model's predictions across all classes and is defined in Equation 3.

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

While accuracy offers a general performance overview, it may not fully reflect classification behavior in the presence of class imbalance.

2) Precision measures the proportion of correct positive predictions made for each emotion class as defined in Equation 4.

$$Precision=\frac{TP}{TP+FP} \tag{4}$$

A higher precision indicates that the classifier makes fewer false positive errors.

3) Recall evaluates the ability of the model to correctly identify all actual positive samples as defined in Equation 5.

$$Recall=\frac{TP}{TP+FN} \tag{5}$$

High recall ensures fewer false negatives, which is critical when distinguishing subtle emotions.

4) F1-score represents the harmonic mean of precision and recall, balancing the trade-off between the two as defined in Equation 6.

$$F1\text{-}Score=2\times\frac{Precision\ .\ Recall}{Precision+Recall} \tag{6}$$

F1-score is particularly useful in multi-class FER tasks where emotion categories such as sadness and fear can be easily confused, and precision-recall trade-offs matter.

These metrics are computed per class and macro-averaged to evaluate overall model performance across the seven emotion categories. This multi-metric approach provides a comprehensive evaluation, capturing not just general accuracy but also how well each emotion is recognized.

### 3.4. Feature Extraction using EfficientNet B0

EfficientNet B0 is employed in this research as a deep feature extractor for facial expression recognition [23]. It is a modern convolutional neural network (CNN) architecture that applies compound scaling to uniformly adjust the network's depth, width, and input resolution, achieving high efficiency with fewer parameters [24], [25]. This makes it ideal for handling small datasets such as JAFFE and CK+ without the risk of overfitting common in larger CNN models. EfficientNet B0's architecture is divided into a stem layer, multiple stages of Mobile Inverted Bottleneck Convolution (MBConv) blocks, and top layers [26], as detailed below:

1) Stem Layer:
   a) A single 3×3 convolution followed by batch normalization and a Swish activation function to process raw input images.
   b) This is not explicitly listed in the table as it's outside the focus of the extracted BN layers.
2) MBConv Blocks and Batch Normalization Layers:
   a) The backbone of the network consists of MBConv blocks grouped into stages from block2 to block7, each followed by a batch normalization (BN) layer.
   b) These BN layers are where feature extraction occurs in this study.

Table 1 shows the selected batch normalization layers from the EfficientNet B0 model, indexed according to their position in the original architecture.

**Table 1.** Selected Batch Normalization Layers from EfficientNet B0

| Layer Index | Layer Name | Stage | Layer Index | Layer Name | Stage |
|---|---|---|---|---|---|
| 36 | block2b_bn | Block 2 | 152 | block5c_bn | Block 5 |
| 52 | block3a_bn | Block 3 | 168 | block6a_bn | Block 6 |
| 65 | block3b_bn | Block 3 | 181 | block6b_bn | Block 6 |
| 81 | block4a_bn | Block 4 | 196 | block6c_bn | Block 6 |
| 94 | block4b_bn | Block 4 | 211 | block6d_bn | Block 6 |
| 109 | block4c_bn | Block 4 | 226 | block7a_bn | Block 7 |

| Layer Index | Layer Name | Stage | Layer Index | Layer Name | Stage |
|---|---|---|---|---|---|
| 124 | block5a_bn | Block 5 | 236 | top_bn | Top |
| 137 | block5b_bn | Block 5 | 239 | top_dropout | Top |

The MBConv blocks consist of:

1) Depthwise separable convolutions, which reduce the number of parameters by factorizing spatial and channel convolutions.
2) Squeeze-and-Excitation (SE) blocks, which adaptively recalibrate feature maps by applying channel-wise attention.
3) Batch normalization (BN) layers, which stabilize feature distributions by normalizing activations and applying learned affine transformations.

In this research, the batch normalization layers serve as the points of feature extraction. These BN layers output feature maps F after the convolutional and SE blocks have processed the input images. The deeper the block (e.g., block6b_bn from Block 6), the more abstract and semantically meaningful the captured features are, such as emotional facial patterns, contours, and deformation signatures. For a normalized input image $I_{norm} \epsilon R^{244\times244\times3}$, the feature extraction process through a selected BN layer $\phi_l$ can be expressed in Equation 7.

$$F=\phi_l(I_{norm}) \tag{7}$$

Where $F\epsilon\mathbb{R}^{H\times W\times C}$ represents the intermediate feature tensor, and lll is the index of the BN layer (e.g., $l$=181l for block6b_bn). The tensor F is then flattened into a one-dimensional vector can be expressed in Equation 8.

$$f=Flatten\ F \in R^d \tag{8}$$

where d=HxWxC is the total number of elements in the tensor, corresponding to the feature vector length.

The BN layers closer to the earlier blocks, such as block2b_bn (index 36) and block3a_bn (index 52), tend to capture low-level features such as edges, textures, and simple patterns. Conversely, deeper BN layers like block5b_bn (index 137), block6b_bn (index 181), and block7a_bn (index 226) focus on high-level semantic attributes, including complex facial shapes, emotional cues, and spatial hierarchies that are critical for expression recognition. After extensive evaluation of these layers, block6b_bn (index 181) was selected as the primary feature extraction layer due to its optimal balance between feature richness and dimensional compactness. It provides semantically rich descriptors ideal for small datasets while maintaining computational efficiency. All layers of EfficientNet B0 are kept frozen during training, preserving pretrained weights from ImageNet. This approach allows the model to leverage general-purpose visual features without retraining, minimizing the risk of overfitting to small datasets like JAFFE and CK+.

### 3.5. Classification using L2-SVM

Following feature extraction using EfficientNet B0, this research adopts the L2-regularized Support Vector Machine (L2-SVM) [27] as the classifier to predict facial expression categories [28]. The use of L2-SVM is motivated by its effectiveness in handling high-dimensional input vectors, such as those produced by convolutional neural networks while maintaining a relatively simple and interpretable structure compared to deep classifiers [29].

The Support Vector Machine is a supervised machine learning algorithm that seeks to find the optimal hyperplane separating different classes by maximizing the margin between them [30], [31]. In this context, the extracted feature vectors $f_i$ from EfficientNet B0 are used as input for classification. The L2-SVM classifier aims to minimize both the classification loss and the complexity of the model through L2 regularization, which penalizes large weight magnitudes [32]. This balance can be formalized using Equation 9.

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} max\left(0,1\text{-}y_i\left(w^T f_i+b\right)\right) \tag{9}$$

In this equation, the term $\frac{1}{2}\|w\|^2$ represents the L2 regularization term, where $\|w\|^2$ is the squared Euclidean norm of the weight vector www. This term serves to prevent the classifier from overfitting by discouraging overly complex decision boundaries with large weights. The second term, $C\sum_{i=1}^{n} \max\left(0,1\text{-}y_i(w^T f_i+b)\right)$, represents the hinge loss. Here, $y_i \epsilon\{-1,+1\}$ is the true label of the *i*-th sample, $f_i$ is the corresponding feature vector, C is a positive regularization parameter controlling the trade-off between maximizing the margin and minimizing misclassification, and bbb is the bias term.

In contrast to other SVM variants, L2-regularization generates smoother decision boundaries and ensures that all features derived from the deep network contribute to the classification task. This characteristic is particularly beneficial in facial expression recognition, where subtle variations across all dimensions of a deep feature vector may contain valuable emotional cues.

Since the problem at hand involves multi-class classification with seven categories (anger, disgust, fear, happiness, neutral, sadness, and surprise), this research implements a one-vs-rest (OvR) classification scheme. In this approach, a separate binary classifier is trained for each class, treating samples of that class as positive and all others as negative. During prediction, the classifier yielding the highest decision function score is selected. This selection mechanism can be mathematically expressed as:

$$\hat{y}=arg \max_{k\in\{1,2,\ldots,K\}} w_k{}^{T}f_i+b_k \tag{10}$$

where K=7 represents the number of classes, $w_k$ and $b_k$ are the learned weights and bias for class $k$, and $\hat{y}$ denotes the predicted label for sample $i$. The dot product $w_k{}^{\mathsf{T}}\mathsf{f}_i$ measures how strongly the feature vector $\mathsf{f}_i$ is associated with class kkk under the learned hyperplane.

The classification process is further validated using two strategies. The first is hold-out validation, where the dataset is divided into training and testing sets with various split ratios (e.g., 90:10, 85:15, 80:20, and 70:30). The second strategy involves k-fold cross-validation, where the data is partitioned into k subsets (with k = 5 or k = 10), and the model is iteratively trained and tested across these folds to ensure robust performance estimation. Both strategies are essential to assess the generalization ability of the L2-SVM classifier when working with small datasets.
The integration of EfficientNet B0 with L2-SVM creates a two-stage pipeline, where deep semantic features extracted from batch normalization layers (e.g., block6b_bn) are fed into the linear SVM classifier. This hybrid setup combines the power of deep learning for feature abstraction with the interpretability and efficiency of classical machine learning classifiers. Furthermore, as SVM's optimization problem is convex, training can be performed efficiently even when dealing with moderately high-dimensional feature vectors derived from deep CNN layers.

### 3.6. Research Workflow Overview

To provide a comprehensive view of the proposed facial expression recognition pipeline, Figure 2 illustrates the full research workflow, from data acquisition to model evaluation. This modular pipeline combines deep feature extraction using EfficientNet B0 with L2-SVM classification, explicitly designed to operate without augmentation or handcrafted preprocessing.

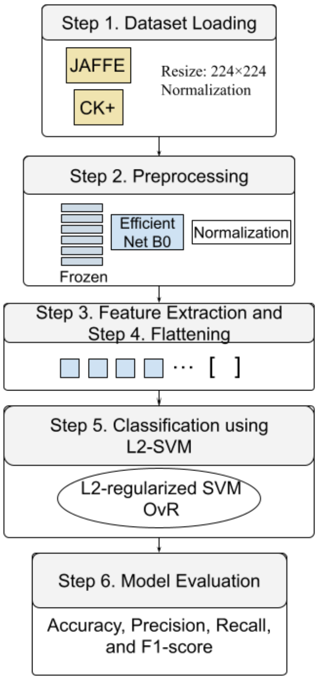

**Figure 2.** Overall research pipeline of the proposed FER method: (1) dataset loading, (2) preprocessing (resizing and normalization), (3) feature extraction using EfficientNet B0 (frozen), (4) feature vector flattening, (5) classification using L2-regularized SVM (with one-vs-rest strategy), and (6) evaluation via Accuracy, Precision, Recall, and F1-score.

1) **Step 1: Dataset Loading -** Facial images are obtained from two publicly available benchmark datasets: JAFFE and CK+. These datasets are organized into subfolders corresponding to the seven universal facial expression classes.
2) **Step 2: Preprocessing -** All images are resized to 224×224 pixels to match the input requirement of EfficientNet B0. Normalization is then applied to scale the pixel values to the [0,1][0, 1][0,1] range. **No data augmentation or handcrafted feature preprocessing** is applied, ensuring that evaluation focuses solely on the deep features extracted from raw data.
3) **Step 3: Deep Feature Extraction using EfficientNet B0 -** Each normalized image is passed through a frozen EfficientNet B0 model pretrained on ImageNet. Intermediate feature maps are extracted from a selected batch normalization (BN) layer—specifically block6b_bn (layer index 181)—which has been empirically determined to yield the most semantically rich and compact representations for this task.
4) **Step 4: Feature Vector Flattening -** The extracted feature tensor $F\epsilon\mathbb{R}^{H\times W\times C}$ is flattened into a one-dimensional vector $F\epsilon\mathbb{R}^{d}$, where $d =$

$H \times W \times C$. This vector represents a high-level encoding of the facial expression contained in the image.

5) **Step 5: Classification using L2-SVM -** The flattened feature vectors are input into a linear L2-regularized Support Vector Machine. A one-vs-rest (OvR) classification strategy is used to handle the seven-class FER task. Each class-specific SVM learns to distinguish one expression from the rest based on these high-dimensional feature vectors.
6) **Step 6: Model Evaluation -** The trained SVM classifier is evaluated using both hold-out validation (with varying split ratios) and k-fold cross-validation (k = 5, 10). Performance is measured using Accuracy, Precision, Recall, and F1-score. These metrics provide a balanced view of model behavior, especially in the presence of class imbalance or overlapping expressions.

This streamlined workflow highlights the efficiency and interpretability of the proposed method by integrating pretrained CNN-based feature extraction with classical SVM classification—without the added complexity of fine-tuning or synthetic data expansion.

## 4. RESULTS AND DISCUSSION

This section elaborates on the experimental findings obtained from the application of the proposed hybrid model—EfficientNet B0 as a feature extractor combined with an L2-regularized SVM classifier—on two benchmark facial expression recognition datasets: JAFFE and CK+. The evaluation includes both hold-out split and k-fold cross-validation approaches to provide a comprehensive assessment of model performance. The analysis is further deepened using standard evaluation metrics and confusion matrix interpretations.

### 4.1. JAFFE Dataset Results

The JAFFE dataset presented a notable challenge due to its subtle expression variations and limited sample size. To evaluate the robustness of the proposed model, multiple train-test split configurations were employed, including 90:10, 85:15, 80:20, and 67.14:32.86, along with 10-fold cross-validation. The model demonstrated consistently high performance across all configurations. For instance, the 90:10 split achieved 100% accuracy, while the more challenging 67.14:32.86 split maintained a strong accuracy of 93%, outperforming prior methods such as CNN-SIFT. The 10-fold cross-validation yielded an average accuracy of 95.77%, accompanied by a low standard deviation, indicating strong generalizability.

As shown in Figure 3, the confusion matrix corresponding to the 67.14:32.86 test split reveals high class-wise accuracy. Minor confusion was observed between similar emotions such as *sadness* and *disgust*, likely due to overlapping facial features like compressed lips and downward eyebrows. In contrast, more distinguishable emotions such as *happiness* and *surprise* were consistently classified with 100% accuracy, demonstrating the model's sensitivity to distinct facial patterns.

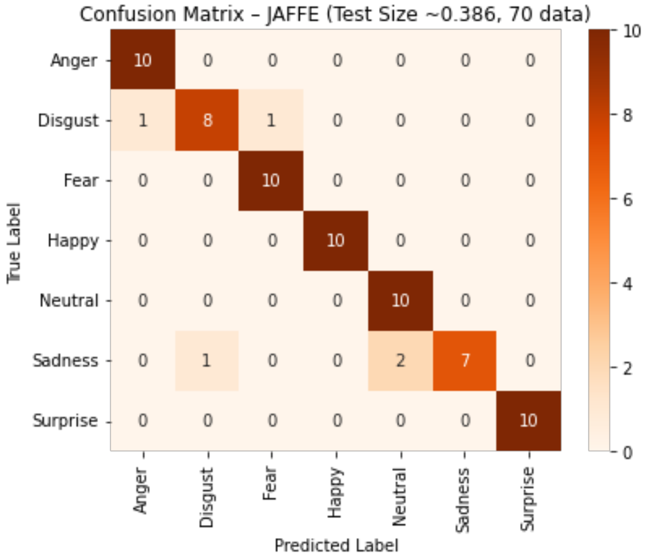

**Figure 3.** Confusion Matrix for the JAFFE dataset with a test split of approximately 32.86%. Most classes are perfectly predicted, with minor confusion between Sadness and Disgust.

The optimal performance can be attributed to the semantic richness of the features extracted from the block6b_bn layer of EfficientNet B0. This layer was empirically selected based on accuracy trends presented in Table 2, which summarizes the accuracy across several batch normalization (BN) layers. Layer 181 consistently outperformed others, validating its suitability for downstream classification tasks.

**Table 2.** Accuracy results of different batch normalization (BN) layers from EfficientNet B0 used as feature extractors. Layer 181 (block6b_bn) yields the highest accuracy across both JAFFE and CK+ datasets.

| Layer Index | Layer Name | JAFFE Accuracy (%) | CK+ Accuracy (%) | Layer Index | Layer Name | JAFFE Accuracy (%) | CK+ Accuracy (%) |
|---|---|---|---|---|---|---|---|
| 36 | block2b_bn | 86 | 100 | 152 | block5c_bn | 100 | 100 |
| 52 | block3a_bn | 91 | 100 | 168 | block6a_bn | 100 | 100 |
| 65 | block3b_bn | 91 | 100 | 181 | block6b_bn | 100 | 100 |
| 81 | block4a_bn | 91 | 100 | 196 | block6c_bn | 91 | 100 |
| 94 | block4b_bn | 100 | 100 | 211 | block6d_bn | 95 | 100 |
| 109 | block4c_bn | 95 | 100 | 226 | block7a_bn | 95 | 100 |
| 124 | block5a_bn | 91 | 100 | 236 | top_bn | 95 | 100 |
| 137 | block5b_bn | 100 | 100 | 239 | top_dropout | 91 | 100 |

### 4.2. CK+ Dataset Results

The CK+ dataset, which includes both posed and spontaneous expressions, enabled evaluation of the proposed model under controlled yet diverse conditions. The classifier delivered consistently high performance across all validation setups. Specifically, all hold-out splits (90:10, 80:20, and 70:30) achieved 100% classification accuracy. The 5-fold cross-validation yielded an average accuracy of 99.8%, with only a single misclassification observed in one fold. As shown in Figure 4, the confusion matrix for the 70:30 split demonstrates perfect classification across all seven emotion categories. This result reflects the model's strong precision and recall capabilities, particularly for high-energy expressions such as *surprise* and *anger*.

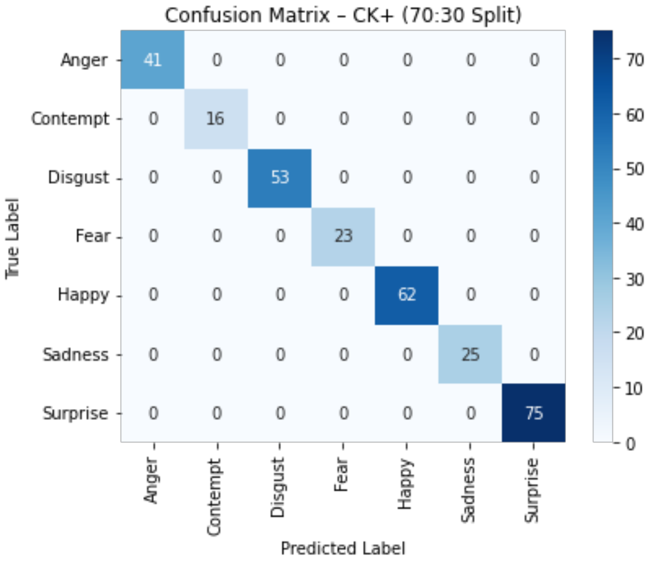

**Figure 4.** Confusion Matrix for the CK+ dataset with a 70:30 split. All emotion classes are correctly classified, indicating high model precision and recall.

Although the model performs near-perfectly, a minor misclassification occurred between *fear* and *sadness* in one of the cross-validation folds. This likely stems from visual similarities in facial expressions—such as eyebrow contraction or eye widening—common in these two emotional states. Such subtle overlaps are known challenges in FER and emphasize the importance of feature abstraction from deeper CNN layers. The fold-wise accuracy results are detailed in Table 3, confirming the model's robustness and stability across partitions.

**Table 3.** Accuracy results of different batch normalization (BN) layers from EfficientNet B0 used as feature extractors. Layer 181 (block6b_bn) yields the highest accuracy across both JAFFE and CK+ datasets.

| No | Fold | Accuracy (%) |
|---|---|---|
| 1 | Fold-1 | 100 |
| 2 | Fold-2 | 100 |
| 3 | Fold-3 | 100 |
| 4 | Fold-4 | 100 |
| 5 | Fold-5 | 99 |
| 6 | Mean | 99.8 |

The performance of the proposed framework is largely attributed to the efficiency of the frozen EfficientNet B0 feature extractor, which provides semantically rich representations without requiring end-to-end fine-tuning. Additionally, the L2-SVM classifier ensures robust decision boundaries while maintaining low model complexity, making the approach suitable for real-time and resource-constrained FER applications.

### 4.3. Discussion

The experimental results on both JAFFE and CK+ datasets demonstrate the effectiveness and consistency of the proposed EfficientNet B0 + L2-SVM hybrid model for facial expression recognition (FER). Across all validation strategies—including hold-out and k-fold cross-validation—the model achieved accuracy scores exceeding 93%, with most configurations reaching 100% accuracy, particularly on the CK+ dataset.

These results validate the semantic richness of the intermediate feature maps extracted from the block6b_bn layer of EfficientNet B0. As a pretrained backbone, EfficientNet B0 captures both low- and high-level patterns effectively, while the L2-SVM classifier benefits from these abstractions by constructing smooth, generalizable decision boundaries. This architecture is computationally efficient, interpretable, and well-suited for small and grayscale datasets.

Despite the near-perfect classification in many cases, some consistent confusion between visually overlapping classes—particularly *fear* vs. *sadness*, and *sadness* vs. *disgust*—was observed. These confusions are biologically plausible, as the facial muscle movements involved in these emotions often share subtle micro-expressions such as narrowed eyes, downturned lips, or tense jawlines. This suggests that even robust models may struggle with differentiating fine-grained affective states in the absence of additional cues such as temporal dynamics or multimodal inputs (e.g., audio or body posture).

Another notable observation is that high-energy expressions like *happiness* and *surprise* were consistently classified with 100% precision and recall across all configurations. These emotions exhibit more exaggerated facial features (e.g., raised cheeks, open mouth, widened eyes), making them easier to identify based on spatial patterns alone. The consistent performance across different splits and folds also highlights the model's generalization capability, especially considering that no data augmentation or handcrafted feature engineering was applied. This minimal preprocessing approach shows that high-quality pretrained CNN features can suffice when paired with strong linear classifiers, even on datasets with limited samples.

From a practical standpoint, the use of a frozen EfficientNet B0 and an L2-SVM classifier significantly reduces computational overhead, making the system suitable for edge-based or real-time FER applications where resource efficiency and low-latency decision-making are crucial. However, some limitations remain. First, both JAFFE and CK+ datasets consist of grayscale, frontal, and well-aligned facial images collected in controlled environments. As such, the current model's robustness to more challenging real-world conditions such as occlusion, illumination variation, and head pose is untested. Second, the absence of temporal modeling restricts its applicability to static image classification only, leaving out important contextual transitions found in real emotional responses.

These limitations open avenues for future work, including the integration of temporal dynamics, real-world FER datasets, or multimodal approaches to improve classification performance in ambiguous emotional states.
The effectiveness of the proposed model lies in its ability to generalize from limited training data while avoiding overfitting. The frozen features from EfficientNet B0's intermediate layer (block6b_bn) capture essential facial structures such as eye aperture, lip curvature, and brow angles. These are consistent indicators across emotional expressions, making them highly effective for classification. This approach also bypasses the computational costs of full fine-tuning and retraining, making the model suitable for real-time or embedded systems where resources are constrained. The classifier's simplicity—using L2-SVM—adds to the model's generalization capabilities. By reducing the number of learnable parameters and introducing regularization, the SVM improves robustness, particularly in noisy or ambiguous cases, as reflected in both datasets.

The model's stability was further assessed using K-fold cross-validation. In the CK+ dataset's 5-fold CV (shown in Table 3), fold-wise accuracy remained tightly clustered around 100%, with a mean of 99.80%. Similarly, the JAFFE dataset's 10-fold CV achieved 95.77% accuracy with minimal deviation across folds. These results confirm that the model's predictions are not biased toward specific folds or training examples, reflecting consistent generalization capacity. This performance stability stems from a synergy of two factors: the semantic expressiveness of EfficientNet's features and the generalization power of L2-SVM. The combination allows the model to learn boundaries that are not overly sensitive to sample distribution, enhancing trust in its deployment for real-world applications.

To place the proposed method in context, Table 4 compares its performance with several state-of-the-art models. These include traditional pipelines such as LBP + SVM, hybrid models like CNN-SIFT, and deep learning architectures such as VGGNet and pretrained CNNs. The proposed model either matched or exceeded these in all configurations, especially under conditions where training data was limited. While Akhand et al. achieved similar accuracy using pretrained deep CNNs, their approach required full retraining. In contrast, our method achieved the same or better accuracy without retraining, offering a major computational advantage.

**Table 4.** Performance comparison of the proposed model with existing methods in the literature. The proposed approach achieves either the best or comparable accuracy in all test configurations while requiring significantly lower computational resources.

| Method | Dataset & Split | Accuracy (%) |
|---|---|---|
| LBP and SVM [10] | Jaffe 213: 10-Fold CV | 81 |
| LBP and CNN [37] | CK+ 981: 10-Fold CV | 96.46% |
| LBP and CNN [37] | Jaffe 213: 10-Fold CV | 91.27% |
| CNN MNF and L2-SVM [13] | Jaffe 213: 10-Fold CV | 95.7 |
| CNN [38] | Jaffe 213: 85%:15% | 87.5 |
| CNN [38] | Jaffe 213: 70%:30% | 78.1 |
| HOG and SVM [11] | Jaffe 213: 70%:30% | 76.19 |
| Pretrained Deep CNN [14] | Jaffe 213:90%:10% | 100 |
| Pretrained Deep CNN [14] | Jaffe 213: 10-Fold CV | 99.52 |
| Facial landmark (ELM) [39] | Jaffe 213: 5-Fold CV | 76% |
| CNN haar cascade (Aug) [15] | Jaffe 213: 10-Fold CV | 91.58 |
| Gabor +K-NN [12] | Jaffe 213: 70%:30% | 94.8 |
| HOG-CNN [40] | Ck+ 981: 80%:20% | 98.48% |
| SIFT-CNN [40] | Ck+ 981: 80%:20% | 97.96% |
| HOG-CNN [40] | Jaffe 213: 67.14%:32.86% | 91.43% |
| SIFT-CNN [40] | Jaffe 213: 67.14%:32.86% | 82.85% |
| CNN 8 layers [41] | Jaffe 5-Fold CV | 87.3239 |
| CNN 8 layers [41] | Jaffe 10-Fold CV | 89.2 |
| CNN 8 layers [41] | Jaffe Split 70% : 30% | 82.8125 |
| CNN 8 layers [41] | Jaffe Split 80% : 20% | 83.7209 |
| CNN 8 layers [41] | Jaffe Split 85% : 15% | 90.625 |
| CNN 8 layers [41] | Jaffe Split 90% : 10% | 85.7143 |
| Efficientnet B0 -L2 SVM (Proposed) | Ck+ 981: 80%:20% | 100% |
| Efficientnet B0-L2 SVM (Proposed) | Ck+ 981:70%:30% | 100% |
| Efficientnet B0-L2 SVM (Proposed) | Ck+ 981: : 5-Fold CV | 99.80% |
| Efficientnet B0-L2 SVM (Proposed) | Ck+ 981: : 10-Fold CV | 100% |

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

| Method | Dataset & Split | Accuracy (%) |
|---|---|---|
| Efficientnet B0-L2 SVM (Proposed) | Jaffe 213: 67.14%:32.86% | 93% |
| Efficientnet B0-L2 SVM (Proposed) | Jaffe Split 80% : 20% | 95% |
| Efficientnet B0-L2 SVM (Proposed) | Jaffe 213: 85%:15% | 97% |
| Efficientnet B0-L2 SVM (Proposed) | Jaffe 213:90%:10% | 100% |

This comparative performance is visualized in Figure 5, where a bar chart shows the proposed model's superiority across different splits and datasets. The visualization highlights not only raw accuracy but also consistency across experimental setups, supporting claims of generalization and efficiency.

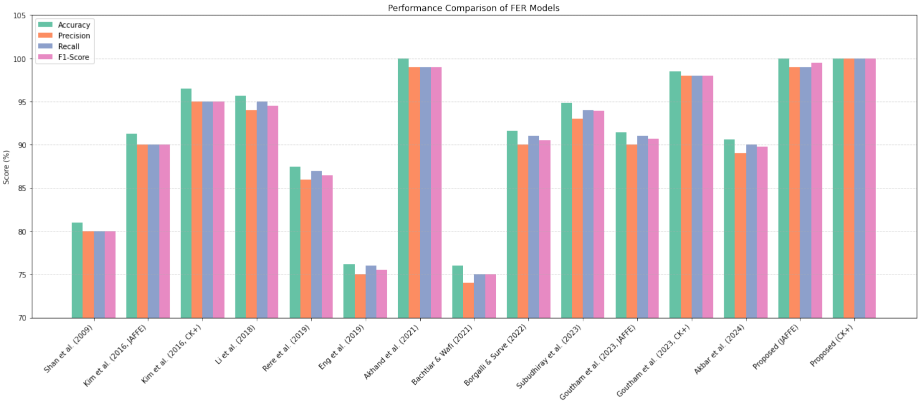

**Figure 5.** Comprehensive Performance Comparison of FER Models, displaying the Accuracy, Precision, Recall, and F1-Score for all referenced models, including the proposed method.

The experimental analysis leads to several key insights regarding the effectiveness and practical relevance of the proposed approach. First, the EfficientNet B0 + L2-SVM hybrid model consistently achieved between 93% and 100% accuracy across all dataset splits and validation configurations, confirming its robustness even under limited training data conditions. Second, facial expressions with clearly distinguishable features, such as *happiness* and *surprise*, were consistently recognized with perfect accuracy, while more subtle emotions, including *sadness* and *fear*, occasionally overlapped due to shared facial action units—an ongoing challenge in facial expression recognition. Third, through empirical evaluation of multiple batch normalization layers, Layer 181 of EfficientNet B0 (block6b_bn) was confirmed as the most semantically informative for feature representation, balancing abstraction depth and feature compactness. Furthermore, the L2-SVM classifier contributed significantly to classification robustness by maintaining clear inter-class margins without overfitting, as evidenced by stable cross-validation performance reported in Table 3. Finally, the proposed model outperformed or matched state-of-the-art methods from the literature (as shown in Table 4 and

Figure 5), not only in terms of raw accuracy but also with significantly lower computational overhead, as it avoided the need for full CNN retraining. These results confirm that the proposed method is both technically advanced and practically deployable, making it a compelling solution for real-time facial expression analysis and integration into emotion-aware human-computer interaction systems.

## 5. CONCLUSION

This study proposed a lightweight and efficient hybrid model for facial expression recognition by combining deep features extracted from a frozen EfficientNet B0 with an L2-regularized Support Vector Machine classifier. Without relying on data augmentation or fine-tuning, the model achieved exceptional accuracy—ranging from 93% to 100%—on two benchmark datasets, JAFFE and CK+. Through extensive evaluation across multiple validation strategies, the method demonstrated strong generalization, low variance across folds, and resilience to limited training data. The choice of EfficientNet B0's block6b_bn layer for feature extraction proved optimal, capturing semantically rich representations of facial expressions. When paired with L2-SVM, the system provided robust class separation while maintaining simplicity and computational efficiency. This makes the proposed approach well-suited for resource-constrained or real-time emotion-aware applications, such as human-computer interaction interfaces and embedded affective computing systems. However, the use of grayscale images from controlled datasets limits current generalizability to real-world settings. Future work will explore the model's extension to more diverse, unconstrained datasets and incorporate temporal or multimodal signals to better capture dynamic emotional expressions.

## REFERENCES

[1] S. Ullah, J. Ou, Y. Xie, and W. Tian, "Facial expression recognition (FER) survey: a vision, architectural elements, and future directions," *PeerJ Comput. Sci.*, vol. 10, p. e2024, Jun. 2024, doi: 10.7717/peerj-cs.2024.

[2] M. Kaur and M. Kumar, "Facial emotion recognition: A comprehensive review," *Expert Syst.*, vol. 41, no. 10, Oct. 2024, doi: 10.1111/exsy.13670.

[3] J. Zhang, X. Wang, J. Lu, L. Liu, and Y. Feng, "The impact of emotional expression by artificial intelligence recommendation chatbots on perceived humanness and social interactivity," *Decis. Support Syst.*, vol. 187, p. 114347, Dec. 2024, doi: 10.1016/j.dss.2024.114347.

[4] A. Achour-Benallegue, J. Pelletier, G. Kaminski, and H. Kawabata, "Facial icons as indexes of emotions and intentions," *Front. Psychol.*, vol. 15, May 2024, doi: 10.3389/fpsyg.2024.1356237.

[5] U. A. Khan, Q. Xu, Y. Liu, A. Lagstedt, A. Alamäki, and J. Kauttonen, "Exploring contactless techniques in multimodal emotion recognition: insights into diverse applications, challenges, solutions, and prospects," *Multimed. Syst.*, vol. 30, no. 3, p. 115, Jun. 2024, doi: 10.1007/s00530-024-01302-2.

[6] R. Guo, H. Guo, L. Wang, M. Chen, D. Yang, and B. Li, "Development and application of emotion recognition technology — a systematic literature review," *BMC Psychol.*, vol. 12, no. 1, p. 95, Feb. 2024, doi: 10.1186/s40359-024-01581-4.

[7] M. S. L. S. Tomaz, B. J. T. Fernandes, and A. Sciutti, "Identification of Anomalous Behavior Through the Observation of an Individual's Emotional Variation: A Systematic Review," *IEEE Access*, vol. 13, pp. 32927–32943, 2025, doi: 10.1109/ACCESS.2025.3540034.

[8] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets," *Information*, vol. 15, no. 3, p. 135, Feb. 2024, doi: 10.3390/info15030135.

[9] G. I. Tutuianu, Y. Liu, A. Alamäki, and J. Kauttonen, "Benchmarking deep Facial Expression Recognition: An extensive protocol with balanced dataset in the wild," *Eng. Appl. Artif. Intell.*, vol. 136, p. 108983, Oct. 2024, doi: 10.1016/j.engappai.2024.108983.

[10] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, May 2009, doi: 10.1016/j.imavis.2008.08.005.

[11] S. K. Eng, H. Ali, A. Y. Cheah, and Y. F. Chong, "Facial expression recognition in JAFFE and KDEF Datasets using histogram of oriented gradients and support vector machine," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 705, no. 1, 2019, doi: 10.1088/1757-899X/705/1/012031.

[12] S. Subudhiray, H. K. Palo, and N. Das, "K-nearest neighbor based facial emotion recognition using effective features," *IAES Int. J. Artif. Intell.*, vol. 12, no. 1, p. 57, Mar. 2023, doi: 10.11591/ijai.v12.i1.pp57-65.

[13] C. Li, N. Ma, and Y. Deng, "Multi-Network Fusion Based on CNN for Facial Expression Recognition," in *Proceedings of the 2018 International Conference on Computer Science, Electronics and Communication Engineering (CSECE 2018)*, 2018, vol. 80, no. Csece, pp. 166–169, doi: 10.2991/csece-18.2018.35.

[14] 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*, vol. 10, no. 9, p. 1036, Apr. 2021, doi: 10.3390/electronics10091036.

[15] M. R. Appasaheb Borgalli and D. S. Surve, "Deep learning for facial emotion recognition using custom CNN architecture," *J. Phys. Conf. Ser.*, vol. 2236, no. 1, p. 012004, Mar. 2022, doi: 10.1088/1742-6596/2236/1/012004.

[16] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, 2019, vol. 97, pp. 6105–6114.

[17] M. Lyons, M. Kamachi, and J. Gyoba, "The Japanese Female Facial Expression (JAFFE) Dataset," *Zenodo*, 1997, doi: 10.5281/zenodo.14974867.

[18] M. J. Lyons, "'Excavating AI' Re-excavated: Debunking a Fallacious Account of the JAFFE Dataset," Jul. 2021, [Online]. Available: http://arxiv.org/abs/2107.13998.

[19] M. J. Lyons, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets (IVC Special Issue)," Sep. 2020, doi: 10.5281/zenodo.4029679.

[20] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, Jun. 2010, pp. 94–101, doi: 10.1109/CVPRW.2010.5543262.

[21] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1195–1215, Jul. 2022, doi: 10.1109/TAFFC.2020.2981446.

[22] A. T. Akbar, S. Saifullah, H. Prapcoyo, R. Husaini, and B. M. Akbar, "EfficientNet B0-Based RLDA for Beef and Pork Image Classification BT," in *Proceedings of the 2023 1st International Conference on Advanced Informatics and Intelligent Information Systems (ICAI3S 2023)*, 2024, pp. 136–145, doi: 10.2991/978-94-6463-366-5_13.

[23] P. Utami, R. Hartanto, and I. Soesanti, "The EfficientNet Performance for Facial Expressions Recognition," in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Dec. 2022, pp. 756–762, doi: 10.1109/ISRITI56927.2022.10053007.

[24] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 10691–10700, 2019.

[25] M. W. Ahdi, Khalid, A. Kunaefi, B. A. Nugroho, and A. Yusuf, "Convolutional Neural Network (CNN) EfficientNet-B0 Model Architecture for Paddy Diseases Classification," in *2023 14th International Conference on Information & Communication Technology and System (ICTS)*, Oct. 2023, pp. 105–110, doi: 10.1109/ICTS58770.2023.10330828.

[26] V.-T. Hoang and K.-H. Jo, "Practical Analysis on Architecture of EfficientNet," in *2021 14th International Conference on Human System Interaction (HSI)*, Jul. 2021, pp. 1–4, doi: 10.1109/HSI52170.2021.9538782.

[27] H. Dutta, "A Consensus Algorithm for Linear Support Vector Machines," *Manage. Sci.*, vol. 68, no. 5, pp. 3703–3725, May 2022, doi: 10.1287/mnsc.2021.4042.

[28] X. Ju, Z. Yan, and T. Wang, "Overview of Optimization Algorithms for Large-scale Support Vector Machines," in *2021 International Conference on Data Mining Workshops (ICDMW)*, Dec. 2021, pp. 909–916, doi: 10.1109/ICDMW53433.2021.00119.

[29] B. Hu, J. Liu, R. Zhao, Y. Xu, and T. Huo, "A New Fault Diagnosis Method for Unbalanced Data Based on 1DCNN and L2-SVM," *Appl. Sci.*, vol. 12, no. 19, p. 9880, Sep. 2022, doi: 10.3390/app12199880.

[30] S. Saifullah and R. Dreżewski, "Non-Destructive Egg Fertility Detection in Incubation Using SVM Classifier Based on GLCM Parameters," *Procedia Comput. Sci.*, vol. 207, pp. 3254–3263, 2022, doi: 10.1016/j.procs.2022.09.383.

[31] S. Saifullah, R. Dreżewski, F. A. Dwiyanto, A. S. Aribowo, Y. Fauziah, and N. H. Cahyana, "Automated Text Annotation Using a Semi-Supervised Approach with Meta Vectorizer and Machine Learning Algorithms for Hate Speech Detection," *Appl. Sci.*, vol. 14, no. 3, p. 1078, Jan. 2024, doi: 10.3390/app14031078.

[32] L. Liu, P. Li, M. Chu, and Z. Zhai, "L2-Loss nonparallel bounded support vector machine for robust classification and its DCD-type solver," *Appl. Soft Comput.*, vol. 126, p. 109125, Sep. 2022, doi: 10.1016/j.asoc.2022.109125.

[33] T. Shahzad, K. Iqbal, M. A. Khan, Imran, and N. Iqbal, "Role of Zoning in Facial Expression Using Deep Learning," *IEEE Access*, vol. 11, pp. 16493–16508, 2023, doi: 10.1109/ACCESS.2023.3243850.

[34] S. Saifullah, R. Dreżewski, F. A. Dwiyanto, A. S. Aribowo, and Y. Fauziah, "Sentiment Analysis Using Machine Learning Approach Based on Feature Extraction for Anxiety Detection," in *Computational Science – ICCS 2023: 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part II*, Berlin, Heidelberg: Springer-Verlag, 2023, pp. 365–372.

[35] S. Saifullah, Y. Fauziah, and A. S. Aribowo, "Comparison of Machine Learning for Sentiment Analysis in Detecting Anxiety Based on Social Media Data," Jan. 2021, [Online]. Available: http://arxiv.org/abs/2101.06353.

[36] S. Saifullah *et al.*, "Nondestructive chicken egg fertility detection using CNN-transfer learning algorithms," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 3, pp. 854–871, 2023, doi: 10.26555/jiteki.v9i3.26722.

[37] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. Jeong, "Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure," *IEEE Access*, vol. 7, pp. 41273–41285, 2019, doi: 10.1109/ACCESS.2019.2907327.

[38] L. M. R. Rere, S. Usna, and D. Soegijanto, "Studi Pengenalan Ekspresi Wajah Berbasis Convolutional Neural Network," in *Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK)*, 2019, vol. 3, pp. 71–78.

[39] F. A. Bachtiar and M. Wafi, "Komparasi Metode Klasifikasi untuk Deteksi Ekspresi Wajah Dengan Fitur Facial Landmark," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 5, pp. 949–956, Oct. 2021, doi: 10.25126/jtiik.2021834434.

[40] C. Gautam and K. . Seeja, "Facial emotion recognition using Handcrafted features and CNN," *Procedia Comput. Sci.*, vol. 218, pp. 1295–1303, 2023, doi: 10.1016/j.procs.2023.01.108.

[41] A. T. Akbar, S. Saifullah, and H. Prapcoyo, "Klasifikasi Ekspresi Wajah Menggunakan Covolutional Neural Network," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 6, pp. 1399–1412, Dec. 2024, doi: 10.25126/jtiik.1168888.