

Impact of Image Quality Enhancement Using Homomorphic Filtering on the Performance of Deep Learning-Based Facial Emotion Recognition Systems

Al Bahri, Maulisa Oktiana, Maya Fitria, Zulfikar

Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Aceh, 23111, Indonesia

ARTICLE INFO

Article history:

Received December 10, 2024

Revised January 25, 2025

Accepted May 02, 2025

Keywords:

Facial Emotion Recognition;

Image Enhancement;

Deep learning;

MobileNet;

InceptionV3;

DenseNet121

ABSTRACT

Facial emotion recognition technology is crucial in understanding human expressions from images or videos by analyzing distinct facial features. A common challenge in this technology is accurately detecting a person's facial expression, which can be hindered by unclear facial lines, often due to poor lighting conditions. To address these challenges, it is essential to improve image quality. This study investigates how enhancing image quality through homomorphic filtering and sharpening techniques can boost the accuracy and performance of deep learning-based facial emotion recognition systems. Improved image quality allows the classification model to focus on relevant expression features better. Therefore, this research contributes to in facilitating more intuitive and responsive communications by enabling system to interpret and respond to human emotions effectively. The testing used three different architectures: MobileNet, InceptionV3, and DenseNet121. Evaluasi kinerja dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Experimental results indicated that image enhancement positively impacts the accuracy of the facial emotion recognition system. Specifically, the average accuracy increased by 1-2% for the MobileNet architecture, by 5-7% for InceptionV3, and by 1-3% for DenseNet121.

This work is licensed under a [Creative Commons Attribution-Share Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



Corresponding Author:

Maulisa Oktiana, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Aceh, CO 23111
Indonesia

Email: maulisaoktiana@usk.ac.id

1. INTRODUCTION

Expressions are the key physiological features that convey human emotions, allowing us to express feelings, thoughts, and emotions through body language, particularly through facial cues [1]-[2]. Common expression include happiness, anger, sadness, boredom, and surprise. Facial expression recognition is a system that identifies individual emotional expression by comparing it to the images stored in the database system. Facial expression recognition has been adopted across various fields, including in security, education, and mental health. In the healthcare sector, this system can assists medical professionals in monitoring mental health condition of the patients and serves as a supporting tools in diagnosis and therapy processes. In security sector, this technology is employed to detect and monitor the anomalous behavior in individuals. Meanwhile, in terms of human-computer interactions, it facilitates more intuitive and responsive communications by enabling system to interpret and respond to human emotions effectively [3]-[4]. Facial expression recognition system typically employs advanced image processing and machine learning techniques to extract and classify facial cues. However, a significant challenge in developing and implementing a facial emotion recognition system is achieving accurate recognition, especially when the quality of the images is poor and facial features

are unclear [5]-[6]. Additionally, the system often struggles to recognize expressions under varying conditions, highlighting the need for methods to enhance image quality [7]-[8].

Image enhancement is a crucial process for reducing noise levels, improving image contrast, balancing lighting, and clarifying edges that may be compromised due to poor image capture or numerous modifications. Overall, image enhancement facilitates deeper analysis of digital images, aids in object identification and isolation, and is essential in various image processing systems. Previous studies have investigated various techniques for enhancing image quality, such as wavelet transform, and Contrast Limited Adaptive Histogram Equalization (CLAHE) [9]-[10]. While these methods are effective in enhancing image contrast and highlighting feature, the methods often fail to preserve the important details of facial part, particularly in the image with extreme lighting conditions. Additionally, the methods do not consider how changes in the image quality impact the feature extraction and subsequently affect the accuracy of deep learning models, resulting in suboptimal performance. To address this challenges, this research proposes the implementation of homomorphic filtering technique combined with sharpening methods for image quality improvement. Homomorphic filtering enhances image sharpness by evenly distributing contrast intensity and sharpening edge details [11]-[12]. Homomorphic filtering has previously been utilized to normalize non-uniform lighting in cross-spectral and cross-distance face and iris recognition, demonstrating significant improvements in recognition accuracy [13]-[14]. This study focuses on the application of Homomorphic filtering for facial emotion recognition. Additionally, the image produced by Homomorphic filtering is enhanced through image sharpening, which highlights facial edges and details to make them clearer. However, the implementation this technique presents challenges due to its high computational complexity and the potential for over-enhancement on the image, which may interfere the feature extraction in deep learning model. Moreover, the method's sensitivity to lighting conditions and environmental variations can also pose difficulties.

The main contribution of this paper lies in improving the accuracy of facial emotion recognition by integrating of image quality improvement techniques, specifically homomorphic filtering, image sharpening, and deep learning. Experimental results indicate that utilizing these image enhancement methods can significantly improve the performance of facial emotion recognition. Thus, this research pioneers the exploration of image quality improvement techniques aimed at enhancing the accuracy of deep learning-based facial emotion recognition. In addition, this research integrates the homomorphic filtering methods with three different deep learning architectures, namely MobileNet, InceptionV3, and DenseNet121, to evaluate computational complexity of the proposed methods and their impact during model development process. This study aims to address the following gaps:

1. Developing image enhancement methods that not only improve the quality of the images but also positively impact accuracy in facial emotion classification.
2. Evaluating the effectiveness of the homomorphic filtering methods and the image sharpening in enhancing the performance of various deep learning architectures.
3. Analyzing the performance of homomorphic filtering methods in comparison to other image enhancement methods.

2. RELATED WORKS

Several related studies have been conducted in the field of Facial Emotion Recognition. Lu et al. used Classical VGGNet and Improved VGGNet models for emotion recognition from the FER2013 dataset. The study results showed that the Classical VGGNet achieved an accuracy of 84.31%, and the Improved VGGNet achieved an accuracy of 84.74% on 35,886 samples, divided into training, testing, and validation data. The limitation of this study is that the dataset only includes grayscale images, and the accuracy improvement between the Classical and Improved models is not significant [15].

Next, study [16] proposed a homomorphic filtering algorithm by utilizing the conventional Canny operator to extract facial edges and highlight important facial features, resulting in more accurate facial feature extraction. The findings showed a 2% increase in accuracy for facial recognition compared to traditional methods on the extended Yale B database. However, the limitation of this study is its focus solely on facial recognition without discussing the algorithm's performance on larger-scale datasets.

Furthermore, study [17] compared multi-scale Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) techniques with homomorphic filtering using a Fuzzy K Nearest Neighbor classifier. The results showed that the DCT classifier achieved an accuracy of 86.5%, and after combining, the accuracy increased to 90%. Meanwhile, the DWT classifier achieved an accuracy of 87.5%, and after combining, the accuracy increased to 89.5%. The computation time in this study was relatively long due to the multi-scale combination load, especially during the pre-processing and classification stages. This is because the use of DWT or DCT for image decomposition and coefficient selection is computationally intensive, leading to a

relatively long process. The limitation of this study is the use of the ORL Face Database, which contains only 400 face images, thus limiting the generalization of the results to a wider range of facial variations. Furthermore, study [18] focused on improving RGB image quality with a combination of Homomorphic Filtering and Butterworth Lowpass Filtering methods, which successfully produced better-quality images by eliminating noise such as salt and pepper. However, the limitation of this study is that it used only one small image in the dataset, which does not demonstrate the effectiveness of the method on high-resolution images, and no image quality metrics such as PSNR and SSIM were provided. Next, a study on a facial recognition system using the Xception architecture [19]. In this study, the FER-13 dataset with 48x48 pixel grayscale images was classified using the Xception architecture with Depthwise Separable Convolutions for efficiency, resulting in an accuracy of 66%. The limitation of this study is that the classification of the "disgust" class was very poor, with a recall of only 5%, due to the very small amount of data, and the system is highly dependent on lighting quality, camera angle, and the distance of the face from the camera, which can affect or even reduce the accuracy.

Lastly, a study on emotion classification using CNN, focusing on technical implementation such as batch size and epoch settings [20]. In this study, the dataset used was from Kaggle, similar to the FER-13 dataset, with grayscale images of 48x48 pixels. The testing results showed that with a combination of batch size 128 and 100 epochs, a training accuracy of 90% and validation accuracy of 65% were achieved. Out of 35 expressions, 28 were successfully classified with a validation accuracy of 80%. The limitation of this study is overfitting, as the difference between training accuracy and validation accuracy is too large, indicating that the model lacks generalization with new data. Additionally, the results showed that the model is highly dependent on the hyperparameter configuration and lacks exploration of learning rate combinations to produce more diverse accuracy data.

3. METHOD

The proposed framework is shown in Fig. 1, which involves four stages: 1) Dataset Input; 2) Image Enhancement using Homomorphic filtering and image sharpening; 3) Model Testing; and 4) Result Analysis.

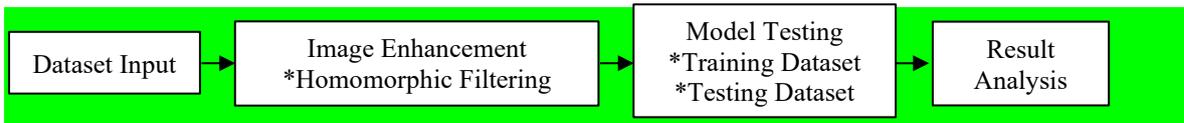


Fig. 1. Proposed Framework

3.1. Dataset Input

In this study, experimental was conduct using USK-FEMO dataset, which contains 1260 images consisting of 5 types of expressions: 255 images of happiness, 255 images of anger, 240 images of boredom, 255 images of sadness, and 255 images of surprise. The dataset, which is 8.25 GB in size, contains expression images divided into two parts: training data and testing data, which will serve as reference data for the system development [21]. Table 1 describes the detailed number of images for each expression class used in the study. The dataset is divided into training data, consisting of 1008 images or 80% of the total images, with 201 images per class, and testing data, consisting of 252 images or 20% of the total images, with 50 images per class.

Table 1. Number of Expression Images

No	Facial Expressions	Number Images	Image Pixel Size	Image Type
1	Happy	255	224×224	JPG
2	Bored	240	224×224	JPG
3	Angry	255	224×224	JPG
4	Sad	255	224×224	JPG
5	Surprised	255	224×224	JPG
	Total	1260	224 pi×224 pi	JPG
	Training Data	1008	201/class expression	
	Testing Data	252	50/class expression	

3.2. Image Enhancement

The image enhancement process consists of Homomorphic filtering and sharpening in order to boost the lighting and enhance details in the images. The Homomorphic filtering performed by importing the Homomorphic Filtering module function. The Homomorphic filtering computations are as follows [22]-[26]:

1. The normal image dataset is imported, and the RGB channels are separated because each image channel will apply Homomorphic filtering separately as illustrated in Fig. 2. The RGB channel separation is performed to enable detailed analysis and processing of each color component individually. Each channel is processed to distinguish between high-frequency (the detail) and low-frequency (the lighting) specifically. In the RGB domain, digital images, with each channel carries information about lighting intensities at specific wavelengths. This separation obtains a detailed analysis of unique characteristic of each channel. The red channel often exhibits dominant lighting intensity, while the blue channel is more prone to noise. The separation process may enable the implementation of filtering algorithm specifically in the channel requiring enhancement without intervene the information in other colors. This approach allows a better flexibility, as the parameter adjustment is done independently in each color to obtain optimal contrast and sharpness in image. Once processing is complete, the improved channel are recombined to form an image with significantly enhanced visual quality compared to original image.

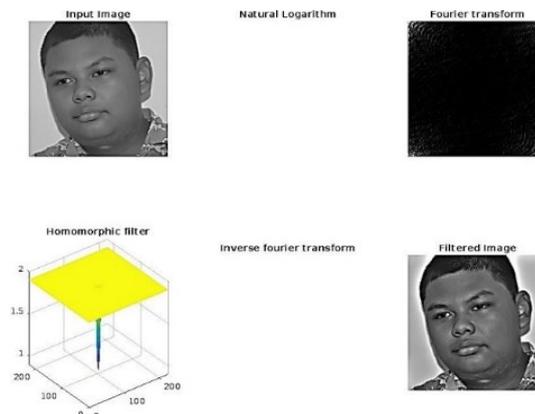


Fig. 2. Visualization of Homomorphic Filtering Application

2. Homomorphic filtering is applied to each image channel by converting each channel to the logarithmic domain to equalize intensity differences in the image. Then, the image is transformed from the spatial domain to the frequency domain using Fast Fourier Transform (FFT) to separate low frequencies (lighting) from high frequencies (details) in the image. Next, a Gaussian mask is applied to the image in the frequency domain by multiplying the FFT result with the Gaussian mask, with parameters low=1.0, high=1.3, and d0 (cutoff frequency)=30. The parameter selection for this technique is carefully conducted to achieve primary objective: balancing the image intensities and enhancing details without producing artifacts or excessive noise. The selection of low, high, and d0 value is performed empirically through the experiments on representative dataset, aiming to enhance image sharpness while avoiding distortion. This parameters are also evaluated to assess their impact on deep learning model accuracy of facial emotion recognition, ensuring performance improvements across various lighting conditions. The low value is set to 1.0 to preserve low-frequency contribution without amplification or attenuation. This ensures the lighting section in the image remains proportional without removing essential information from original lighting. The high value is set to 1.3 to amplify high-frequency components (the image details) aiming to increase the sharpness and image details proportionally without causing distortion. The value of 1.3 is selected as a balance compromise to clarify the details while preventing over-enhancement, which could lead to noise or artifacts. Meanwhile, the Cutoff frequency determines the boundary where the separation between high- and low-frequency occurs. A value of 30 is chosen for d0 based on the average size of the important features in the image, such as face expression, whose the frequency details typically between medium to high. By using the Cutoff value, the low-frequency components related to global lighting are reduced, while preserving the details related to face expression. This choice ensures that minor noise and very fine details are not excessively amplified, maintaining a balance between enhancement and noise suppression. The process of Homomorphic filtering are calculated using the formula of the Image Multiplication Model, Logarithmic Transformation, Fourier Transformation, and Gaussian Mask process as shown in (1), (2), (3), and (4), respectively. The computation of Homomorphic filtering process begin with the calculation of Image Multiplication model, which presents as in (1).

$$F(x, y) = L(x, y) \cdot R(x, y) \quad (1)$$

where $F(x, y)$ is the image intensities of original image in the (x, y) coordinate, $L(x, y)$ represents lighting level in the image. $R(x, y)$ is reflectance component as intrinsic properties of the reflected object. Subsequently, the logarithmic transformation in (2) is calculated by using $L(x, y)$ and $R(x, y)$.

$$\ln F(x, y) = \ln(L(x, y)) + \ln(R(x, y)) \quad (2)$$

where \ln represents natural logarithmic. This logarithmic transformation can ease the filtering process as the lighting and reflectance components are separated additively. In addition, image is transformed performing Fourier Transformation as in (2).

$$G(u, v) = F \{ \ln F(x, y) \} \quad (3)$$

where $G(u, v)$ represents the image in domain frequency, and $F \{ \ln F(x, y) \}$ is Fourier Transformation of logarithmic image $\ln F(x, y)$. The coordinate in frequency domain is described as u, v . Frequency domain aims to allow the manipulation in high and low-frequency.

$$1 - e^{-\frac{(u^2+v^2)}{d_0^2}} \quad (4)$$

where $H(u, v)$ is filter functions that modify intensity of image frequency based on the coordinate u and v . The low is minimum low-frequency intensity, while the high represents the number of high-frequency to be amplified. The calculation of $u^2 \pm v^2$ describes the distance of a certain frequency to the origin $(0,0)$.

3. After applying the mask, the image is returned from the frequency domain to the spatial domain using Inverse Fast Fourier Transform as in (5). Finally, the logarithmic transformation applied earlier is reversed, so the image pixel intensities return to a linear scale by using equation (6). The homomorphic filter process as illustrated in Fig. 3.

$$\ln F'(x, y) = F^{-1} \{ Z(u, v) \} \quad (5)$$

$$F'(x, y) = e^{\ln F'(x, y)} \quad (6)$$

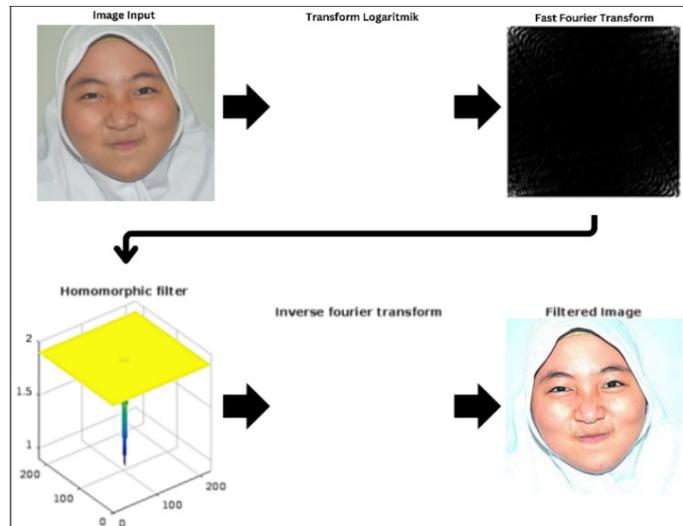


Fig. 3. Homomorphic Filtering results

4. The filtered image channels are then combined. At this stage, the filtered image is normalized to ensure that pixel intensity values fall within the visualizable range, i.e., between 0 and 255. Following the Homomorphic filtering process, image sharpening is performed to enhance edge details and improve the overall sharpness in the image. The image sharpening technique employed in this research is Convolutional Kernel Sharpening, utilizing 3×3 of kernel size. The technique is subset of spatial domain filtering, aiming to enhance image sharpness by strengthening image details and edges. The 3×3 kernel is a small matrix applied to the image through a convolutional operation, where the intensity value of pixels in a certain area

are multiplied by kernel value and summed to produce sharper pixel. This approach is particularly effective because the Homomorphic filtering repairs the lighting and image contrast by separating the illumination and reflectance components. While this process evens out lighting, it can sometimes reduce some fine details. Thus, the 3×3 kernel is essential to restore and enhance the fine details, particularly in the edges and facial textures. The combination of the Homomorphic filtering and image sharpening produce images with balanced optimal lighting, and sharp and clear detail [27]-[28]. Fig. 4 illustrates the Image Enhancement Results obtained through this combined approach. The image results are utilized as the input in classification process, providing a strong foundation for improved accuracy in emotion facial recognition.

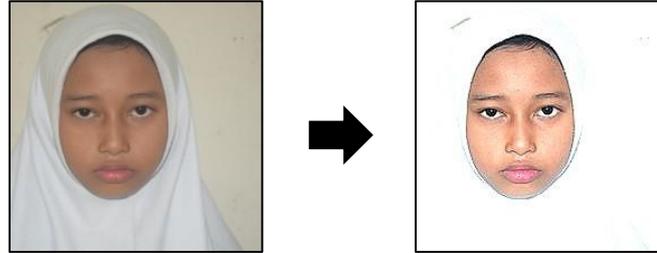


Fig. 4. Image Enhancement Results

- Image quality assessment was performed by comparing the normal image and the filtered image using Peak Signal-to-Noise Ratio (PSNR) to measure the signal-to-noise ratio, SSIM to evaluate perceptual changes in structural information, and MSE to calculate the mean squared error between the original pixel values and the filtered pixel values. The measurement results indicated that the filtered image exhibited better quality compared to the image without quality enhancement [29]. Equations (7), (8), and (9) are used to calculate PSNR, SSIM, and MSE, accordingly. Table 2 demonstrates the quality measurement values of several sampe image using PSNR, SSIM, and MSE. PSNR quantifies the ratio between the maximum possible pixel value and the noise or error affecting the image. Higher PSNR values indicate better image quality (closer to the original). PSNR values above 30 dB are typically considered good and low noise for image quality. while <20 dB indicates poor quality, high noise, and $20-30$ dB indicates moderate quality. SSIM is a metric designed to evaluate image quality based on human visual perception. It measures the structural similarity between the original and reconstructed images by considering luminance, contrast, and structure. SSIM values range from 0 to 1: 0 where 0 represents no similarity between the images and 1 represents perfect similarity (images are identical). SSIM is more sensitive to structural and perceptual changes than MSE or PSNR, making it better for evaluating visual quality. MSE is a fundamental metric that measures the average squared difference between the pixel values of the original and reconstructed images. It represents the reconstruction error. Smaller MSE values indicate better image quality (closer to the original). However, MSE has limitations as it does not account for structural or perceptual aspects of the image, making it less aligned with human perception. The table demonstrates that the image quality improves significantly, with a PSNR value exceeding 25 dB, indicating a high signal-to-noise ratio and minimal distortion. Additionally, the average SSIM value, which is approximately 0.9 and close to 1.0, reflects a strong structural similarity between the original and filtered images. Furthermore, the MSE values, which are nearly zero, confirm that the filtered images retain excellent quality. Based on these three measurement parameters, the filtered images exhibit superior quality, which contributes to enhancing the performance of the model in classifying facial expressions accurately.

$$PSNR = 10. \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (7)$$

where MAX is the maximum possible pixel value in the image (e.g. 255 for 8-bit images), and MSE is Mean Squared Error between the original and reconstructed images.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (8)$$

where μ_x, μ_y is mean pixel intensities of images x and y , $\sigma_x^2 + \sigma_y^2$ is variances of pixel intensities in x and y , σ_{xy} is covariance between x and y , and C_1, C_2 represents small constants to prevent division by zero.

$$MSE = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N [I_{original}(i, j) - I_{reconstructed}(i, j)]^2 \quad (9)$$

where M, N is dimensions of the image, and $I_{original}$ and $I_{reconstructed}$ represents pixel values of the original image and reconstructed image, respectively. Quality measurement values of several images shown in Table 2.

Table 2. Quality Measurement Values of Several Images

Original Image	Filtered Image	Quality Measurement Parameters		
		PSNR	SSIM	MSE
		28.79 dB	0.9806	0.0013
		31.08 dB	0.9858	0.0008

To determine the most effective image enhancement in this research, several evaluations are performed employing CLAHE, Homomorphic filtering, and Wavelet Transform, with the results presented in Table 3. The enhanced images using CLAHE and Wavelet Transform tend to appear darker and contain excessive noise, which negatively impact the accuracy. This is primarily due to suboptimal parameter settings, leading to uneven contrast enhancement and amplified noise. Moreover, in the Wavelet Transform method, artifacts were observed during the decomposition and reconstruction process, particularly when the decomposition level or wavelet basis did not align with characteristics of image. In contrast, Homomorphic filtering method obtained better results in improving image contrast without causing excessive noise, showing its ability to separate high- and low-frequency components to enhance image details while maintaining overall quality of image.

Table 3. Enhancement technique using Homomorphic Filtering, CLAHE, and Wavelet Transform

Image Type	Testing Model	Learning Rate	Accuracy	
Original Image		MobileNet	10-2	80.63%
Image with Homomorphic Filtering		MobileNet	10-2	81.74%
Image with CLAHE		MobileNet	10-2	72.61%
Image with Wavelet Transform		MobileNet	10-2	76.98%

3.3. Model Testing and Training

The experimental was conducted on the Google Colab platform, using training and testing data on three expression classification model architectures: MobileNet, DenseNet121, and InceptionV3. The selection of MobileNet as the architecture in this research was considered because it is a lightweight architecture optimized for devices with limited resources, using depthwise separable convolutions to efficiently process input images and provide high performance for mobile and embedded vision applications. InceptionV3 emphasizes computational efficiency with fewer parameters, combining various convolutional operations to extract spatial and contextual features across multiple scales while preventing overfitting through reduction and regularization

techniques. DenseNet121 maximizes feature reuse by connecting each layer to all preceding layers, enhancing feature utilization without increasing model complexity, making it highly effective for extracting detailed features from input images.

The hyperparameters used included 100 epochs, image size of $224 \times 224 \times 3$, the Adam optimizer, a batch size of 32, and three learning rate variations: 0.01 (10^{-2}), 0.0001 (10^{-4}), and 0.000001 (10^{-6}) [29], [30]. An epoch represents a complete iteration of training that covers the entire dataset, while the learning rate is a parameter that controls the speed at which the model learns the problem during training. A batch size of 32 divides the dataset into smaller subsets, with each batch containing 32 iterations. The loss function used is cross-entropy, commonly applied in classification models. Pre-trained weights initialized with ImageNet serve as a starting point for transfer learning, leveraging existing knowledge. Frozen layers in each expression classification system refer to various layers within the model whose parameters are not modified or updated during training, as they are considered sufficiently adept at recognizing general features in images, such as edges, shapes, and textures, learned from the large ImageNet dataset [11], [14], [31], [32], [33]. Frozen layer is a layer in neural network model whose the parameters (weights and biases) remain unchanged during the training process. In transfer learning, the pre-trained models like ImageNet often freeze the initial layers, as they capture basic features like edges and textures common across images. This reduces computation, improves stability, and allows retaining only the later layers to adapt to specific tasks, such as facial expression detection.

In the MobileNet model, the first 15 layers are frozen, while DenseNet121 also freezes several initial layers. In InceptionV3, 249 layers are frozen, covering most of the early and middle layers responsible for detecting features within images. Categorical Crossentropy is used as the loss function to measure the difference between the predicted probability distribution and the expected target distribution [34]-[35]. Testing was conducted using both the filtered expression dataset and the normal expression dataset. The accuracy results of the two datasets were then compared to evaluate the impact of image quality enhancement.

3.4. Evaluation Methods

Performance were analyzed to determine the effectiveness of using Homomorphic Filtering in enhancing expression classification performance of deep learning-based expression classification systems. The model's performance were analyzed using matrix of accuracy, precision, recall, F1-score, and multiclass confusion matrix. Accuracy measures the proportion of correct predictions out of all predictions made by the model. However, it can be misleading with imbalanced datasets, as the model may achieve high accuracy by predicting only the majority class correctly. Precision measures how many positive predictions are actually correct, while recall measures the model's ability to detect all positive samples. F1-score, the harmonic mean of precision and recall, balances both metrics and is particularly useful for imbalanced datasets. It is crucial in tasks like pattern recognition or facial expression classification, where both precision and recall are equally important. A high F1-score indicates a good balance between accurate detection and minimal errors. The calculation of accuracy, precision, recall, and F1-Score are exhibited in equations (10)-(13).

$$Accuracy = \frac{All\ predicted\ image}{Total\ Data} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

4. RESULTS AND DISCUSSION

4.1. Results

Fig. 5 illustrates the training results of MobileNet with the normal and enhanced dataset. In normal dataset, accuracy consistently increases for both training and validation data over 100 epochs, stabilizing after approximately 20 epochs. The graph shows a significant decrease in loss during the early stages of training, eventually stabilizing. However, there are higher fluctuations in accuracy and loss for the validation data compared to the training data, indicating the potential for overfitting. These results demonstrate that MobileNet achieves good performance, but further improvements are needed to address the fluctuations in validation.

Besides, the training results of the MobileNet model using the enhanced dataset indicates the model's performance increases rapidly at the beginning of training and stabilizes after approximately 20 epochs. The validation accuracy also improves, although it exhibits higher fluctuations compared to the training data. In the loss graph, the loss value decreases sharply during the early stages of training and stabilizes after 20 epochs; however, the loss on the validation data remains higher and more fluctuating than on the training data. Overall, the application of the filtered dataset appears to enhance the model's accuracy, but the fluctuations in the validation data suggest that further improvements may be needed to address issues of overfitting or data imbalance in validation. The blue curve in the accuracy graph represents training accuracy, showing that MobileNet architecture achieved accuracy of 98% with the normal dataset, and 99% with the enhanced dataset. Meanwhile, the orange curve describes validation accuracy, indicating 86.5% accuracy for the normal dataset, and 87.69% for the enhanced dataset. In the loss graph, it can be seen that the blue curve shows training loss, where the model recorded 5% loss with the normal dataset and 4% with the enhanced dataset. The orange curves represents validation loss, with values of 58% for the normal dataset, and 48% for the improved dataset.

Fig. 6 illustrates the training results of InceptionV3 with the normal and enhanced dataset. In the normal dataset, the training accuracy increases rapidly at the beginning and stabilizes after approximately 20 epochs, while the validation accuracy is more fluctuating, albeit showing improvement. In the enhanced dataset, the training loss decreases sharply in the early stages of training and stabilizes after several epochs, whereas the validation loss remains more fluctuating and consistently higher than the training loss. These results indicate that the application of the filtered dataset helps improve the accuracy of the InceptionV3 model. However, the fluctuations in validation data suggest that there is room for further improvement, particularly in addressing performance instability in validation data and potential overfitting.

Fig. 7 illustrates the training results of the DenseNet121 model using the normal and enhanced dataset. In normal dataset, the training and validation accuracy gradually increase as the number of epochs increases, although the validation accuracy is slightly more fluctuating compared to the training data. This indicates that the model is able to learn well; however, there is some instability in the predictions on the validation data. In the loss graph, both the training and validation data experience significant decreases in loss during the training process, with the training loss consistently being lower than the validation loss. This indicates that although the model demonstrates reasonably good performance overall, the fluctuations in the validation data suggest potential overfitting or instability that needs to be addressed further to improve its performance on unseen data. Otherwise, in enhanced dataset, the training data accuracy consistently increases as the epoch progresses, while the validation data accuracy shows significant fluctuations, particularly in the early training phases, before reaching stability. The loss graph demonstrates a significant decrease in both datasets, with the training loss being lower compared to the validation data. Although the enhanced dataset helps improve the model's performance on the training data, the fluctuations observed in the validation data indicate that the model still faces challenges in maintaining stability and generalization, which may require further adjustments to reduce the risk of overfitting.

Fig. 8 shows the Confusion Matrix for the MobileNet, InceptionV3, and DenseNet121 model. Each matrix represents the distribution of the model's predictions compared to the actual labels of the test data. MobileNet model performs reasonably well in classifying certain classes, as indicated by the high number of correct predictions along the diagonal of the matrix (high values at positions corresponding to True Labels and Predicted Labels). However, some misclassifications are evident from the off-diagonal predictions, indicating that certain classes are still confused by the model. Overall, the Confusion Matrix provides a detailed view of the model's strengths and weaknesses in recognizing and classifying expressions from the dataset, helping to identify areas for improvement, particularly for classes with higher error rates. In the InceptionV3 model successfully recognizes some expressions well. However, several misclassifications are evident from the numbers outside the diagonal, showing that certain expressions are still incorrectly identified. While the model performs adequately for some classes, the higher misclassification rates in certain classes suggest the need for further improvements, either in model adjustments or in the quality of the training data. Overall, this Confusion Matrix helps pinpoint specific areas where the model's performance can be enhanced. The DenseNet121 model in normal dataset, the model demonstrates the ability to classify some expressions well, particularly in certain classes such as "happy" and "surprised," which have a higher number of correct predictions along the diagonal of the matrix. However, some misclassifications are observed in other classes, such as "angry" and "bored," indicating that the model struggles to distinguish between certain expressions. Otherwise, in enhanced dataset, the model's performance appears slightly more consistent, showing improved classification accuracy in some classes like "happy" and "sad." Nevertheless, misclassifications are still evident in several other classes, albeit with fewer errors compared to the normal dataset. Overall, using the enhanced dataset seems to have a positive

impact on the model's ability to classify expressions, but there remain areas that require improvement to achieve better accuracy across all classes.

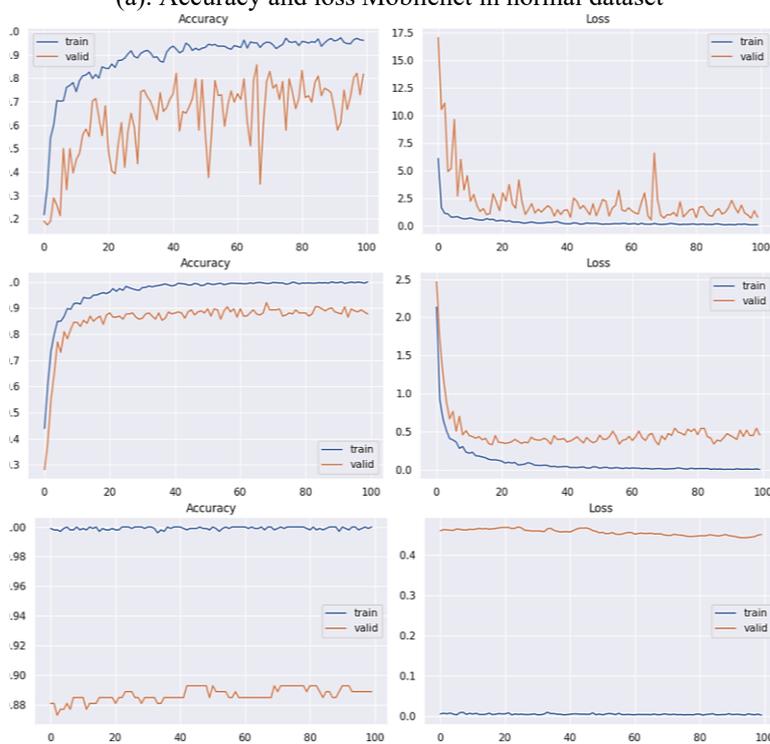
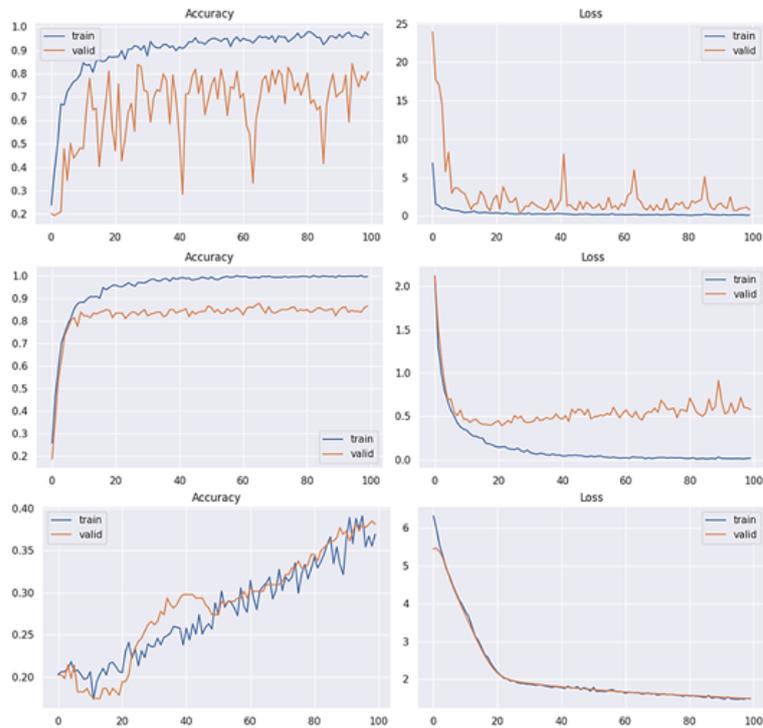
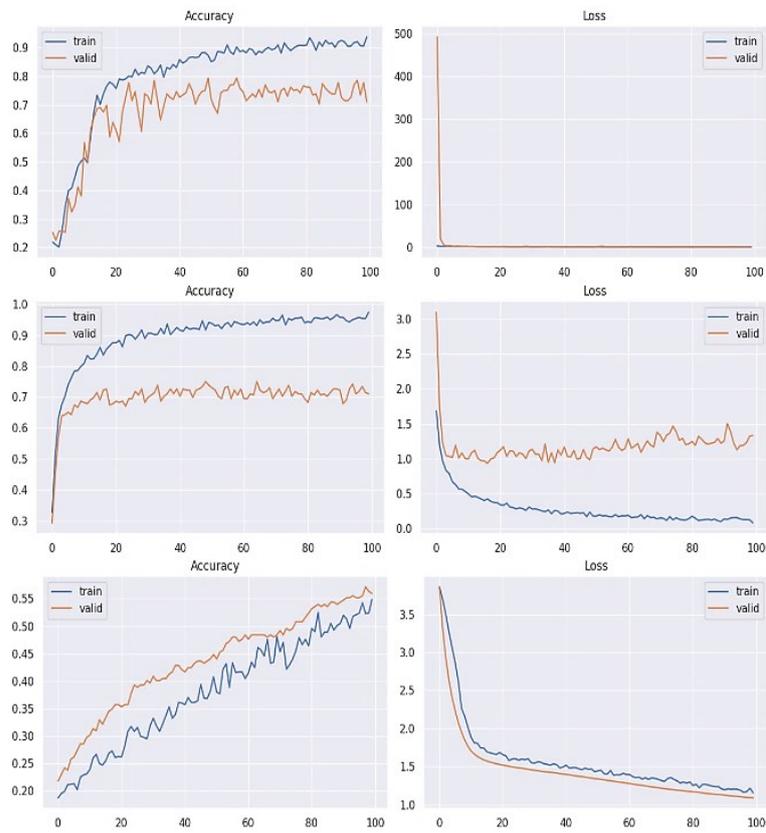
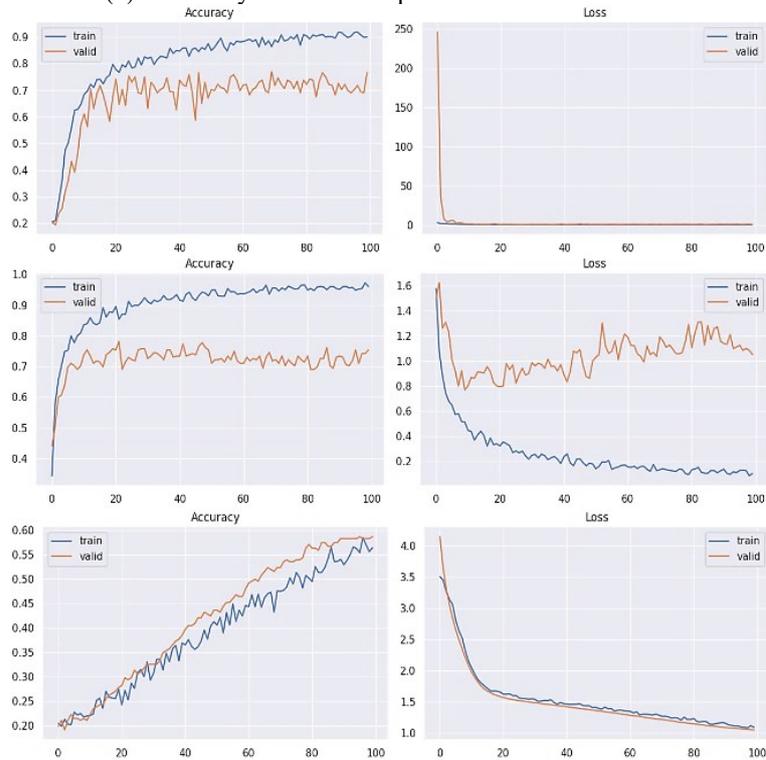


Fig. 5. The training results of MobileNet with the normal (a) and enhanced dataset (b)

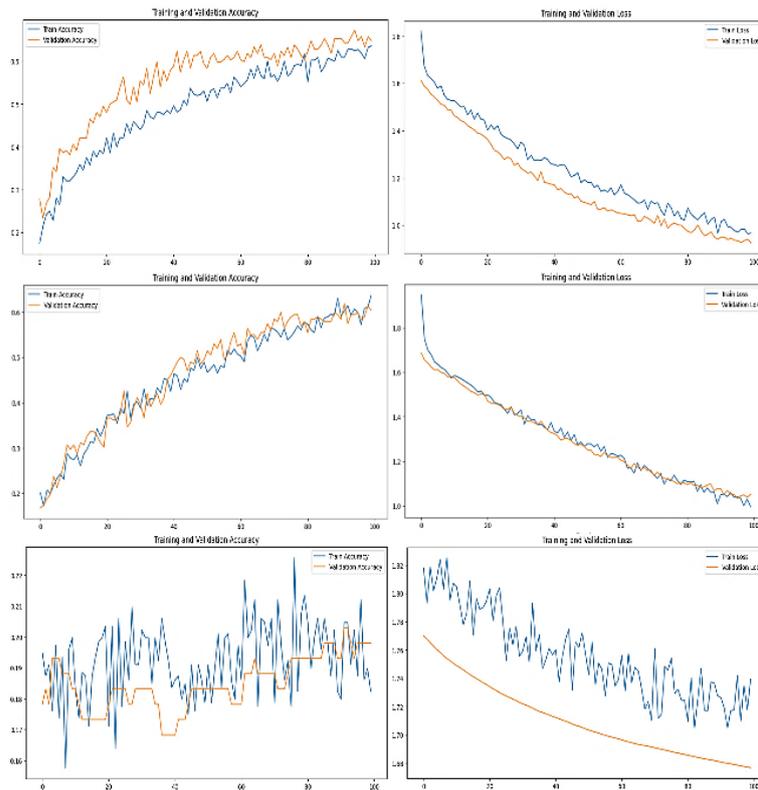


(a). Accuracy and loss InceptionV3 in normal dataset

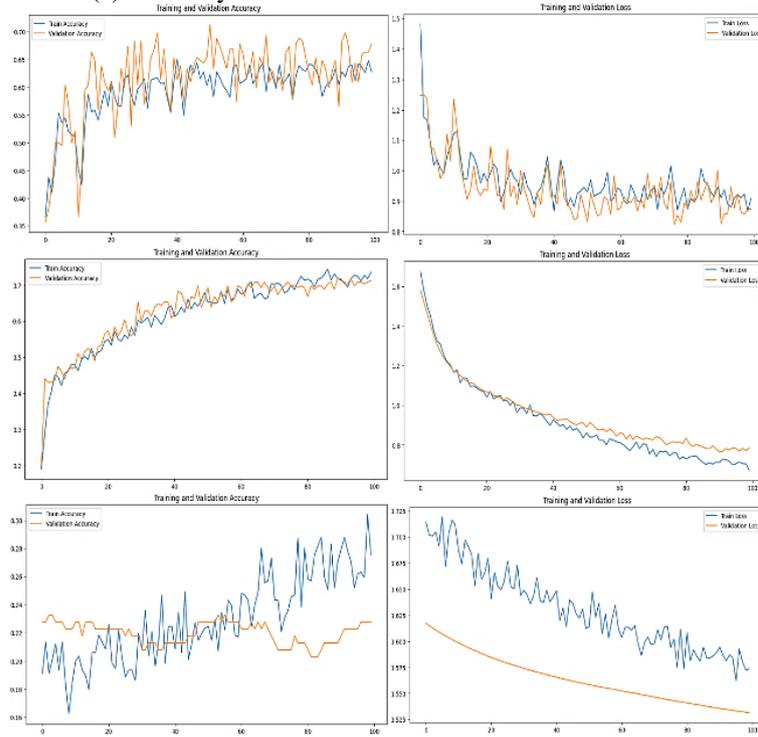


(b) Accuracy and loss InceptionV3 in enhanced dataset

Fig. 6. The training results of InceptionV3 with the normal and enhanced dataset



(a) Accuracy and loss DenseNet121 with normal dataset



(b) Accuracy and loss DenseNet121 with enhanced dataset

Fig. 7. The training results of DenseNet121 with the normal and enhanced dataset

Three matrices shown that most of the failure in classification occurs in “sad” and “bored” expressions, as both of them often share similar facial features. However, Homomorphic filtering contributed in reducing this by enhancing contrast and edges in facial lines. The graphic of accuracy and validation loss in reveals

fluctuations and over-fitting in the three models, particularly in MobileNet and DenseNet121, primarily due to relatively small dataset. This issue can be addressed by employing data augmentation technique, such as rotation, cropping, and color adjustment, to expand the number of dataset. Additionally, the model's sensitivity to enhanced features requires careful hyperparameter tuning, as parameters like low, high, and d0 have a significant impact on the results. Further testing is necessary to optimize values like learning rate and dropout to minimize overfitting across different datasets.

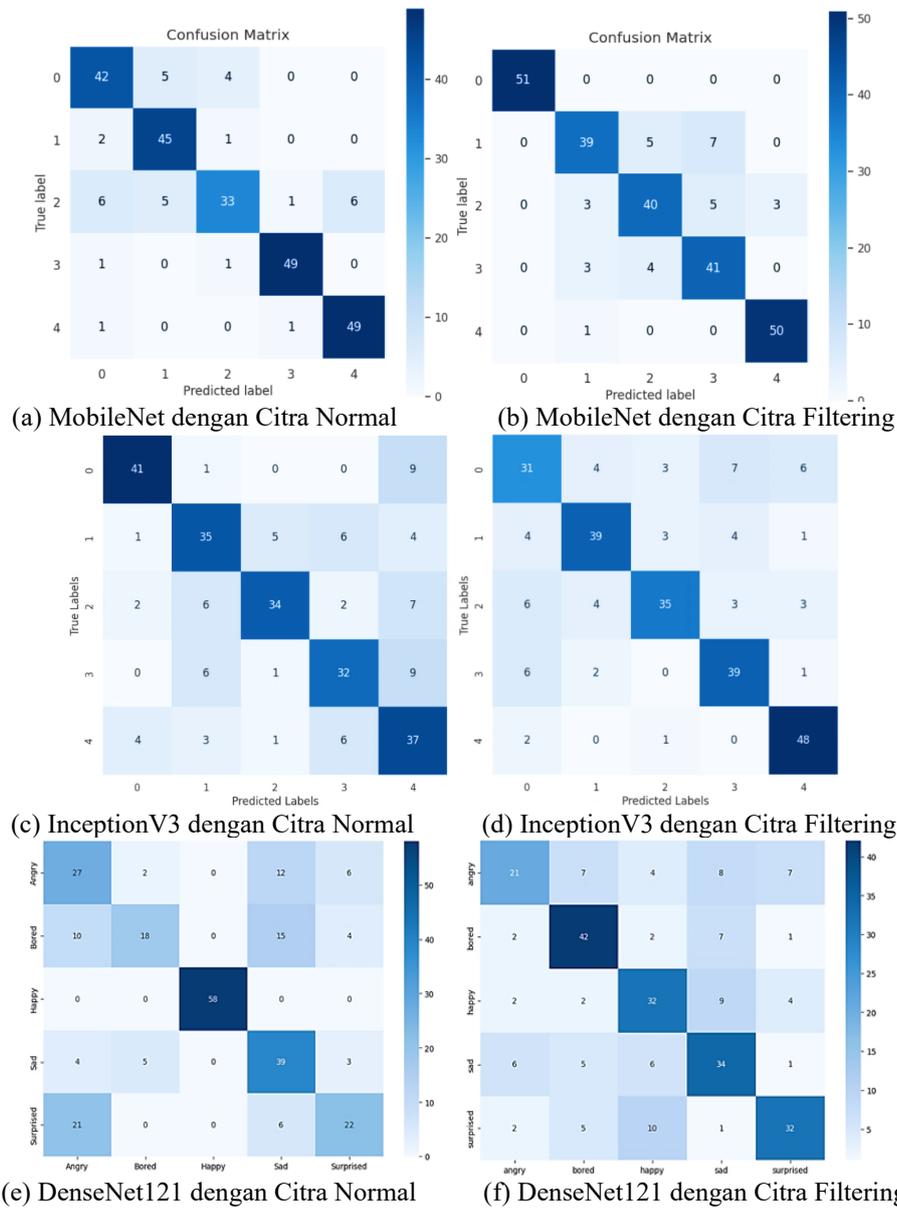


Fig. 8. Confusion Matrix of MobileNet (a and b), Inception V3 (c and d), and DenseNet121models (e and f)

4.2. Discussion

The performances of each model were evaluated in term of accuracy, precision, recall, and F-1 score. The addition of evaluation metric, including PSNR, SSIM, and MSE, also contributed to quantitatively prove the improvement of image quality after pre-processing. Table 4-Table 7 demonstrate the evaluation results of the MobileNet, InceptionV3, and DenseNet121 model. MobileNet with normal images already performs very well, especially at learning rates of 10-2 and 10-4, with precision, recall, F1-Score, and accuracy in the range of 0.80-0.87. However, there is a decrease at learning rate 10-6. Compared to normal images, filtered images enhance MobileNet's performance across all learning rates. At learning rate 10-6, the filtered images improved

precision, recall, F1-Score, and accuracy up to 0.89. This makes MobileNet the best model because it consistently maintains high performance with both normal and filtered images. A high accuracy shows the ability of model in classifying majority data correctly. However, model bias towards majority class may possibly occurs, as the accuracy only measure the proportion of correct predictions out of the total data. The small recall value indicates the model struggles to detect many true positives for certain classes, often due to a bias toward the majority class while ignoring the minority class. Although filtering enhances image quality, excessive modification or missing features can make it harder for the model to recognize specific classes. On the other hand, MobileNet achieves high precision due to its depthwise separable convolutions, which effectively extract key features like edges and textures. This filtering improves the clarity of dominant features, enhancing precision for certain classes. The balance between high precision and relatively stable recall contributes to a strong F1-score, as MobileNet effectively captures important features.

Next, InceptionV3 shows a consistent performance when using normal images at learning rates 10^{-2} and 10^{-4} , with accuracies reaching 0.71 and 0.72. However, at learning rate 10^{-6} , performance drastically drops to 0.56. Enhanced images show a significant improvement compared to normal images. At learning rates 10^{-2} and 10^{-4} , accuracy reaches 0.77 and 0.75, and at 10^{-6} , its accuracy is still better than with normal images, although the accuracy is lower compared to the other two learning rates. This makes InceptionV3 the second best performing model after MobileNet. Higher precision in InceptionV3 is due to its ability to capture important features across various scale, making it better in detecting certain classes with clear, dominant feature. Filtering further enhanced the dominant features, increasing the model's confidence in predicting positively correct. However, the recall is relatively low because InceptionV3 struggles to capture features from minority or less dominant classes, partly due to data homogenization after filtering. Additionally, InceptionV3 often prioritizes precision over recall, particularly in the case of imbalanced data distribution. This leads to a decent F1-score value, though not as high as MobileNet.

Finally, the performance of DenseNet121 with normal datasets shows a decline in performance at learning rates 10^{-2} through 10^{-6} , with other evaluation metrics decreasing and accuracy dropping from 0.64 to 0.20. Meanwhile, enhanced datasets show slightly better performance at learning rates 10^{-2} and 10^{-4} , with precision and accuracy remaining around 0.65 before dropping drastically at 10^{-6} . This makes DenseNet121 the model with the poorest performance compared to the other models, although using the filtered image dataset still improves its accuracy compared to using the normal dataset. DenseNet121 shows low precision due to overfitting on an insufficient variety of dataset. This architecture requires a large and diverse dataset to learn features effectively, but high homogeneity of filtered dataset leads to many incorrect positive prediction. Moreover, the filtering may result in the loss of fine details required by DenseNet121, reducing its ability in accurately detecting certain classes. Recall is also low because DenseNet121 struggles to capture positive samples from minority classes due to its reliance on rich and detailed features, which are diminished in the filtered dataset. Efforts to improve precision may further sacrifice recall, leaving some positive samples undetected. The imbalance between low precision and recall results in a poor F1-score, indicating suboptimal performance on both enhanced and normal datasets. The evaluation also reveals that an excessively small learning rate (10^{-6}) hinders learning progress, preventing the model from effectively capturing patterns.

Table 7 provided shows a performance comparison of different models when applied to both the normal dataset and the enhanced dataset. The table illustrates the impact of dataset enhancement on the performance of three different models: MobileNet, InceptionV3, and DenseNet121. The enhanced dataset significantly improves the performance of MobileNet, with accuracy increasing from 38.09% to 88.88%. This demonstrates that MobileNet benefits the most from the image enhancement process, making it more effective in extracting features from the improved input data. InceptionV3 shows a modest accuracy improvement from 71.03% to 78.17% with the enhanced dataset. This indicates that while the enhancement helps, InceptionV3 already performs well on the normal dataset due to its more advanced architecture. DenseNet121 experiences only a slight accuracy increase, from 63.89% to 65.48%, when using the enhanced dataset. This suggests that DenseNet121 may already have strong feature extraction capabilities, making the enhancement process less impactful.

Table 4. The evaluation results of the MobileNet

Testing Evaluation Parameters	Normal Image			Filtering Image		
	10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}
Precision	0.80	0.86	0.39	0.84	0.88	0.89
Recall	0.80	0.87	0.38	0.82	0.88	0.89
F1-Score	0.80	0.86	0.37	0.82	0.88	0.89
Accuracy	0.81	0.87	0.38	0.82	0.88	0.89

Table 5. The evaluation results of the InceptionV3

Testing Evaluation Parameters	Normal Image			Filtering Image		
	10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}
Precision	0.76	0.73	0.56	0.78	0.75	0.60
Recall	0.71	0.72	0.56	0.77	0.75	0.59
F1-Score	0.72	0.72	0.56	0.77	0.75	0.58
Accuracy	0.71	0.72	0.56	0.77	0.75	0.59

Table 6. The evaluation results of the DenseNet121

Testing Evaluation Parameters	Normal Image			Filtering Image		
	10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}
Precision	0.65	0.64	0.22	0.66	0.67	0.22
Recall	0.64	0.63	0.20	0.64	0.63	0.21
F1-Score	0.64	0.63	0.19	0.63	0.63	0.21
Accuracy	0.64	0.63	0.20	0.65	0.65	0.22

Table 7. Performance Comparison of the Models

Model	Learning Rate	Accuracy of Normal Dataset	Accuracy of Enhanced Dataset
MobileNet	10^{-6}	38.09%	88.88 %
InceptionV3	10^{-4}	71.03 %	78.17 %
DenseNet121	10^{-2}	63.89%	65.48 %

Although Homomorphic filtering is adequate to improve visual quality and model accuracy, the filtering technique may lead to increased computational load due to complexity in Fourier Transformation. Fourier Transformation requires significant computational time for a large dataset particularly. To address this issue, the utilization of GPU or parallel implementation is required to reduce overhead. Additionally, the risk of overfitting occurs in DenseNet121 as this architecture has high number parameter and requires more diverse data. Limited data augmentation or variation makes it difficult for the model to recognize patterns under different conditions. Overfitting can be reduced through data augmentation technique. In addition, hyperparameter sensitivity, particularly parameters like low, high, and d0, impact final results, requiring further testing to determine optimal values for various datasets. Overall, the results show that image enhancement significantly benefits complex architectures like InceptionV3, which excels in capturing detailed features. However, the enhancements for MobileNet and DenseNet121 are more limited, indicating that image enhancement effects are not always directly proportional to model performance.

5. CONCLUSION

This research aims to explore the impact of image quality enhancement using homomorphic filtering and sharpening to improve accuracy and support the performance of facial emotion recognition based on deep learning. Image quality enhancement is used to improve the quality of input images, allowing the classification model to focus on relevant expression features. The testing was conducted using MobileNet, InceptionV3, and DenseNet121 architectures. The study demonstrates that image quality enhancement significantly improves facial emotion recognition performance. The three architectures tested—MobileNet, InceptionV3, and DenseNet121—showed an increase in accuracy. MobileNet emerged as the most stable model, with an accuracy increase of 1-2%, achieving the best result of 89% at learning rate 10^{-6} . The consistency of this model in maintaining accuracy across various learning rates highlights MobileNet's ability to handle enhanced quality images. InceptionV3 also showed positive results, with a more significant accuracy improvement of 5-7%, reaching the highest accuracy of 78.17% at learning rate 10^{-2} . Although its performance was lower with normal images, applying filtering to the input images substantially improved its performance, especially at lower learning rates. DenseNet121 also experienced an accuracy increase of 1-3%, but it still lagged behind the performance of the other two models, particularly at learning rate 10^{-6} , where its accuracy dropped significantly. Overall, the results of this study emphasize the importance of image quality enhancement in improving the accuracy of expression classification systems. Models trained with enhanced images proved to be more effective than those trained with non-enhanced images. Therefore, image enhancement techniques such as homomorphic filtering and image sharpening can be relied upon to improve image quality and support the development of more accurate and reliable classification systems, particularly in applications involving low-quality images. While challenges remain, these results offer a strong basis for advancing more reliable and efficient expression classification systems in the future. To further improve, extensive testing with datasets

featuring diverse resolutions and expressions, along with real-world scenarios involving varied lighting and angles, is essential.

Acknowledgments

We would like to express our gratitude to Universitas Syiah Kuala for their funding support for this research. This research was carried out in accordance with the Research Implementation Assignment Agreement for the Lecturer Year 2024, with Number: 460/UN11.2.1/PG.01.03/SPK/PTNBH/2024, issued on May 3, 2024. This support has been invaluable in ensuring the success of all stages of the research.

REFERENCES

- [1] S. Dwijayanti, M. Iqbal, and B. Y. Suprpto, "Real-Time Implementation of Face Recognition and Emotion Recognition in a Humanoid Robot Using a Convolutional Neural Network," *IEEE Access*, vol. 10, pp. 89876–89886, 2022, <https://doi.org/10.1109/ACCESS.2022.3200762>.
- [2] N. Zhou, R. Liang and W. Shi, "A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection," in *IEEE Access*, vol. 9, pp. 5573-5584, 2021, <https://doi.org/10.1109/ACCESS.2020.3046715>.
- [3] A. Dutta, H. Singh, A. Majumdar, A. Maiti, "Improved Facial Emotion Recognition using Convolutional Neural Network," *Conference: The 14th Inter-University Engineering, Science & Technology Academic*, 2024, https://www.researchgate.net/publication/381772613_Improved_Facial_Emotion_Recognition_using_Convolutional_Neural_Network.
- [4] R. Sowmiya, G. Sivakamasundari, V. Archana, "Facial Emotion Recognition using Deep Learning Approach," *International Conference on Automation, Computing and Renewable Systems (ICACRS)*, pp. 1064-1069, 2022, <https://doi.org/10.1109/ICACRS55517.2022.10029092>.
- [5] F. Zhao, J. Li, L. Zhang, Z. Li, S. G. Na, "Multi-view face recognition using deep neural networks," *Future Generation Computer Systems*, vol. 111, pp. 75-380, 2020, <https://doi.org/10.1016/j.future.2020.05.002>.
- [6] L. Zahara, P. Musa, E. P. Wibowo, I. Karim, and S. B. Musa, "The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi," in *Fifth International Conference on Informatics and Computing (ICIC)*, pp. 1–9, 2020, <https://doi.org/10.1109/ICIC50835.2020.9288560>.
- [7] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in *Proceedings of the International Conference on Image Processing (ICIP)*, pp. 2889–2893, 2020, <https://doi.org/10.1109/ICIP40778.2020.9191247>.
- [8] A. Singh, M. Kaur, "Deep Residual Networks Including Transfer Learning for Facial Emotion Identification," In *International Conference on Cognitive Computing and Cyber Physical Systems*, pp. 29-39, 2025, https://doi.org/10.1007/978-981-97-7371-8_3.
- [9] P. Giannopoulos, I. Perikos, I. Hatzilygeroudis, "Deep Learning Approaches for Facial Emotion Recognition: A Case Study on FER-2013," In *Advances in hybridization of intelligent methods: Models, systems and applications*, pp. 1-16, 2007, https://doi.org/10.1007/978-3-319-66790-4_1.
- [10] S. Daeng, "Pengaruh Histogram Equalization untuk Perbaikan," *Jurnal Simetris*, vol. 7, no. 1, pp. 177–182, 2016, <https://doi.org/10.24176/simet.v7i1.502>.
- [11] R. Al Sobhahi and J. Tekli, "Low-light homomorphic filtering network for integrating image enhancement and classification," *Signal Process Image Commun*, vol. 100, p. 116527, 2022, <https://doi.org/10.1016/j.image.2021.116527>.
- [12] N. A. Olobo, "Image Contrast Enhancement Using General Histogram Equalization and Homomorphic," *Filtering. Asian Journal of Science, Technology, Engineering, and Art*, vol. 3, no. 1, pp. 1-31, 2024, <https://doi.org/10.58578/ajstea.v3i1.4244>.
- [13] F. Arnia, M. Oktiana, K. Saddami, K. Munadi, R. Roslidar, and B. Pradhan, "Homomorphic Filtering and Phase-Based Matching for Cross-Spectral Cross-Distance Face Recognition," *Sensors* vol. 21, no. 13, p. 4575, Jul. 2021, <https://doi.org/10.3390/S21134575>.
- [14] M. Oktiana *et al.*, "Cross-spectral iris recognition using phase-based matching and homomorphic filtering," *Heliyon*, vol. 6, no. 2, Feb. 2020, [https://www.cell.com/heliyon/fulltext/S2405-8440\(20\)30252-8](https://www.cell.com/heliyon/fulltext/S2405-8440(20)30252-8).
- [15] Z. Yue, F. Yanyan, Z. Shangyou and P. Bing, "Facial Expression Recognition Based on Convolutional Neural Network," *IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, pp. 410-413, 2019, <https://doi.org/10.1109/ICSESS47205.2019.9040730>.
- [16] Y.-T. Han, G.-J. Lin, L.-J. Zhao, X.-L. Tang, Y. Huang, and H. Jiang, "An Improved Homomorphic Filtering Algorithm for Face Image Preprocessing," *J Comput (Taipei)*, vol. 32, no. 6, pp. 66–82, 2021, <http://www.csroc.org.tw/journal/JOC32-6/JOC3206-06.pdf>.
- [17] A. Thamizharasi, "Performance analysis of face recognition by combining multiscale techniques and homomorphic filter using fuzzy K nearest neighbour classifier," in *International Conference on Communication Control and Computing Technologies*, pp. 394–401, 2010, <https://doi.org/10.1109/ICCCCT.2010.5670584>.
- [18] H. Hafidz, A. Ananda, M. Akbar, "Perbaikan Citra RGB dengan Metode Homomorphic Filtering Menggunakan Butterworth Filter| Jurnal Komputer Terapan," *Jurnal Komputer Terapan*, vol. 1, no. 1, pp. 1-9, 2015, https://web.archive.org/web/20180426023059id_/https://jurnal.pcr.ac.id/index.php/jkt/article/viewFile/6/5.

- [19] P. Musa *et al.*, "Pembelajaran Mendalam Pengklasifikasi Ekspresi Wajah Manusia dengan Model Arsitektur Xception pada Metode Convolutional Neural Network," *Rekayasa*, vol. 16, no. 1, pp. 65-73, 2023, <https://doi.org/10.21107/rekayasa.v16i1.16974>
- [20] P. A. Nugroho, I. Fenriana, R. Arijanto, "Implementasi Deep Learning Menggunakan Convolutional Neural Network (CNN) Pada Ekspresi Manusia," *In Journal of Physics: Conference Series*, vol. 2, no. 1, pp. 12–20, 2020, <https://jurnal.buddhidharma.ac.id/index.php/algior/article/view/441>.
- [21] M. Muhajir *et al.*, "USK-FEMO: A Face Emotion Dataset using Deep Learning for Effective Learning," *Proceeding 2nd International Conference on Computer System, Information Technology, and Electrical Engineering: Sustainable Development for Smart Innovation System, COSITE 2023*, pp. 199–203, 2023, <https://doi.org/10.1109/COSITE60233.2023.10249834>.
- [22] T. I. Rizky, N. A. Hasibuan, and R. Syahputra, "RGB Image Improvement Using Homomorphic Filtering Method with Butterworth Filter," *Komik (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, pp. 41–48, 2019, <https://doi.org/10.30865/Komik.V3i1.1565>.
- [23] M. J. Seow and V. K. Asari, "Ratio Rule and Homomorphic Filter for Enhancement of Digital Colour Image," *Neurocomputing*, vol. 69, no. 7–9, pp. 954–958, 2006, <https://doi.org/10.1016/J.Neucom.2005.07.003>.
- [24] P. Yugander, C. H. Tejaswini, J. Meenakshi, K. S. Kumar, B. V. N. S. Varma, and M. Jagannath, "MR Image Enhancement Using Adaptive Weighted Mean Filtering and Homomorphic Filtering," *Procedia Comput Sci*, vol. 167, pp. 677–685, 2020, <https://doi.org/10.1016/J.Procs.2020.03.334>.
- [25] A. Chavarin, E. Cuevas, O. Avalos, J. Galvez, and M. Perez-Cisneros, "Contrast Enhancement in Images by Homomorphic Filtering and Cluster-Chaotic Optimization," *IEEE Access*, vol. 11, pp. 73803–73822, 2023, <https://doi.org/10.1109/Access.2023.3287559>.
- [26] U. Hairah, M. Wati, A.P. A. Masa, A. Septiarini, N. Puspitasari, O. Gultom, "Enhanced Corn Leaf Disease Classification Using DenseNet121 and Hybrid Machine Learning Models," *International Conference on Mechatronics Engineering (ICOM)*, pp. 417-422, 2024, <https://doi.org/10.1109/ICOM61675.2024.10652337>.
- [27] A. Priyanka, I. M. S. Kumara, "Classification of rice plant diseases using the convolutional neural network method," *Jurnal Ilmiah Teknologi Informasi*, vol. 12, no. 2, p. 123, 2021, <https://pdfs.semanticscholar.org/45b7/280b0a93d2a70870e22864cbf7bc3ace9036.pdf>.
- [28] D. A. Dharmawan, L. Listyalina, "Performance Analysis of Lung Cancer Diagnosis Algorithms on X-Ray Images," *Journal of Electrical Technology*, vol. 2, no. 2, pp. 34-40, 2018, <https://journal.umy.ac.id/index.php/jet/article/view/7429>
- [29] K. V. Jamuna, V. B. Ambhore, A. Amudha, J. Singh, N. T. Velusudha and N. Vashisht, "A Comparative Analysis of Noise Reduction Techniques in Magnetic Resonance Imaging," *International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-6, 2024, <https://doi.org/10.1109/ICCCNT61001.2024.10724152>.
- [30] W. Jiang and D. Xue, "Infrared Image Enhancement for Photovoltaic Panels Based on Improved Homomorphic Filtering and CLAHE," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14495, pp. 348–361, 2024, https://doi.org/10.1007/978-3-031-50069-5_29.
- [31] A. Malik *et al.*, "Design and Evaluation of a Hybrid Technique for Detecting Sunflower Leaf Disease Using Deep Learning Approach," *Journal of Food Quality*, p. 9211700, 2022, <https://doi.org/10.1155/2022/9211700>.
- [32] X. Fu and X. Cao, "Underwater Image Enhancement with Global-Local Networks and Compressed-Histogram Equalization," *Signal Process Image Commun*, vol. 86, p. 115892, 2020, <https://doi.org/10.1016/J.Image.2020.115892>.
- [33] I. N. Purnama, "Herbal Plant Detection Based On Leaves Image Using Convolutional Neural Network With Mobile Net Architecture," *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, vol. 6, no. 1, pp. 27-32, 2020, <https://doi.org/10.33480/jitk.v6i1.1400>.
- [34] Y. ; Hu *et al.*, "Detail Enhancement Multi-Exposure Image Fusion Based on Homomorphic Filtering," *Electronics*, vol. 11, no. 8, p. 1211, Apr. 2022, <https://doi.org/10.3390/ELECTRONICS11081211>.
- [35] S. Gamini and S. S. Kumar, "Homomorphic filtering for the image enhancement based on fractional-order derivative and genetic algorithm," *Computers and Electrical Engineering*, vol. 106, p. 108566, Mar. 2023, <https://doi.org/10.1016/J.COMPELECENG.2022.108566>.
- [36] S. D. Lalitha and K. K. Thyagarajan, "Micro-Facial Expression Recognition Based on Deep-Rooted Learning Algorithm," *arXiv preprint arXiv:2009.05778*, 2020. [Online]. Available: <https://arxiv.org/abs/2009.05778>.
- [37] Y. Hu *et al.*, "Detail Enhancement Multi-Exposure Image Fusion Based on Homomorphic Filtering," *Electronics*, vol. 11, no. 8, p. 1211, Apr. 2022. [Online]. Available: <https://www.mdpi.com/2079-9292/11/8/1211>.
- [38] M. Bentoumi, M. Daoud, M. Benaouali, and A. Taleb-Ahmed, "Improvement of Emotion Recognition from Facial Images Using Deep Learning and Early Stopping Cross Validation," *Multimedia Tools and Applications*, vol. 81, pp. 29887–29917, 2022. [Online]. Available: <https://link.springer.com/article/10.1007/s11042-022-12058-0>.
- [39] G. Balachandran, S. Ranjith, T. R. Chenthil, and G. C. Jagan, "Facial Expression-Based Emotion Recognition Across Diverse Age Groups: A Multi-Scale Vision Transformer with Contrastive Learning Approach," *Journal of Combinatorial Optimization*, vol. 49, 2025. [Online]. Available: <https://link.springer.com/article/10.1007/s10878-024-01241-8>.

- [40] S. Li and W. Deng, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2019. [Online]. Available: <https://doi.org/10.1109/TIP.2018.2868382>.
- [41] Y. Han, G. Lin, L. Zhao, X. Tang, Y. Huang, and H. Jiang, "An Improved Homomorphic Filtering Algorithm for Face Image Preprocessing," *Journal of Computational Science*, vol. 32, pp. 102–113, 2021. [Online]. Available: <http://www.csroc.org.tw/journal/JOC32-6/JOC3206-06.pdf>.
- [42] S. S. Jarande, P. K. Kadbe and A. W. Bhagat, "Comparative analysis of image enhancement techniques," *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pp. 307-309, 2016, <https://doi.org/10.1109/ICEEOT.2016.7755249>.
- [43] G. Balachandran, S. Ranjith, T. R. Chenthil, and G. C. Jagan, "Facial expression-based emotion recognition across diverse age groups: a multi-scale vision transformer with contrastive learning approach," *Journal of Combinatorial Optimization*, vol. 49, no. 1, Dec. 2024, <https://doi.org/10.1007/s10878-024-01241-8>.
- [44] C. Saravanan, M. Poonkodi, and P. Sankar, "Comprehensive Exploration of Facial Emotion Recognition using Conventional Machine Learning and Transfer learning Models," *Research Square (Research Square)*, pp. 161-172, 2024, <https://doi.org/10.21203/rs.3.rs-4304090/v1>.
- [45] S. Bobojanov, B. M. Kim, M. Arabboev, and S. Begmatov, "Comparative analysis of vision transformer models for facial emotion recognition using augmented balanced datasets," *Applied Sciences*, vol. 13, no. 22, p. 12271, Nov. 2023, <https://doi.org/10.3390/app132212271>.
- [46] N. Yalçın and M. Alisawi, "Introducing a Novel Dataset for Facial Emotion Recognition and Demonstrating Significant Enhancements in Deep Learning Performance through Pre-processing Techniques," *Heliyon*, vol. 10, no. 20, p. e38913, Oct. 2024, <https://doi.org/10.1016/j.heliyon.2024.e38913>.
- [47] S. Roka, D. B. Rawat, "Fine tuning Vision Transformer model for facial Emotion recognition: Performance Analysis for Human-Machine Teaming," *24th International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 134-139, 2023, <https://doi.org/10.1109/IRI58017.2023.00030>.
- [48] A. Chaudhari, C. Bhatt, A. Krishna, and P. L. Mazzeo, "ViTFER: Facial Emotion Recognition with Vision Transformers," *Applied System Innovation*, vol. 5, no. 4, p. 80, Aug. 2022, <https://doi.org/10.3390/asi5040080>.
- [49] J. Soni, N. Prabakar, and H. Upadhyay, "Vision Transformer-Based Emotion Detection in HCI for enhanced interaction," in *Lecture notes in computer science*, pp. 76–86, 2024, https://doi.org/10.1007/978-3-031-53827-8_8.
- [50] A. K. Roy, H. K. Kathania, and A. Sharma, "Improvement in Facial Emotion Recognition using Synthetic Data Generated by Diffusion Model," *arXiv.org*, Nov. 16, 2024. <https://arxiv.org/abs/2411.10863>.
- [51] N. N. Urnisha, S. I. Bithi, M. M. S. Rafee, N. I. Remon, M. M. Hasan, P. R. Chowdhury, "A Transfer Learning Approach for Facial Emotion Recognition Using a Deep Learning Model," *International journal of research and scientific innovation*, vol. 11, no. 4, pp. 274-284, 2024, <https://doi.org/10.51244/IJRSI.2024.1104022>.
- [52] A. Chowanda, "Implementing Vision Transformer to Model Emotions Recognition from Facial Expressions," *International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, pp. 48-53, 2023, <https://doi.org/10.1109/AiDAS60501.2023.10284712>.
- [53] Mejia-Escobar, C, Cazorla, M, Martinez-Martin, E. "Improving facial expression recognition through data preparation and merging," *IEEE Access*, vol. 11, pp. 71339-71360, 2023, <https://doi.org/10.1109/ACCESS.2023.3293728>.
- [54] R. Helaly, S. Messaoud, S. Bouaafia, M. A. Hajjaji, and A. Mtibaa, "DTL-I-ResNet18: facial emotion recognition based on deep transfer learning and improved ResNet18," *Signal Image and Video Processing*, vol. 17, no. 6, pp. 2731–2744, Feb. 2023, <https://doi.org/10.1007/s11760-023-02490-6>.
- [55] K. Desai, H. N. Patel, D. Dashlaniya, P. Desai, M. Shah and P. Goel, "A Transfer Learning Framework for Facial Emotion Recognition: Leveraging Pre-Trained Convolutional Neural Networks," *2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, pp. 663-667, 2024, <https://doi.org/10.1109/InCACCT61598.2024.10551190>.
- [56] S. B. Punuri, S. K. Kuanar and T. K. Mishra, "Facial Emotion Recognition in Unconstrained Environments through Rank-Based Ensemble of Deep Learning Models using 1-Cycle Policy," *2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM)*, pp. 1-8, 2023, <https://doi.org/10.1109/IC-RVITM60032.2023.10435159>.
- [57] T. Dar, A. Javed, S. Bourouis, H. S. Hussein and H. Alshazly, "Efficient-SwishNet Based System for Facial Emotion Recognition," in *IEEE Access*, vol. 10, pp. 71311-71328, 2022, <https://doi.org/10.1109/ACCESS.2022.3188730>.
- [58] U. Chinta and A. Atyabi, "Facial Expression Recognition Using Robust Attention-Based CNN," *2023 International Conference on Machine Learning and Applications (ICMLA)*, pp. 858-863, 2023, <https://doi.org/10.1109/ICMLA58977.2023.00126>.
- [59] J. Wang, "Evaluation and analysis of visual perception using attention-enhanced computation in multimedia affective computing," *Frontiers in Neuroscience*, vol. 18, Aug. 2024, <https://doi.org/10.3389/fnins.2024.1449527>.
- [60] K. Zakiieldin *et al.*, "ViTCN: Hybrid Vision Transformer with Temporal Convolution for Multi-Emotion Recognition," *Int J Comput Intell Syst*, vol. 17, p. 64, 2024, <https://doi.org/10.1007/s44196-024-00436-5>.

BIOGRAPHY OF AUTHOR

Al Bahri received his bachelor's degree in 2014 and his master's degree in 2018 from the School of Electrical Engineering and Informatics, Institut Teknologi Bandung (ITB). Since March 2019, he has been a member of the Department of Electrical Engineering, Faculty of Engineering, Syiah Kuala University. His research interests include digital media technology, games, and computers. He has conducted several research projects such as the SeaCyberclass Question Bank in 2013, geographic information systems for schools and hospitals in Bandung, and the implementation of the Fisher-Yates shuffle algorithm in e-learning systems. Email: albahri@usk.ac.id, ORCID: <https://orcid.org/0009-0004-0519-9634>.



Maulisa Oktiana earned her Bachelor's degree in Electrical Engineering (S.T.) from Syiah Kuala University (USK) in 2013. She obtained her Ph.D. in Electrical and Computer Engineering from Syiah Kuala University in 2020. She received a scholarship from the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia through the PMDSU scheme. In November 2018, she visited Chiba University as an exchange student. Currently, she is a lecturer at the Department of Electrical Engineering and Computer Science at Syiah Kuala University. Her research interests include image processing, biometrics, and pattern recognition. Email: maulisaoktiana@usk.ac.id, ORCID: <https://orcid.org/0000-0001-7453-7183>.



Maya Fitria began her career as a lecturer and member of the Department of Electrical Engineering and Computer Science at Syiah Kuala University in 2017 and has been there ever since. She earned her Bachelor's degree in Computer Science from the University of Indonesia (UI) in 2012. In 2013, she continued her studies at the Department of Computer Engineering, University of Duisburg-Essen, Germany, specializing in Interactive Systems and Visualization. During her studies, she received support from the DAAD-LPSDM Aceh Scholarship. She completed her master's degree in 2016, earning a Master of Science in Computer Engineering. Her research interests include Human-Centered AI and IoT for Diagnostic and Interactive Systems. Email: mayafitria@usk.ac.id, ORCID: <https://orcid.org/0000-0002-6593-212X>.



Zulfikar, a graduate of SMAN 1 Seunagan, is currently pursuing a degree in Electrical and Computer Engineering, starting from the 2020 cohort. His research interests include image processing and classification. Email: zul20@mhs.usk.ac.id.