Facial image recognition using hybrid filtering models and convolutional neural networks (CNNs)

Dading Oktaviadi Resmiranta[1], Bambang Krismono[2], Khasnur Hidjah[3]
[1,2,3,]Bumigora University
dading@universitasbumigora.ac.id[1],bkrismono@universitasbumigora.ac.id[2],khasnur72.h@universitasbumigora.ac.id[3]

Sent: July 3, 2025.; Revised: July 19, 2025 .; Accepted: October 30, 2025

Abstract

The face is an important object in the biometric identification system. However, low image quality due to uneven lighting, noise, and variations in facial expressions can interfere with the accuracy of the recognition system. The study investigated the use of Convolutional Artificial Neural Network (CNN) combined with hybrid screening techniques to improve image quality, thereby improving the accuracy of facial recognition systems. Filters used include weight mean filtering, median filtering, Contrast-Limited Adaptive Histogram Equalization and gaussian filtering, wavelet filtering. The pre-processed image was then trained using image denoising measurements of the Structural Similarity Index, Mean Squared Error, and Peak Signal to Noise Ratio. The main objective of this study is to evaluate the best filtration combination to produce high accuracy in face classification. The datasets used were 55 classes and 100 images per class. The inceptionV3 architecture model is used for classifications with a number of epochs of 10. Evaluation was carried out on a facial data set with an 80%:20% scheme. The results of the experiment showed that the hybrid method produced the best performance with 94.5% validation accuracy, 94.2% precision, and 94.6% recall, an increase of +1.4% compared to baseline. The (original) baseline itself recorded 93.1% validation accuracy, 92.8% precision, and 93.2% recall. In addition, the loss graph shows that the pre-process model has faster and more stable convergence than the non-pre-processing model. These results confirm that the application of preprocessing, especially the hybrid approach, is able to improve the accuracy and stability of the model in image classification tasks.

Keywords*: Facial Recognition, Hybrid Screening, CNN, Image Classification, Deep Learning.*

1. Introduction

The face is one of the typical parts of the human body that has individual qualities that store various important information, such as identity and emotional expressions. Through facial expressions, a person can display self-actualization in social interactions, so emotional expression is an important aspect in building effective communication [1], [2].

In the context of technology, facial recognition is included in the field of computer vision, it has been widely applied to various systems such as social media platforms, security systems, and biometric authentication and verification. Images or photographs of human faces contain important information that can be used to identify or classify individuals [3].

Recognition of objects and patterns through computer vision, face detection is a very important basic step [4]. This technology is increasingly crucial in the digital age, especially in applications related to security and biometrics. However, the accuracy of the facial recognition system is highly dependent on the quality of the input image. Common problems such as noise, low resolution, and uneven lighting are often the main causes of failures in the identification process [5].

This challenge is compounded by the limitations of traditional preprocessing techniques, which only work well in certain situations, making this task even more difficult. Its real-world application is limited by its reliance on artificial noise in facial photographs, which are incapable of depicting the complexity of natural environmental degradation. Forensic identification relies heavily on the legitimacy of the data source, but many of the image augmentation methods currently in use risk altering the substantial content of the original facial image, making it unsuitable for this kind of use. The study specifically examines the impact of image processing on modern deep facial recognition systems, such as the more robust and widely used MXNet architecture. This study shows through quantitative and comparative research that the denoising approach is more specifically, Gaussian screening is superior to enhancement strategies when it comes to improving performance in real-world situations. The study offers more practical and reliable insights to improve facial recognition performance in forensic identification by focusing on real-world data and cutting-edge deep learning systems [6]. One of the approaches used in facial recognition is artificial neural networks, especially deep learning. Deep learning technology is known to be able to achieve high accuracy in object recognition due to its ability to learn complex patterns from large amounts of data without explicit programming [7]. However, deep learning has limitations, such as the need for large, quality training data. Data bias is data that does not represent the population fairly or evenly, so it can cause the results of model analysis or training to be inaccurate or biased can reduce model performance. Additionally, deep learning models are often considered "black boxes" because they are difficult to track the decision-making process, and require high computational resources [8].

Previous research by [6]. Demonstrate the importance of the pre-processing stage in improving image quality and investigate more advanced noise reduction and augmentation methods. To ensure practitioners in this field can adapt to changing barriers in facial recognition technology, this opens up opportunities for future research to improve and expand the devices they have access to, such as the use of gaussian filtering to reduce noise and clarify facial contours in forensic images. Another study by [9]. It applies CNN for face shape detection with an accuracy of 74% of 5,000 images, while VGG-Face achieves an accuracy of up to 97% in cross-age facial recognition. These findings drive the need for optimization in denoising techniques, dataset expansion, and model tuning. Another study by [10] Research showed that the facial recognition and verification system based on computer vision achieved a high accuracy rate of 98.5714% after testing. The study used large datasets made up of pre-labeled images, which are important for training facial recognition models. A total of 400 face images are categorized into four classes: the faces of Serena, Salma, Silvio, and Spencer, with each class containing 100 images. The training dataset consists of 70 images per class, while the testing dataset includes 30 images per class, summing up to 400 images of faces for evaluation. The images used in this study are in JPG format. To improve detection accuracy, future research on facial recognition and verification should look at changes in lighting conditions. During shooting, learning different angles and perspective angles can improve the resilience of the recognition system. It is recommended to conduct performance analysis using various metrics to evaluate the effectiveness of various algorithms. Combining facial recognition and verification algorithms can lead to better results in practical applications. Increasing the diversity of Datasets to include broader facial features and expressions can improve model training. Another study by [11] This study used 6107 datasets for training.412 test images were used for image testing. Deep Neural Network (DNN) for face detection. FaceNet for facial recognition Where the FaceNet model achieved 97,48% accuracy on 6107 data sets. The accuracy of the test image is 97% with 412 test images. The accuracy of the real-time testing varies: 90% normal, 83% glasses, 70% masks. The accuracy of the cover condition is 81%, all accessories are 66%. The system operates as expected for facial recognition tasks. The impact of lighting conditions on accuracy needs further exploration. Performance with a wide range of accessories requires additional testing and analysis. Real-time processing efficiency under various conditions is not fully addressed.

Unlike previous research, this study focuses on optimizing hybrid screening techniques through the hyperparameter tuning process, to obtain a more accurate facial recognition model. Adjusting parameters such as the number of times, the proportion of data (training and testing), and the selection of optimizers are important factors in improving model performance. In addition, image quality was evaluated through denoising measurements using three main metrics: Mean Square Error (MSE), Structural Similarity Index (SSIM), and Peak Signal-to-Noise Ratio (PSNR). Thus, the main objective of this study is to improve the accuracy of face classification through a series of experiments involving a combination of training schemes and the best hybrid screening methods, based on the evaluation of image quality and model performance.

2. Method

The methodology used to understand the research process, is shown in Figure 1. stages carried out to obtain results in this study. The initial stage starts from the data collection process, followed by dataset processing, pre-processing, character extraction, training and model testing process, to evaluation of results. The following is an explanation of each stage that is carried out systematically in this study.

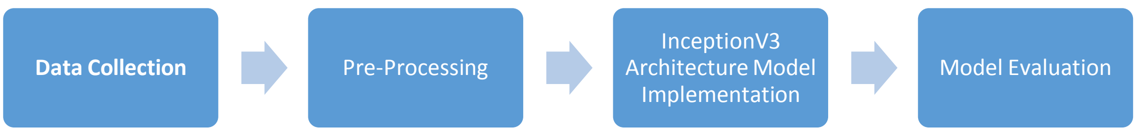


Figure 1. Research Flow

2.1 Data Collection

Table 1 of this research stage is to use primary data from real people's faces, and secondary data from the images of Celebrity Images taken will be used as data for model training and testing. The primary dataset of 27 people 1 person took 100 pictures of faces in various positions, the secondary dataset of Celebrity images of 28 people 1 person took 100 pictures of faces. This research is aided by this additional data set of celebrity photos, which offers selected visual data sources that are often used in facial recognition, expression analysis, and visual identity modeling investigations. His accurate depiction of diverse lighting scenarios, perspectives, and facial expressions that reflect the real difficulties in advances in computer vision technology makes it relevant. These secondary datasets are often used for face detection, attribute recognition, and face editing. You can access and download it https://www.kaggle.com/Datasets/hemantsoni042/Celebrity-images-for-face-recognition.

Table 1 Dataset (Face Image Data)

NOT	Dataset Face Images		
	Class(label)	**Photograph**	**Sum**
1	Aamir Khan		The 'Aamir Khan' class has 191 images.
2	Abhay Deol		The 'Abhay Deol' class has 128 images.
…	....	....	....
55	Zelda_Suryatmi		Class 'Vivi_Uyainah' has 194 images.

2.2 Pre-processing

Pre-processing is one of the important stages to prepare a dataset to facilitate the next process. At this stage, the pre-processing of the dataset is analyzed before being used to the characteristic extraction stage. The pre-processing stages are Croping, Resize, Image Augmentation, Grayscale, and Filtering.

Figure 2 Pre-Processing Stage

Figure 2 The pre-processing stage describes the stages in the digital image pre-processing process before the analysis or evaluation of image quality. The first stage is cropping, which is cropping the image to get a complete image of the face. This process aims to obtain a complete image of the face, so that the facial features can be clearly recognized. The image will be cropped on several sides, leaving a portion of the face image on the image [12]. Furthermore, the cropped image undergoes a resizing process, resizing is a procedure that changes the dimensions of the size of the image in pixels. In resizing, the pixels that make up the image can be reduced or added to achieve the desired resizing change. This stage is done to maintain consistency in the pixel intensity distribution, so that noise is easily identified and eliminated [13]. After that, the image is converted to a grayscale format, grayscale is an imaginary archiving attempt that changes the appearance of the image's center pixels to a grayish image. Since Grayscale fantasy is a fantasy where each pixel contains the world plus the algae spirit life view 0 is visible 255, the story of the Grayscale fantasy pixel life view can be represented by a matrix so that the computing power of Grayscale's imagination is limited. The next exercise. The Grey Scale formula can be seen in Equation 1. [14].

$$Grayscale = \frac{R+G+B}{3} \tag{1}$$

Description: R = Red, G = Green, B = Blue

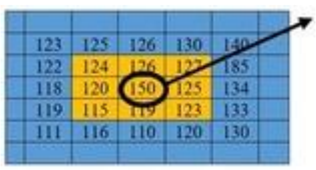
The next stage is hybrid filtering, which is the application of a combined filtering method to reduce noise in the image while retaining important details. Filtering used Weight-average filtering In this method, each pixel in the filtering window not only calculates the average, but is also given a different weight based on its position and characteristics. This allows the algorithm to better retain important details in the image while reducing noise. This weight is usually greater for pixels closer to the center pixel and smaller for the farther pixels. In this way, when an image is affected by gaussian noise, weighted average filtering attempts to isolate the points that affect the noise by giving less weight to those pixels, thereby reducing the impact of noise on the final result [15]. An example of Kernel Average Filtering using a 3 x 3 matrix can be seen in Figure 3. To understand the role of each Filter in the process, the formula and the name of the Filter used can be seen in Equation 2.

$$\frac{1}{16} \times \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Figure 1 Kernel Weight Average Filtration

$$\text{g (x, y)} = \frac{\sum_{s=-a}^{a} \sum_{t=-b}^{b} w\,(s,t) f\,(x+s, y+t)}{3 \sum_{s=-a}^{a} \sum_{t=-b}^{b} w\,(s,t) f} \tag{2}$$

f(x+s, y+t) is the gray value of the pixels around the midpoint (x, y), g(x, y) is the approximate gray value of the middle pixel in the filter, and w(s, t) is the weight that corresponds to the pixels f(x+s, y+t) in the filter. Next Filter Median filtering function to reduce noise and smooth digital images. It is said to be nonlinear because the way this window or filter works does not fall under the category of convolutional operations. Non-linear operations are calculated by sorting the values of a group of pixels or environmental values, and then replacing the value of the processed pixels with the middle or median values of all of their neighbors [16]. An example of Median Filtering solving using the m xn matrix, namely a 3 x 3 matrix, can be seen in Figure 4. The Filter formula used can be seen in Equation 3.



What pixels are you looking for

Figure 2 Matrix Median Filtering

$$g(x, y) = median\{f\ (x - i, y\ - j), i, j\ \varepsilon\ w\} \tag{3}$$

f(x,y) = Original image, g(x,y) = Result of each image, W = Matrix m x n (odd value) such as 3x3, 5x5 and so on. Next: The Contrast Limited Adaptive Histogram Equalization filter was created to address the two main weaknesses of AHE to handle excessive noise amplification and contrast. CLAHE not only applies Histogram Equalization locally, but also limits the resulting contrast to prevent excessive contrast effects on certain parts of the image. To avoid excessive intensity values, or clip limits, the local histograms on each tile are bounded before adjusting. If the number of pixels on each tile exceeds a certain limit, the histogram will be limited or "truncated". This reduces noise amplification and ensures contrast does not exceed acceptable limits [17]. The Filter formula used can be seen in Equation 4.

$$\beta = \frac{M}{N}\left\{1 + \frac{\alpha}{100}\left(S_{max} - 1\right)\right\} \tag{4}$$

M = Area size, N = Gray scale value (0-256), ɑ = Clipfactor 0-100, Smax = maximum pixel value. Filters Next, gaussian filtering results in a smoother image. This smoothing technique uses a Gaussian distribution to reduce noise in the image. This filter works by contorting the image with a gaussian kernel, which produces a smoothing effect by reducing the variation in intensity between adjacent pixels [18]. The Filter formula used can be seen in Equation 5.

$$G(x.y) = e - \frac{x^2+y^2}{2\sigma^2} \tag{5}$$

e = 2.71 (Euler constant), σ = Gaussian function width, g(x,y) = Convolutional result image, x2 and y2 respectively represent the distance between the other surrounding pixels and the surrounding middle pixels, and represent the standard deviation. The results of this screening are then evaluated using a denoising measurement, namely Mean Square Error (MSE). This is the sigma result of the total error that occurred between the Filter result image and the original image. The MSE value is generated from the difference in the resulting image that has the same pixel location as the original image. The higher the MSE value, the greater the difference between the original image and the final image that can be seen in Equation 6.

$$MSE = \frac{1}{WH}\sum_{w=1}^{W} \sum_{h=1}^{H} \left[Ii(x,y) - lo(x,y)\right]^2 \tag{6}$$

(x,y) is the pixel coordinate, WH is the image size, expressing the enhanced image, and Ii(x,y) lo(x,y) expresses the original image. The next measurement of denoising is the Structural Similarity Index (SSIM). The method used to measure the similarity between two images, is usually the reference image and the distorted version. It aims to provide a more accurate assessment of perceived image quality compared to traditional metrics by taking into account the changes in structural information, lighting, and contrast that can be seen in Equation 7.

$$SSIM(x,y) = \frac{(2\mu_{xy} + c1)(2\,\sigma_{xy} + c2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)} \tag{7}$$

μx is the average value for the original image, μy is the average value for the noise-muffled image, σx is the standard deviation of the original image, σy is the standard deviation of the noise-muffled image, and σxy = μxy - μx μy is the covariances. C1 and C2 are two variables that must be avoided to be divided by zero. Signal-to-Peak Noise Ratio (PSNR). Works by calculating the signal from the noise. PSNR is calculated by dividing the maximum value that a pixel can represent, divided by the square root of the MSE. The higher the PSNR value, the better the quality of the resulting image can be seen in Equation 8.

$$PSNR = 20\log_{10}\left(\frac{255}{\sqrt{MSE}}\right) \tag{8}$$

Based on the measurement results, the image then goes through the image data denoising process, which is the final stage to remove the noise residue that still exists in the image, resulting in a cleaner and more optimal finish for use in the next digital image processing application. This process is iterative and integral to improve image quality before it is used in further analysis or visual-based systems.

2.3 InceptionV3 Architecture Model Implementation

Inception V3 is a convolutional neural network architecture from the Inception family that makes several improvements including using Label Smoothing, Factorized 8x8 convolutions, and the use of additional classifers to propagate Label information down the network (along with the use of normalization batches for layers on the sidehead). Each facial image information processed at this stage will result in 2048 new features [19].

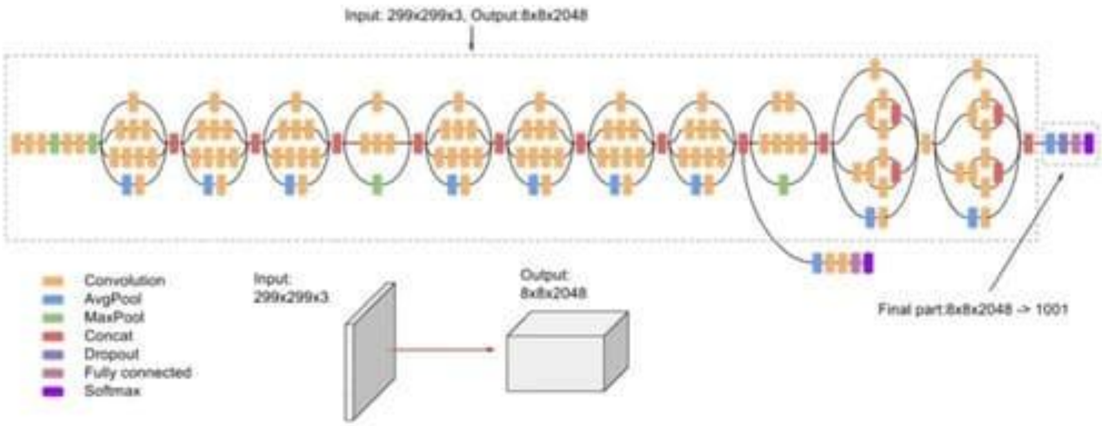


Figure 3 InceptionV3 Architecture

2.4 Evaluation of Results

Once the model has finished being trained, the next stage is performance evaluation. The evaluation was conducted to assess the model's ability to recognize faces using *test datasets* that were not used during training. The accuracy, precision, recall and F1-Score metrics, which are used for evaluation, can be seen in Equations 9-12.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{9}$$

Accuracy The percentage of a correctly recognized facial image.

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

Used to measure the accuracy and sensitivity of the model in recognizing faces.

$$F1\ Score = \frac{1}{\frac{1}{recall}+\frac{1}{Precision}} \tag{12}$$

The harmonious average of *Precision and Recall*, gives a more balanced picture between the two.

Table 2 Confussion Matrix

Predictions \ Current	*Positive*	*Negative*
Positive	*True* Positive (TP)	*False* Positive (FP)
Negative	*False* Negative (FN)	*True* Negative (TN)

Here are the terms used to describe the model's performance:
1. *True* Positive (TP) is the actual *data of Positive* that is correctly predicted.
2. *True* Negative (TN) is actual *Negative* data that is correctly predicted.
3. *False* Positive (FP) is actual *Negative data* that is incorrectly predicted as *Positive*.

3. Results and Discussion

Evaluation of post-pre-processing image quality showed significant differences between the filtration methods. The MSE (Mean Squared Error), PSNR (Peak Signal-to-Noise Ratio), and SSIM

(Structural Similarity Index) metrics are used as a measure of the difference between the Filter image and the original reference image. This study follows the flow as shown in Figure 1, starting from data collection, pre-processing, application of the CNN model based on the InceptionV3 architecture, to the evaluation of classification results. Each stage contributes to the final quality of the facial recognition system.

3.1. Data Collection

The data used consisted of 55 face classes, each consisting of 100 images obtained from a combination of primary (student faces) and secondary (celebrity face images) datasets. This research is aided by this additional data set of celebrity photos, which offers selected visual data sources that are often used in facial recognition, expression analysis, and visual identity modeling investigations. His accurate depiction of diverse lighting scenarios, perspectives, and facial expressions that reflect the real difficulties in advances in computer vision technology makes it relevant.

3.2. Data Pre-processing

Each image goes through a pre-processing stage to ensure optimal image quality before being extracted using the CNN model. The main focus is on the application of various filtering techniques, each of which makes a unique contribution to improving the quality of facial features, including cutting, resizing, grayscale conversion, and the application of five filtering methods: CLAHE, gaussian, Weighted Mean, Median, Hybrid (combined), and baseline (unfiltered). The main goal of this stage is to reduce noise as well as three image quality metrics: SSIM, MSE, and PSNR. and clarify facial features before being extracted by CNN.

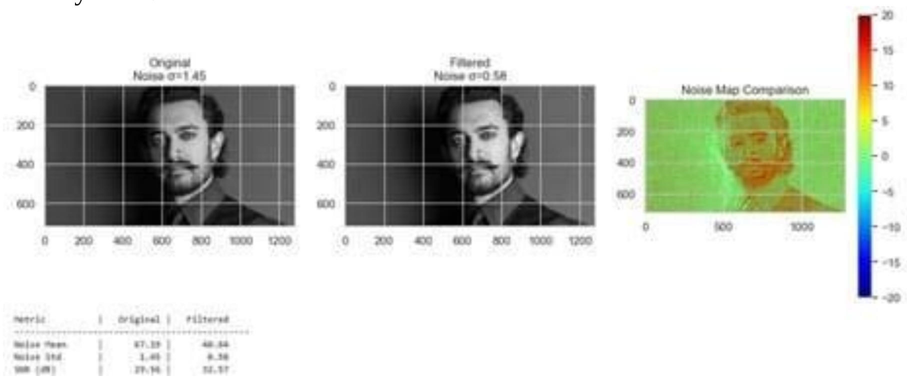


Figure 4 Comparison of Original Noise VS Filter

Figure 6 Original VS Filter Noise Comparison shows a visual comparison between the original face image and the filtered image using one of the hybrid methods. In the left image, the original image has a standard noise deviation (σ) level of 1.45, while in the middle image after filtering, the noise value is reduced to 0.58. This change suggests that the filtering method manages to significantly reduce noise without removing important details on the face. This is reinforced by the Signal-to-Noise Ratio (SNR) value which increased from 29.56 dB to 32.57 dB after the filtering process. Comparison of noise maps between original images and filtered images. Green areas indicate areas with low noise differences (stable), while red and blue indicate areas with changes in high noise intensity. It can be seen that most of the face, especially in the eye area and contours, is maintained, while the background area has experienced a fairly drastic reduction in noise. The Noise Mean value also decreased from 67.19 to 40.64, which indicates that the general noise distribution was successfully suppressed.

This visualization provides empirical evidence that the pre-processing process is able to effectively reduce noise noise noise without causing the loss of important structural features, supporting the results of previous metric evaluations such as SSIM and PSNR. These findings are also in line with the results of the classification which showed that images with low noise but preserved structures were easier to recognize by the CNN model.

Context Limited Adaptive Histogram Equalization, or CLAHE, adaptively enhances local contrast in facial imagery, which is particularly useful in photos with irregular illumination. By limiting histogram amplification, CLAHE avoids over-amplification and clarifies faces without generating artifacts. This technique facilitates the model's ability to extract features with more precise information from light and dark areas. The image is smoothed and Gaussian noise is reduced using Gaussian filtering. When used alone, Gaussian blur can blur edges but is effective in eliminating random pixel fluctuations. This technique is a non-linear technique that is very efficient to eliminate salt-pepper noise. Stabilizes the entire image structure to prevent random noise from interfering with CNN. Significant facial edges and features, including the contours of the mouth and eyes, are maintained by the median filter. Eliminates excess noise while maintaining crucial edge characteristics. A more selective smoothing effect is produced by a weighted average, which gives greater weight to the core pixels and lower weight to the surrounding pixels. Maintain the structural integrity of the face by smoothing the image proportionally. Combining the previously mentioned filtration techniques is known as hybrid filtration. By combining the benefits of each filter, hybrid filtering improves contrast, reduces noise, and preserves important facial features, all of which contribute to a significant improvement in CNN model accuracy.

Original	**The abu-abu scale**	**Clahe**	**Gaussian**	**Median**	**weighted_mean**	**Hybrid**

Figure 7 Face Image Processing

Figure 7 Face Image Processing Compares a series of image processing techniques applied sequentially to improve visual quality. The first column displays the original image in grayscale as the pre-processing basis. The second column shows the results of the implementation of Contrast Limited Adaptive Histogram Equalization (CLAHE), an adaptive contrast correction technique that enhances local detail without amplifying excessive noise. The third column visualizes the effect of gaussian filtration, which serves to dampen high-frequency noise while maintaining edge structure. The fourth column shows the results of the median filter, a nonlinear method that effectively eliminates impulsive noise (salt and pepper) without significantly obscuring the edges. The last column shows the weighted average result, which combines the advantages of the previous technique to achieve the optimal balance between noise smoothing and detail preservation. This gradual process illustrates a hybrid approach to image enhancement, where each stage makes a unique contribution to improving image quality.

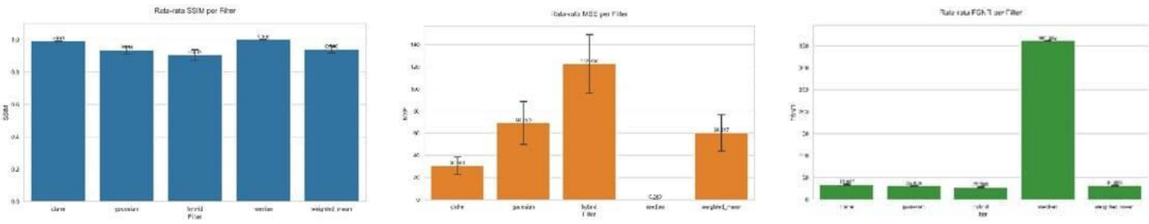


Figure 8 results of SSIM, MSE, and PSNR Average Measurements

Figure 8 contains the average SSIM, MSE, and PSNR values of the five types of denoisesing filters—clahe, gaussian, hybrid, median, and weighted_mean—that show varying performance in maintaining facial image quality. The median filter consistently performs best, with a perfect SSIM value (1,000), MSE 0.000, and the highest PSNR (361,202), signifying its ability to remove noise without damaging the original image structure or quality. The filter clahe also showed good results with high SSIM (0.991) and low MSE (30.601), while weighted_mean recorded a fairly stable performance with SSIM values of 0.940 and PSNR 31.026. In contrast, hybrid filters showed the least optimal results, with the highest MSE values (122,460) and the lowest PSNR (27,599), suggesting that the combination of techniques in such filters was less effective in this context. Overall, the information in the table confirms that the median filter was the most effective method for the face denoising process in this study.

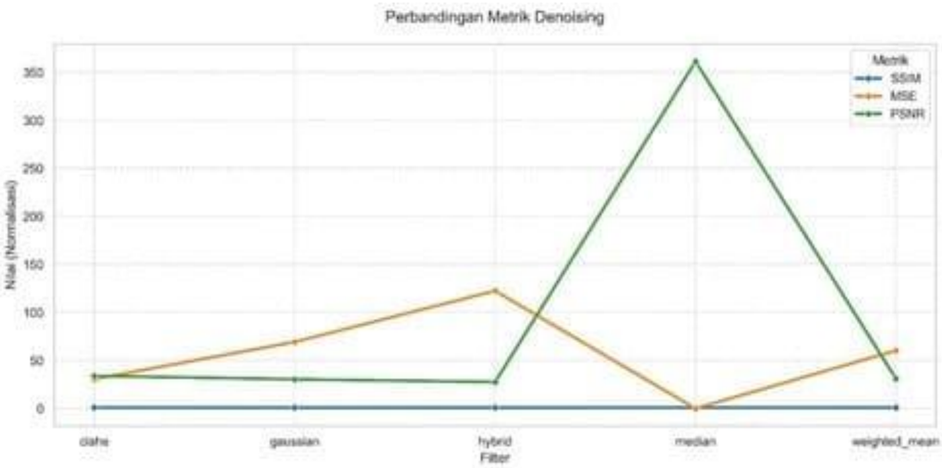


Figure 9 Denoising Metrics Comparison Chart

Figure 9 The denoising metric comparison graph shows that the median filter significantly excels in two main metrics, i.e. having the highest PSNR value and the lowest MSE. A very high PSNR value indicates that the median filter produces an image with excellent clarity and minimal distortion, while an MSE value close to zero indicates that the difference between the original image and the noise cancelling image is very small. In contrast, hybrid filters perform the worst with the highest MSE values and the lowest PSNR, which means that denoise images have large deviations from the original image as well as low visual quality. Meanwhile, the clahe and weighted_mean filters showed moderate performance with relatively high SSIM values, but still below the median in terms of PSNR and MSEs. Overall, this graph shows that median filters are the most effective noise cancellation method in maintaining image structure and quality, whereas the use of hybrid filters is less recommended for the facial image cases in this study.

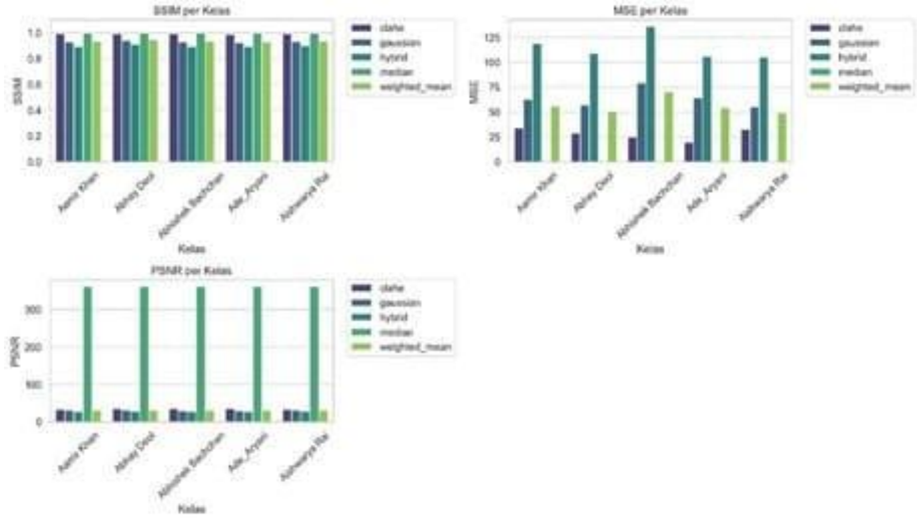


Figure 10 Measurements in the Classroom

Figure 10 Measurements on Classes is reinforced with a comparison graph of metrics and observations of five face classes (such as Aamir Khan, Abhay Deol, and Vivi Uvaina), which show a similar pattern in terms of median Filter advantage. Thus, the median filter proves to be the most effective method in suppressing noise while maintaining the visual quality and structure of the face image. The value of each face measurement can be seen in table 3.

Table 3 Average denoising in class

Class	**Denoise Average Measurement**			
	Filter	**Yes**	**MSE**	**PSNR**
Aamir Khan	clahe	0.991025	33.918026	33.755094
	Gaussian	0.928634	62.851207	30.860821
	Hybrid	0.893522	119.108817	27.686588

	median	1	0	361.201999
	weighted_mean	0.934731	56.123324	31.3663
Abhay Deol	clahe	0.991911	28.898423	34.814995
	Gaussian	0.94046	57.048588	31.431291
	Hybrid	0.909435	108.757443	28.104969
	median	1	0	361.201999
	weighted_mean	0.945839	50.582188	31.962781
………………	*Clahe*	……..	…….	……
	Weighted Average	……..	…….	……
	Gaussian	……..	…….	……
	Median	……..	…….	……
	Hybrid	……..	…….	……
Vivi_Uyainah	clahe	0.987668	16.803105	35.900834
	Gaussian	0.948402	41.764083	32.080079
	Hybrid	0.928419	66.727898	29.982435
	median	1	0	361.201999
	weighted_mean	0.953972	35.087025	32.833916

In the Pre-Pocessing Data method, the hybrid filtering method is smaller than the other filters, although the facial features shown in the SSIM measurement are still 90%, and the filtering data can be continued in the recognition process using the InceptionV3 architecture

3.3. CNN Model Implementation

The implementation stage of the inceptionV3 model is able to produce a representation of 2048 features per image. The model is trained for 10 epoches with a data-sharing scheme of 80% training and 20% testing

3.4. Evaluation of Results

. Evaluations are carried out based on classification: accuracy, precision, recall and F1-Score.

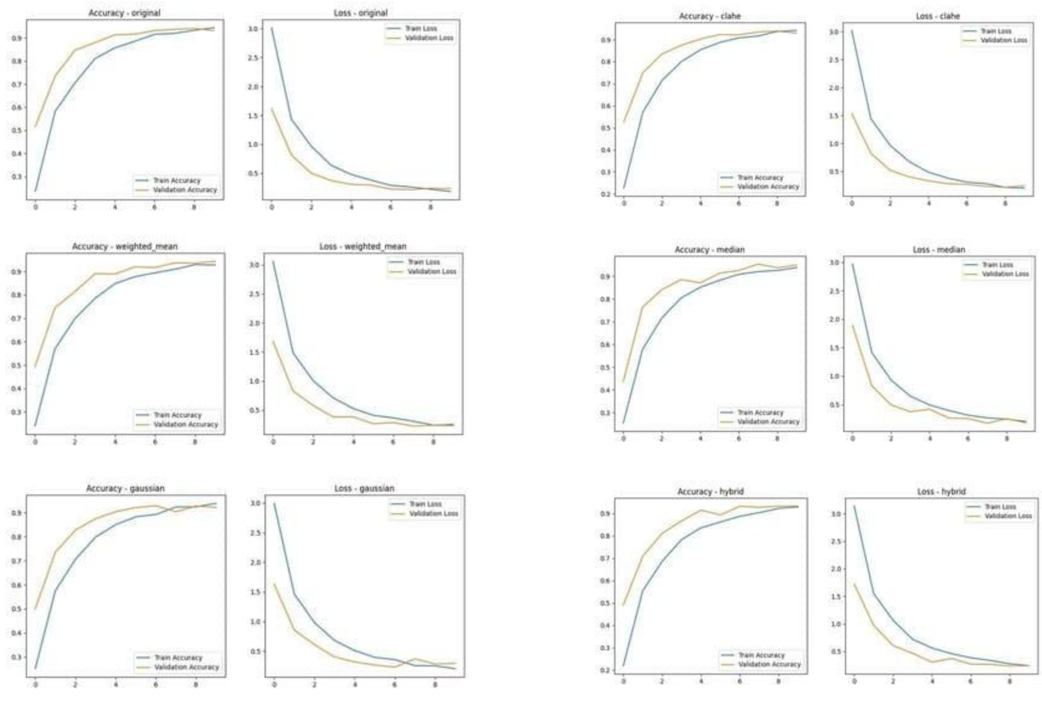
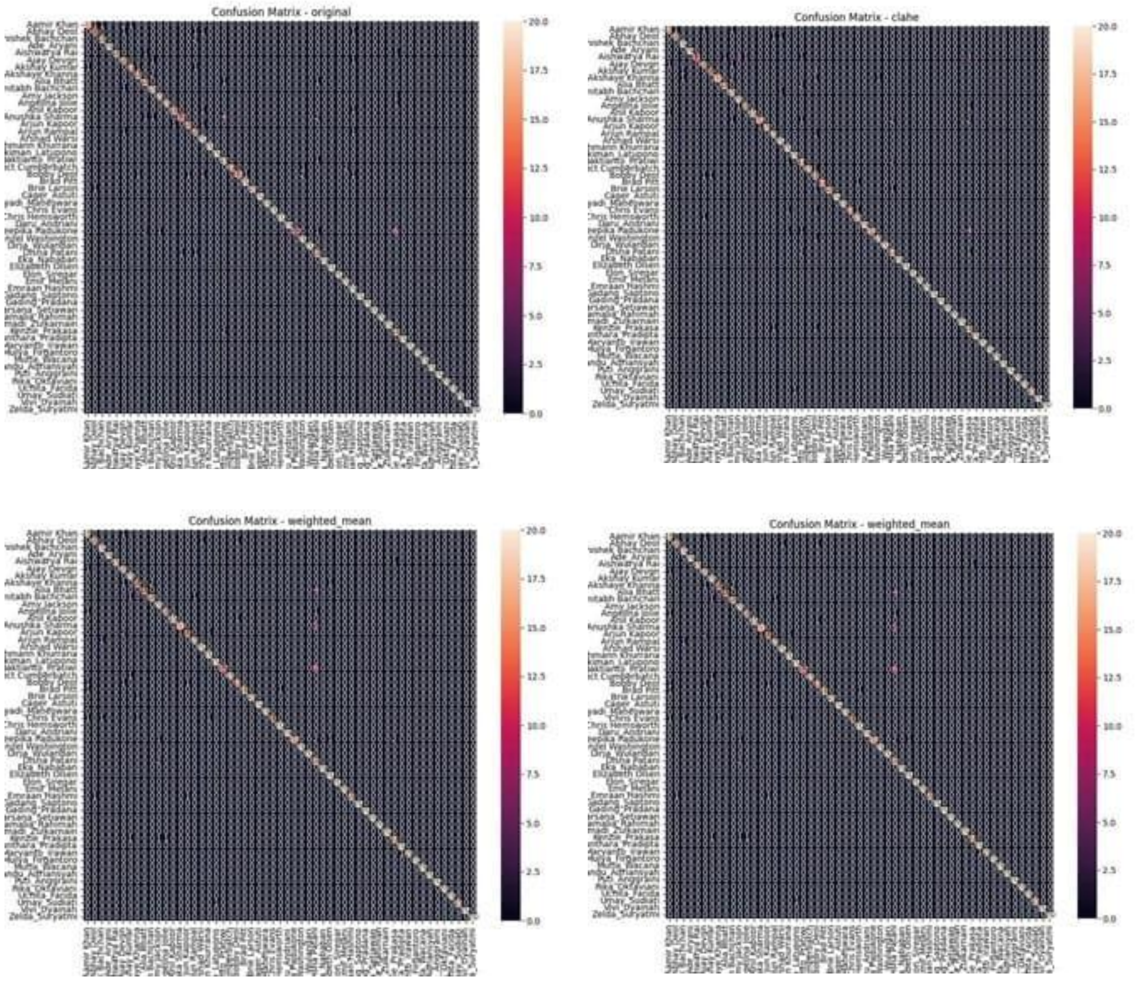


Figure 11 Training results

Figure 11 From the training results, the model with hybrid pre-processing produced the highest validation accuracy of 94.5%, followed by the weighted mean (94.3%), median (94.2%), CLAHE (94.1%), and Gaussian (94.0%) methods, while the model without preprocessing (original) obtained 93.1%. The difference in the highest accuracy increase compared to the base (original) model was +1.4%, which was achieved by the hybrid method. In addition to accuracy, the hybrid model also recorded a precision value of 94.2% and a recall of 94.6%, higher than the baseline which only recorded a precision of 92.8% and a 93.2% recall. When compared to the study by [9], which used CNN with an accuracy of only 74% on 5,000 images, the achievement of this model was significantly higher, even close to the performance of the VGG-Face model in the same study (97%). The study by [10], which used 400 images with a facial recognition accuracy of 98.57%, did note higher performance, but used a manually labeled dataset with a balanced distribution per class, in contrast to this study which still obtained an accuracy above 94% with preprocessing as the main variable. Meanwhile, a study by [11] showed that the FaceNet model obtained 97.48% accuracy on static data, but dropped drastically on testing the condition of accessories (masks: 70%, glasses: 83%). In this context, the results of the training model in this study showed stable performance during the training process (train loss and validation loss decreased uniformly), and did not experience significant overfitting—in contrast to some previous studies that did not explicitly display the training curve. Therefore, judging from the results of the initial training and evaluation, the model applied in this study was able to show high performance consistency, even with a simpler architecture and dataset compared to the previous more complex research.



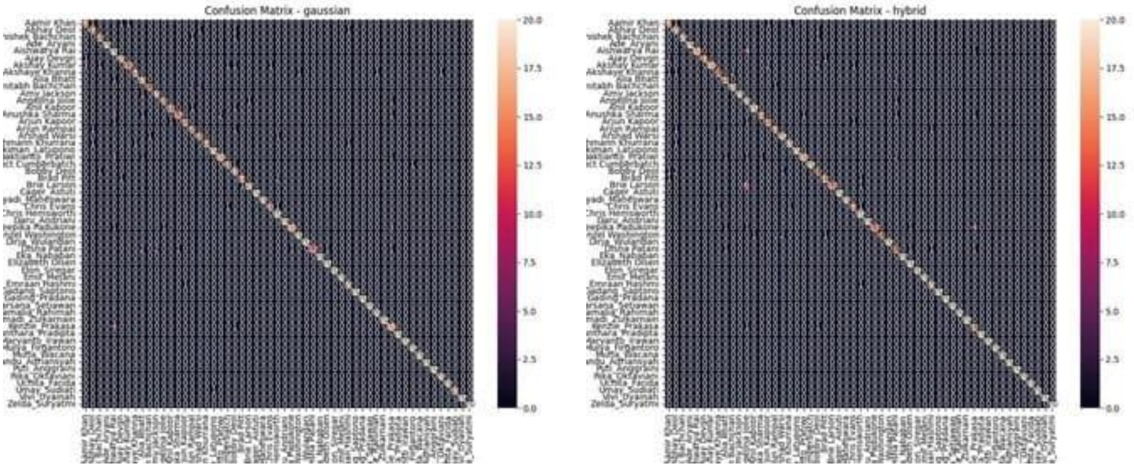


Figure 12 Confusion Metrics

The confusion matrix image in Figure 12 represents the performance of the face classification model against several image pre-processing techniques, namely original, CLAHE (Contrast Limited Adaptive Histogram Equalization), gaussian filter, median filter, hybrid method, and weighted average. Each row in the matrix shows the actual number of predictions for each class, while the columns show the predictions given by the model. From the results of visual observations, the CLAHE, hybrid, and Gaussian methods showed a relatively high and stable number of correct predictions (diagonal values), suggesting that these techniques were able to store information about important facial features that contributed to classification. In contrast, the original and weighted average methods show a wider distribution of classification errors beyond the diagonal, especially in some classes such as "Aishwarya Rai" and "Chris Hemsworth", which show feature ambiguity due to the lack of contrast enhancement or residual noise. Overall, the CLAHE and hybrid methods appear to provide more accurate classification results than the other methods, suggesting that local improvements in image contrast play an important role in improving the performance of facial recognition systems.

4. Conclusion

Based on the results of the research that has been conducted, it can be concluded that the application of image pre-processing techniques using hybrid screening methods consisting of Clahe, gaussian, median, and weighted mean has been proven to be effective in improving the visual quality of facial images and the performance of Convolutional Neural Network (CNN)-based classification systems. Tests using SSIM, MSE, and PSNR metrics show that the median filter has the most optimal ability to remove noise without damaging the original image structure. However, the results of the classification performance evaluation using a confusion matrix showed that the Clahe and hybrid methods were able to provide higher accuracy compared to other methods, including the original unfiltered image. This suggests that local contrast enhancement and the combination of multiple screening techniques can amplify important facial features that are influential in the classification process. The implementation of the CNN InceptionV3 architecture with proper pre-processing support results in validation performance above 94%, which reflects the model's good generalization capabilities. Therefore, the integration of hybrid filtering-based pre-processing can be recommended as a strategic approach in the development of reliable facial recognition systems, especially on datasets that have complex variations in noise and image quality. In this study, we only discussed a few filters that are hybrid in nature as a comparative material for the measurement of the input face image before being processed into an introductory model, some limitations here for the primary face data to be taken only at the front of the face. Future research suggestions are to try different hybrid filter methods again and get better results in terms of measurements, increasingly diverse datasets, and better facial recognition architecture methods.

Reference

[1] D. Setiawan, S. Widodo, T. Ridwan, and R. Ambari, "Human Emotion Detection Design Based on Facial Expressions Using the Vgg16 Algorithm," Vol. 11, No. 01, pp. 1–11, 2022.

[2] P. Musa, W.K. Anam, and SB Musa, "Deep Learning of Human Facial Expression Classifiers with Xception Architecture Model on Convolutional Neural Network Methods," Vol. 16, No. 1, pp. 66–73, 2023.
[3] M. Fansyuri, "Gender Classification Based on Facial Image Using Neural Network Algorithms," Vol. 2, No. 4, pp. 3784–3795, 2024.
[4] M. Adrezo, M. E. Ardiansyah, and PS Informatics, "Gender Detection Based on Face Using the Yolov8 1,2 Method," Vol. 7, No. 2019, pp. 1757–1762, 2024.
[5] M. A. Aziz, R. Wulanningrum, and D. Swanjaya, "Hand Movements Using the Histogram Equalization Method with an Adaptive Histogram Comparative Study of Improving the Image Quality of Hand Movements Using the Histogram Equalization Method," Vol. 6, No. 2, pp. 161–170, 2021.
[6] J. Zeng, X. Qiu, and S. Shi, "The Effect of Image Processing on Deep Facial Recognition Systems," vol. 18, January No., pp. 1187–1200, 2021, doi: 10.3934/mbe.2021064.
[7] R. Mardhiyyah, "Facial Landmark-Based Detection," Vol. 5, No. 2, pp. 144–148, 2021.
[8] Y. K. Bintang and H. Imaduddin, "Development of a Deep Learning Model for the Detection of Diabetic Retinopathy Using Transfer Learning Methods," Vol. 9, No. 3, pp. 1442–1455, 2024.
[9] I. Maulana, N. Khairunisa, and R. Mufidah, "Facial Shape Detection Using Convolutional Neural Network (CNN)," vol. 7, no. 6, pp. 3348–3355, 2023.
[10] F. H. Laia *et al.*, "AI-Based Facial Recognition Detection," Vol. 15, No. 1, pp. 62–72, 2023.
[11] W. Sugeng and D. Barus, "Examining Passport Photos Using the Dnn and Facenet Methods as Facial Recognition," pp. 169–180, 2023.
[12] I. P. Sari, F. Ramadhani, A. Satria, and D. Apdilah, "Implementation of Digital Image Processing in Facial Recognition Use," 2023.
[13] T. D. A. R. Putra, Y. F. Riti, dan D. Anggelia, "Noise Image Enchancement in Ultrasound Imagery," vol. 28, hlm. 149–160, 2023.
[14] L. Maximillian, Y. F. Riti, M. Anugraha, and Y. J. Palis, "Comparison of Sobel and Canny Algorithms for Detection of Aloe Vera Leaf Image Edges," Vol. 12, No. 2, 2023.
[15] H. A. N. Xiao, B. Guo, H. Zhang, and C. Li, "Parallel Algorithm of Image Average Filtering Based on Opencl," pp. 65001–65016, 2021, doi: 10.1109/access.2021.3068772.
[16] Amru, Yasir, Welnof, Satria, Putri, and Yuanda, "Digital Image Processing Median Screening Methods and Deep Aperture Morphology," vol. 17, pp. 1687–1701, 2023.
[17] S. H. Majeed, N. O. R. Ashidi, and M. A. T. Isa, "Repeated Adaptive Entropy Clip Limit Histogram Equalization for Poor Contrast Images," No. LCD, 2020, doi: 10.1109/access.2020.3014453.
[18] A. Yulistia, G. Aulia, H. Al, M. Murini, and N. Intan, "Evaluation of High-Pass and Low-Pass Filter Performance on High-Resolution Images Using Matlab," Vol. 02, No. 02, pp. 190–196, 2024.
[19] M. Noor, A. Azam, and C. Darujati, "Introduction to Frontal Facial Imagery Using Hirarikal