Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

Face Gender Classification for Public Facility Access Control using EfficientNet with Penalized-Entropy Loss

Sabrina Adinda Sari*[1], Faidhil Nugrah Ramadhan Ahmad[2], Miftahul Adnan Rasyid[3], I Gede Manggala Putra[4], Fauzan Ramadhan[5]

[1,2,3,4,5]School of Computing, Telkom University, Indonesia

Email: [1]sabrinaas@telkomuniversity.ac.id

Received: Jan 29, 2026; Revised: Feb 14, 2026; Accepted: Feb 21, 2026; Published: Jun 15, 2026

Abstract

Access to public facilities that are restricted based on gender, such as toilets and changing rooms, requires a strict security system because there are still many cases of abuse by irresponsible parties if only gender signs are relied upon. CCTV integrated with facial recognition is becoming more sophisticated every day, but it is limited if the face is covered by attributes such as masks. This is because the less visible the area is, the more difficult it is for the model to determine the label. To overcome this, this study proposes a gender classification approach for faces that may be covered by accessories such as masks, by adding Penalized Entropy loss as a loss function to the EfficientNet-B0 model. This loss function adds a penalty for incorrect predictions even if they are fairly accurate. The evaluation results show that the proposed model, with a penalty weight of 0.5, improved the accuracy by 3% from 90% to 93%. The experimental results show that the determination of the penalty weight has a significant impact on model performance, where a weight of 0.5 produces optimal performance because it provides a balance between penalizing overconfident predictions and the model's ability to maintain relevant feature discrimination; too small a weight does not sufficiently suppress overconfidence, while too large a weight actually reduces classification ability. The proposed method has demonstrated improvements in generalization and reduced overconfidence in gender classification systems. This method contributes to the development of reliable biometric systems suitable for uncontrolled real-world environments.

Keywords: *CCTV recordings, EfficientNet-B0, face mask, gender classification, overconfidence, Penalized Entropy Loss.*

1. INTRODUCTION

Privacy and security in public areas are becoming increasingly important issues in modern society. Toilets, changing rooms, nursing rooms, and other similar spaces are primarily designed to promote users' safety and comfort, particularly for women and other vulnerable populations. However, it is not uncommon for facilities that are supposed to provide protection to be misused. The most common example is through the installation of hidden cameras or recording devices without authorization [1][2][3][4]. Such incidents constitute violations of legal provisions and ethical norms and have the high potential to cause deep fear and significant psychological pressure on the victims.

One effort to maintain the safety and comfort of users when they are in public facilities related to gender is to create a gender-based access restriction system. For example, public toilet areas are separated for men and women with visual markers to provide more comfort for each visitor [5]. Unfortunately, this traditional system can still be abused by irresponsible parties [5].

With the development of technology, particularly in the fields of artificial intelligence and computer vision, visual identification-based access control systems have begun to be widely developed

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

[5]. Facial recognition technology has become one of the most popular methods in security systems [6] due to its ability to identify individuals quickly and with relatively high accuracy.

As a biometric method, facial recognition technology works by utilizing computer software to process and analyze unique facial features as a source of data [7][8][5]. Computer vision applications for facial image processing have been widely used, including individual identification [9][10][11], person re-identification [12][13], facial expression recognition [14][15][16], and age and emotion prediction [17][18][19]. Meanwhile, face-based gender classification systems have been widely developed [20][5][21], one of which is by Avishek et al. [22] using a conventional Convolutional Neural Network (CNN) architecture. This model showed fairly high accuracy in limited testing scenarios. However, the dataset used was still relatively simple, front-view facial images without obstructing attributes such as masks or other accessories, and thus failed to represent real-world conditions.

Additionally, the research by Jawad et al. also conducted face gender recognition using HOG and SVM, achieving the highest accuracy of 98.75% [23]. Although the accuracy is very high, this study still has limitations in terms of the dataset, which is taken from aligned frontal views. If implemented in CCTV, the results would drop significantly. Furthermore, the use of the HOG method is less robust against occlusion and pose variation [24]. This is because HOG only captures edge features and local gradient orientation, so if the face image is mostly covered by attributes like a mask, the feature information becomes less, thereby reducing accuracy.

In practice, facial images used for access control in public spaces vary greatly. Differences in shooting angles, lighting, camera quality, and the presence of face coverings can affect system performance. These conditions have the potential to reduce model accuracy, especially when the system is deployed in uncontrolled real-world environments.

The reality on the ground is increasingly complex. Since the COVID-19 pandemic, wearing masks has become the norm in a variety of daily activities [25][26]. In addition to masks, accessories such as glasses, and other objects are commonly worn in public places. The presence of these varied accessories reduces the amount of visual data that can be analyzed, lowering the accuracy of conventional face-based classification systems.

Another challenge to consider is the limitation of the availability of public datasets that specifically describe faces with various covering attributes. Most facial recognition and gender classification datasets only contain clear images of unmasked faces. As a result, models trained based on these datasets tend to be less capable of generalizing when faced with real-world conditions with occlusion [5]. This situation contributes to the low accuracy of gender classification systems in CCTV recordings in public facilities.

This study utilizes a self-constructed dataset specifically customized for imitating real-word settings. The dataset contains a variety facial images, some of which are covered by masks and glasses, and others of which are not. The model can learn to determine the difference between men and women by looking at this dataset, even if part of the face is not entirely visible. EfficientNet was chosen as the fundamental architecture due to its advantages in combining parameter efficiency with high accuracy and low computation cost compared to other CNN models [27][28]. This model has a lightweight architecture [27][28], making it ideal for use in real-time CCTV surveillance systems.

To address these limitations, this research suggests a modification to the loss function by implementing a Penalized Entropy Loss function that combines cross entropy loss with Wrong & Certain Regularization ($Loss_{WC-REG}$) to improve the performance of the EfficientNet-B0 model in addressing challenging samples caused by occlusion. In general, $Loss_{WC-REG}$ is designed to impose a greater penalty on incorrect predictions, but with a high level of confidence [29][30][31]. Meanwhile, the addition of cross-entropy-based is expected to stabilize the model's output probability distribution, resulting in more consistent and accurate classification performance [32].

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

This research presents three key contributions:

- It proposes a combination of the penalized-entropy loss function with EfficientNetB0 to improve the performance of gender classification models, particularly for occluded facial areas.
- This research created its own dataset taken from CCTV recordings in which individuals captured by the camera might be using facial coverings such as masks, glasses, and others, making it resemble real-life situations.
- This research designs a robust and lightweight model to be applied to CCTV in real-time, making it very suitable as a gender-based access control system for public facilities such as toilets, changing rooms, and others.

Therefore, this research aims to build a gender classification system that combines EfficientNet-B0 with Penalized Entropy Loss to improve the performance and confidence of the model, especially for datasets where facial images taken from uncontrolled CCTV are obstructed by facial barrier attributes such as masks and others.

The rest of this paper is organized as follows: Section 2 outlines the method; Section 3 shows the results; Section 4 presents the discussions; and Section 5 provides the conclusions.

2. METHOD

The proposed method aims to identify the gender of a person who is about to enter a specific public service facility environment from an image or CCTV recording, as shown in Figure 1.

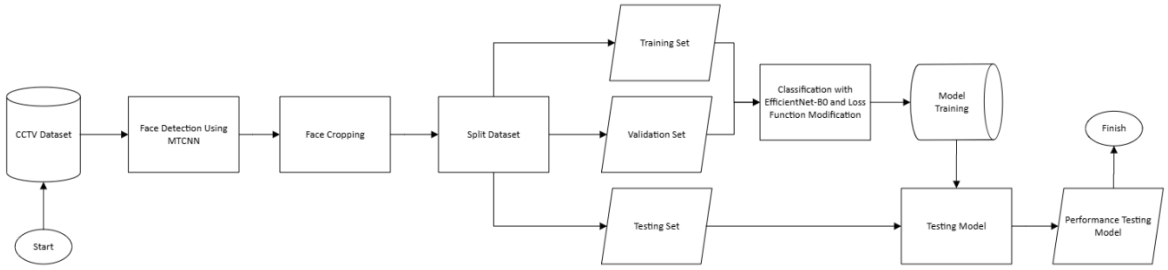

Figure 1. Flowchart Design System

2.1. Dataset

The dataset used is a self-installed CCTV dataset. It is designed to represent actual conditions in public service facilities. Each individual recorded by the CCTV wears face-obscuring attributes such as masks, hats, glasses, or a combination thereof, and some wear no obstructions at all. The dataset consists of 81 videos classified as men and women.

2.2. Face Detection Using MTCNN and Face Cropping

MTCNN constructs picture pyramids by resizing images to multiple dimensions [8][33][34]. The three-stage synthesized structure employs this visual pyramid as an input in three separate methods [35]. The initial phase involves generating a boundary box regression vector for each prospective face window with the fully convolutional Proposal Network (P-Net). The candidate is then calibrated using the obtained boundaries box regressive vector. Subsequently, non-maximum suppression (NMS) is employed to consolidate several overlapping candidates. The second phase involves sending each candidate to a distinct CNN referred to as the Refine Network (R-Net), which performs Non-Maximum Suppression (NMS), calibration by bounding box regression, and the subsequent elimination of a substantial proportion of incorrect candidates. Third phase: This stage parallels the second, but aims to identify face areas necessitating greater scrutiny using the Output Network (O-Net). It generates the coordinates of five facial landmarks.

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

CCTV video recordings are converted into images and collected at a frequency of one frame per second. MTCNN is used for face detection. The input includes images that have been processed from CCTV recordings, and produces face position coordinates. A bounding box is created around a specific area on the face. Next, images containing the face bounding box are cropped to only cover the face area and resized to 224 x 224 x 3 pixels to be used as conditional input for the EfficientNet-B0 model.

2.3. Split Dataset

The CCTV dataset has been cropped and resized to 224 x 224 x 3 and divided into a train set, validation set, and test set with a ratio of 70%:10%:20%. Details of the image dataset for the entire process are in Table 1.

Table 1. The Proportion of the Training, Validation, and Testing Datasets

Split Dataset	Men	Women
Training (70%)	601	831
Validation (10%)	86	118
Testing (20%)	172	237

2.4. EfficientNet-B0

EfficientNet was first introduced by Tan and Le in their research [36]. They stated that EfficientNet is one of the most efficient models, capable of achieving high accuracy on ImageNet and transfer learning for image recognition [36]. EfficientNet is a convolutional neural network architecture optimized through a consistent scaling process across three main dimensions: depth, width, and resolution. EfficientNet has eight variants, from EfficientNet-B0 to EfficientNet-B7. Increased block usage results in a greater number of parameters generated, which simultaneously enhances accuracy. Consequently, the higher the variant, the greater the computational complexity and inference time required.

EfficientNet-B0 was chosen as the EfficientNet family's base model because it offers the best balance of performance and efficiency [27][28]. This model uses a hybrid scalability strategy that balances depth, width, and resolution, resulting in high accuracy with a relatively small number of parameters and minimal computational cost [27]. Furthermore, the adoption of the Mobile Inverted Bottleneck Convolution (MB-Conv) layer, as well as squeeze-and-excitation optimization, allows for enhanced feature representation while reducing computational cost [28][37].

Compared to other variants, efficientnetb0 has the fewest parameters [27] and is computationally lightweight, making it highly suitable for integration with real-time CCTV. In the context of CCTV systems that require fast and stable inference processes, computational efficiency is a primary consideration so the system can run near real-time without requiring high-spec hardware.

Table 2. Model Configuration and Computational Environment

Component	Specification
Batch Size	32
Learning rate	0.001
Epoch	50
Optimizer	Adam
Platform	Google Colab Pro
GPU	NVIDIA Tesla T4
RAM	± 25 GB

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

Although mobilenet is also a lightweight CNN model, efficientnet uses a more structured compound scaling strategy, making it superior in capturing feature representations [38]. Therefore, efficientnetb0 is a highly appropriate choice due to its advantages of computational efficiency and reliability in capturing more structured features, making it highly suitable for integration with CCTV.

The computational specifications used for all experiments are listed in Table 2. A batch size of 32 was chosen because it is computationally efficient and fits the GPU memory used in the training process. The Adam optimizer was chosen because of its two advantages: memory efficiency and lower processing resource requirements compared to other conventional optimizers [39]. This results in a faster and more stable training process.

2.5. Penalized Entropy Loss Function

The primary issue in this gender classification research is that images of the face lose significant features in the mouth and nose regions because of obscuring objects such as mask. This leads to the model frequently mispredicting gender. Standard loss functions such as categorical cross-entropy ignore prediction uncertainty, even though this aspect is crucial in the case of masked faces. Therefore, this study adopts and adapts penalized-entropy-loss to estimate uncertainty in CCTV-based masked face gender classification.

- **Entropy Measurement as an Estimate of Uncertainty**

  Suppose the training dataset is defined as $Data_{train} = \{X, Y\} = \{x_i, y_i\}_{i=1}^{N}$, where x_i is the i^{th} face image, $y_i \in \{1, K\}$ is the class label (1 = Women, 2 = Men), N is the total number of training data. Additionally, the testing dataset is expressed $Data_{test} = \{X^*, Y^*\} = \{x_i^*, y_i^*\}_{i=1}^{M}$ [30]. The softmax function in Equation (1) produces a probability distribution for each class in the EfficientNet-B0.

$$p(\hat{y} = K|x^*, X, Y) = \frac{1}{T}\sum_{t=1}^{T} p(\hat{y} = K|x^*, w_t) \qquad (1)$$

  In stochastic forward passes T, $\hat{y}$ represents the predicted class and w_t represents the set of model parameters in t^{th} phase. Shannon Entropy, as defined in Equation (2), is used for calculating the amount of uncertainty in the predicted outcome.

$$E(x^*) = -\sum_K p(\hat{y} = k|x^*) \cdot \log p(\hat{y} = k|x^*) \qquad (2)$$

  In this case, the number of classes K=2. Entropy is high when the model is uncertain (for example, the probability is close to 0.5: 0.5), and low when the model is very confident (for example, 0.9: 0.1). Thus, entropy is used as an indicator of prediction uncertainty in each face image.

- **Cross-Entropy Loss**

  Cross-Entropy loss is used to make sure that the model continues learning to classify correctly [30], which is stated as Equation (3).

$$Loss_{CE} = -\frac{1}{N}\sum_{i=1}^{N} \log p(\hat{y} = y|x_i) \qquad (3)$$

  This function aims to calculate how far the model's prediction deviate from the expected labels, commonly known as the loss function. The better model is at classifying images, the lower value of loss function. Unfortunately, this loss function does not address confidence level information.

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

- **Wrong & Certain Regularization (WC-reg)**
  To add an element of uncertainty to the learning process, a regularization term called Wrong & Certain Regularization (WC-reg) is introduced [30]. First, the certainty value for each sample is calculated as $C(x_i) = 1 - \frac{E(x_i)}{K}$. Then, the regularization function is defined as Equation (4).

$$Loss_{WC-REG} = \frac{1}{N}\sum_{i=1}^{N}[-\log p(\hat{y} = y|x_i)] \cdot C(x_i) \quad (4)$$

  This function penalized false predictions provided with high confidence. Conversely, if the model is inaccurate but has a low confidence level (indicating that the model is unclear), the penalty is reduced. This is extremely consistent with the features of veiled faces, which are naturally more ambiguous.
- **Penalized Entropy Loss**
  The final loss function (Penalized Entropy Loss) used in this study is a combination of two components as shown in Equation (5).

$$Loss_{PE} = Loss_{CE} + \delta \cdot Loss_{WE-REG} \quad (5)$$

  In this study, the δ value is used to regulate the extent of uncertainty's influence on the model's learning process [30]. The combination of penalized entropy loss as a loss function replacing cross entropy loss in the EfficientNet-B0 model really helps the model to predict more accurately, be more validated, dan be more adaptable to situations when the dataset experiences occlusion (some important features on the face are covered by attributes such as masks and glasses.

2.6. Performance Testing Model

Performance testing model includes scenario testing and hyperparameter modification to confirm experimental outcomes. Analyzing experimental outcomes will inform the selected hyperparameter for optimal performance in future experiments. Confusion matrix evaluation is used in this study. Confusion matrix measures classification results with real value [36]. Accuracy is assesses the frequency of true predictions made by the model overall, as described in equation (6). Precision is defined as the ratio of True Positives (TP) to the total number of instances predicted as positive, as seen in equation (7). Recall is the proportion of True Positives (TP) to the total number of actual positive data, as expressed in equation (8). The F1-Score represents the harmonious average of accuracy and recall, as seen in equation (9).

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$F1 - Score = 2 \times \left(\frac{Precision \times Recall}{Precision+Recall}\right) \quad (9)$$

3. RESULT

In this chapter, we perform each step according to the system design flowchart in Figure 1 and conduct experiments to determine the optimal penalty weight (delta) value.

The first stage involves obtaining a dataset from CCTV footage. The dataset is then processed using MTCNN to detect facial regions, followed by a cropping process to capture only the facial region, so that only the facial area becomes the focus of the model input, as shown in Figure 2.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

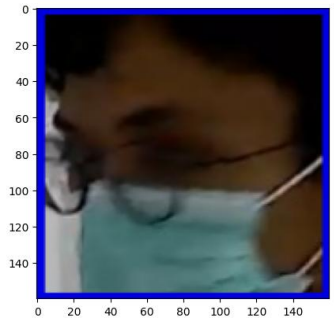

Figure 2. Example of an image with the face cropped

The dataset is then divided into a training set, a validation set, and a testing set. The training and validation data are used in the training stage of the EfficientNet-B0 model with a modified loss function, while the testing data is used in the final performance testing stage of the model. Figure 3 shows an example of the output of the created gender classification system

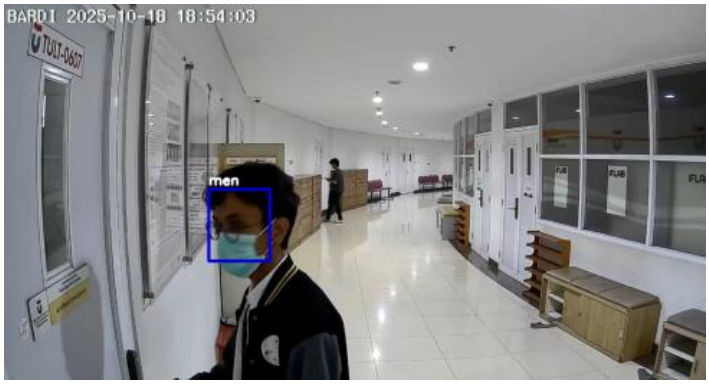

Figure 3. The example of the gender classification system's outcomes

Nine experiments were conducted to analyze the hyperparameter δ, or entropy weight, on the performance of a gender classification model on masked facial images. The hyperparameter δ serves as an entropy penalty weight that suppresses prediction errors with a high degree of confidence. The nine experiments included a baseline and various δ values (0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.7, and 0.8).

To measure model performance, the experiments were divided into two parts. The first section examines the training process, considering accuracy, loss values, and curve stability from the training and validation processes. The second section focuses on the testing process, which yields performance indicators such as accuracy, precision, recall, and f1-score.

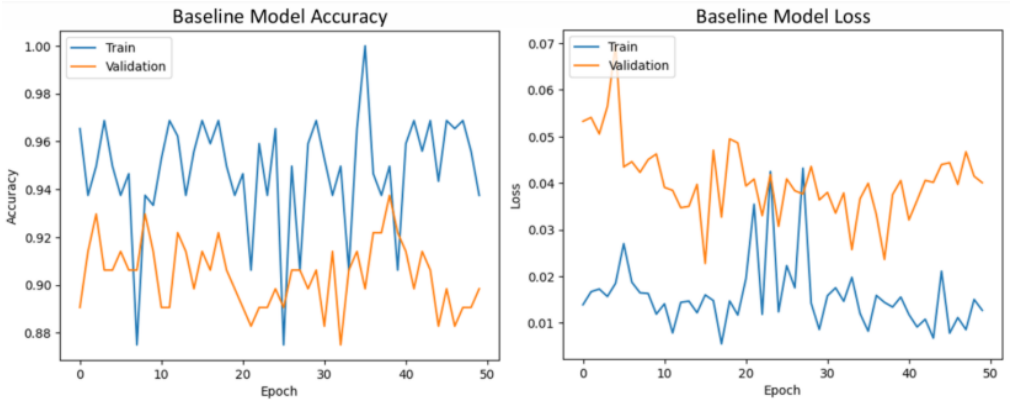

Figure 4. Performance Graph of the Model for the First Experiment

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

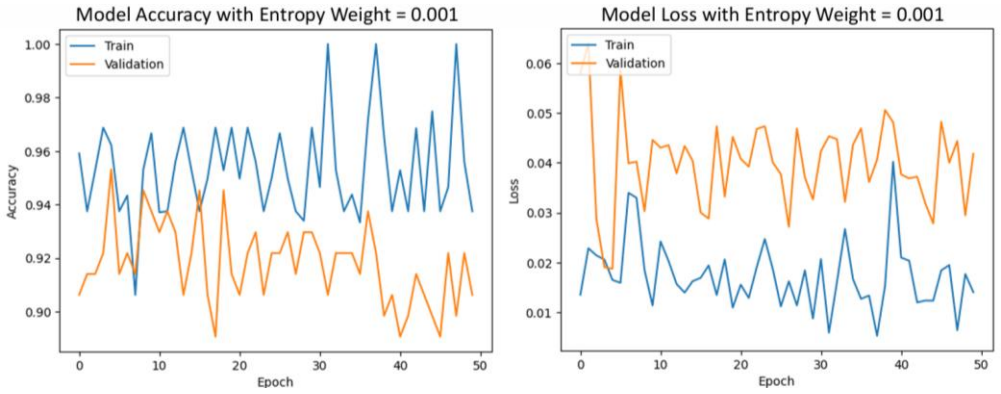

Figure 5. Performance Graph of the Model for the Second Experiment

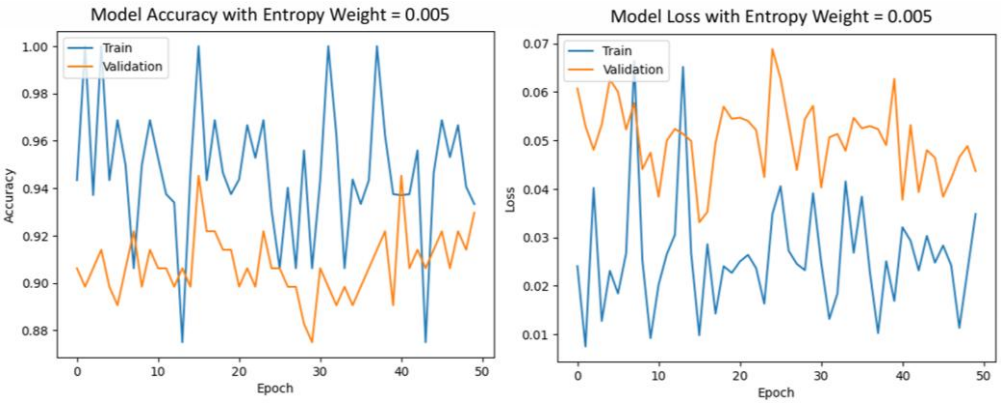

Figure 6. Performance Graph of the Model for the Third Experiment

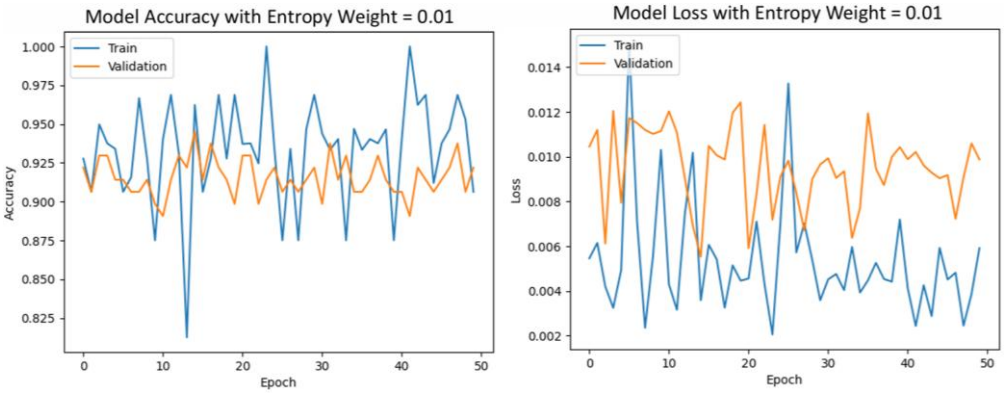

Figure 7. Performance Graph of the Model for the Fourth Experiment

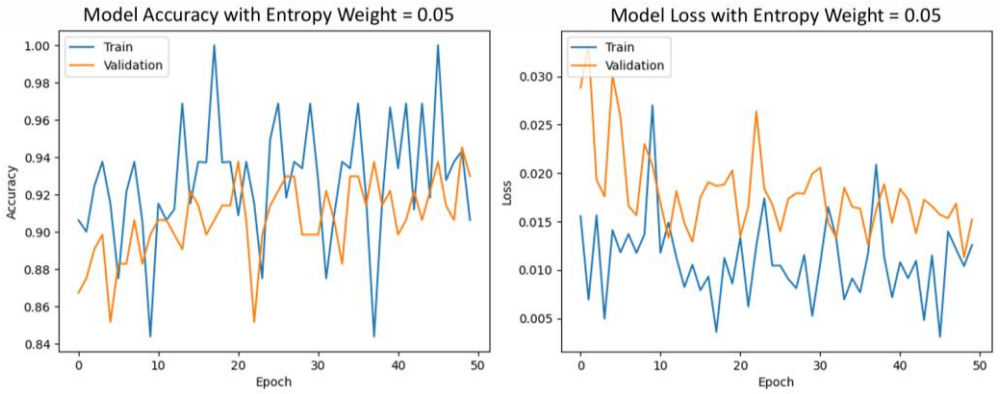

Figure 8. Performance Graph of the Model for the Fifth Experiment

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

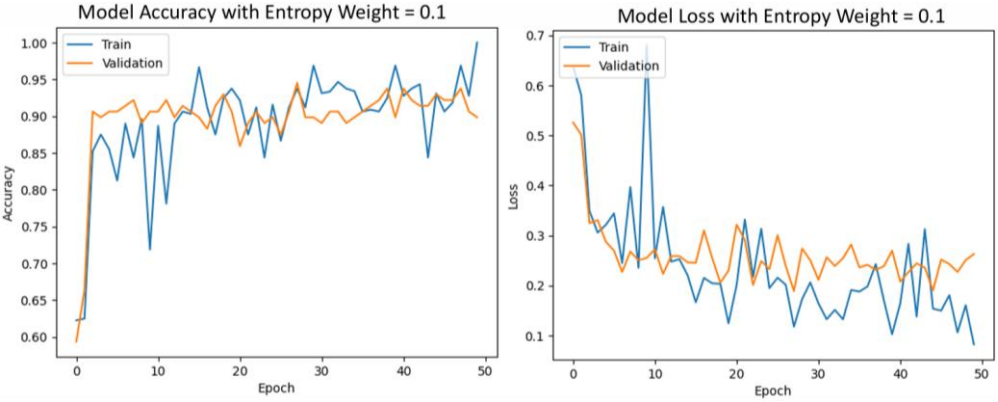

Figure 9. Performance Graph of the Model for the Sixth Experiment

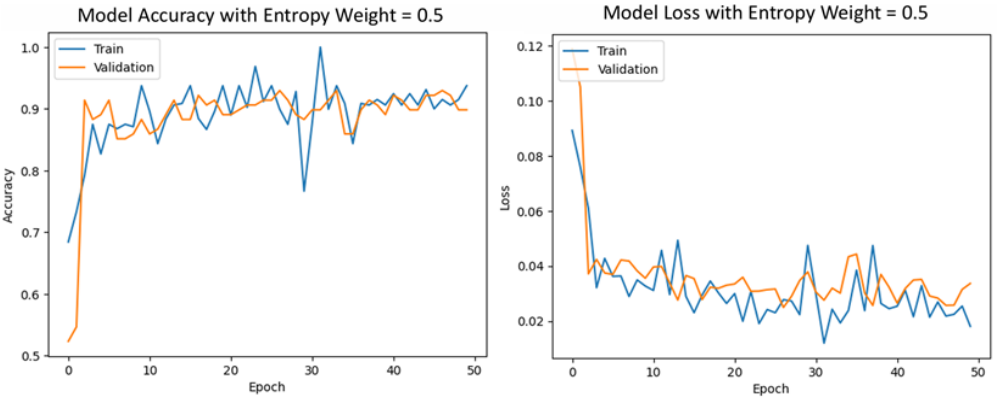

Figure 10. Performance Graph of the Model for the Seventh Experiment

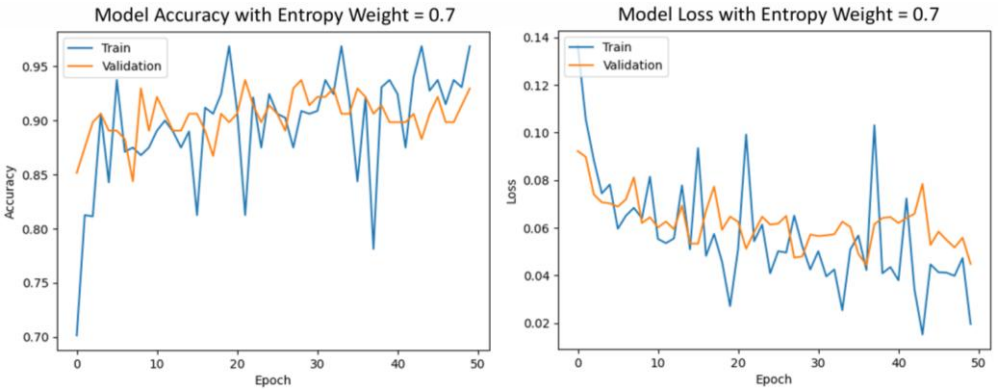

Figure 11. Performance Graph of the Model for the Eighth Experiment

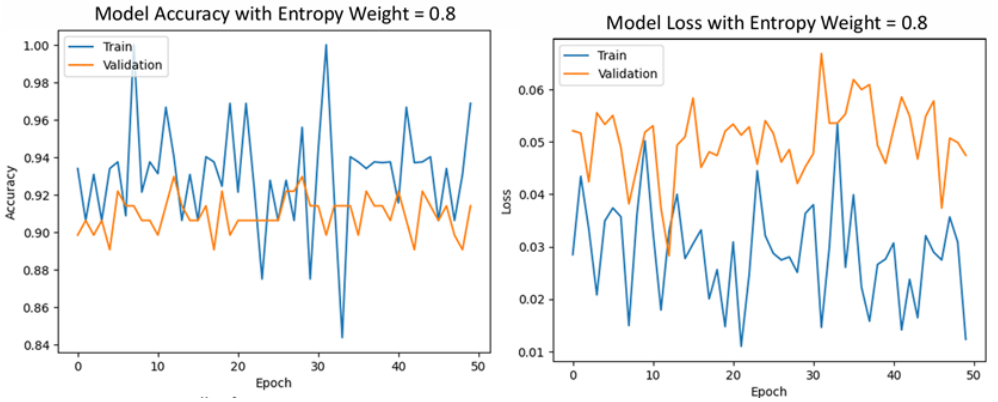

Figure 12. Performance Graph of the Model for the Ninth Experiment

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

Figures 4 to 12 show graphs of accuracy and loss during the training and validation processes for each experiment. In the baseline model, training accuracy increased rapidly, but there was a significant shift in the loss curve, indicating overfitting.

Meanwhile, experiments with very small entropy weight values (δ = 0.001 and δ = 0.005) showed better stability, although the difference between training and validation was still clearly visible. This indicates that very small entropy weight values are still ineffective. More specifically, at small δ values, a fluctuating validation loss phenomenon occurs, where the validation loss curve fluctuates unstably. This occurs because the penalty for incorrect but highly confident predictions is still too weak, resulting in the model still producing a sharp probability distribution (overconfidence). As a result, when the model encounters validation samples that differ slightly from the training data (for example, variations in masks or lighting), predictions become unstable and the loss fluctuates.

In experiments with intermediate entropy weight values (δ = 0.01 and δ = 0.05), the gap between the training and validation curves begins to narrow, but the validation loss remains unstable. This indicates that entropy regularization is beginning to work to control the model, but it is not yet optimal in balancing overconfidence control and discriminatory ability.

In experiments with an entropy weight of 0.1, the training and validation processes became more stable. This indicates that overfitting is starting to be suppressed, although accuracy tends to converge. However, performance has not yet reached its peak, as the model begins to lose clarity in class distinction.

A closer look at the experiment with delta = 0.5 reveals that the accuracy curves between training and validation are the closest compared to other experiments, and the loss decreases steadily and consistently. This indicates that the loss function is able to make the model more cautious in classifying classes by suppressing the model from being overconfident when predicting labels in occluded data. This is what causes the model to produce the best performance. At higher entropy weights (δ = 0.7 and δ = 0.8), model performance decreases. Training and validation accuracy are lower, and validation loss increases, indicating over-regularization. In this condition, the penalty is too strong, causing the probability distribution to become too flat (over-smoothing), and the model becomes too hesitant in making decisions.

Table 3. Performance Results for All Experiments

Experiment	Accuracy	Precision	Recall	F1-Score
Baseline Model	0.900	0.910	0.800	0.850
$Loss_{PE}(\delta = 0.001)$	0.900	0.900	0.900	0.900
$Loss_{PE}(\delta = 0.005)$	0.920	0.920	0.920	0.920
$Loss_{PE}(\delta = 0.01)$	0.910	0.910	0.910	0.910
$Loss_{PE}(\delta = 0.05)$	0.900	0.910	0.900	0.900
$Loss_{PE}(\delta = 0.1)$	0.910	0.910	0.910	0.910
$\boldsymbol{Loss_{PE}(\delta = 0.5)}$	**0.930**	**0.930**	**0.930**	**0.930**
$Loss_{PE}(\delta = 0.6)$	0.920	0.920	0.920	0.920
$Loss_{PE}(\delta = 0.8)$	0.910	0.910	0.910	0.910

Table 3. shows the performance results on the testing data. From the table, it can be seen that the highest performance was achieved when the entropy weight experiment = 0.5 with the same accuracy, precision, recall, and f1-score values of 0.93. Based on Figure 4 to 12 and Table 3, from all the experiments conducted, the best results were obtained when the entropy weight = 0.5.

Overall, these results show that performance improvement is not only determined by the accuracy rate alone, but also by the stability of the curve, the consistency of the validation loss, and the model's ability to control the level of prediction confidence in conditions of faces experiencing occlusion

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

4. DISCUSSIONS

In the discussion chapter, there are three main topics, namely error analysis, analysis of the impact of penalty weight value on model perfomance, and comparison with previous studies.

4.1. Error Analysis

In the model baseline for test cases, examples were found where several men faces were classified as women. This error falls under the category of false positives, which are predictions that do not match the actual labels.

During the baseline model testing experiment, several male faces were classified as female. This error falls into the false positive category, meaning predictions that do not match the actual label. Based on Figure 13, which presents the confusion matrix results, the number of errors in the "Male" class is higher than in the "Female" class.

Analytically, this can be explained by the characteristics of the visual features. Many samples were mispredicted because the face was covered by a mask and other attributes such as glasses or a hat, and because the lighting conditions were suboptimal. In these conditions, the area of the face that the model could explain was limited to the eyes, eyebrows, and forehead. Important information such as jaw structure, chin shape, and lower facial contours were not available due to the mask. This lower facial area is a strong morphological differentiator between the "men" and "women" genders.

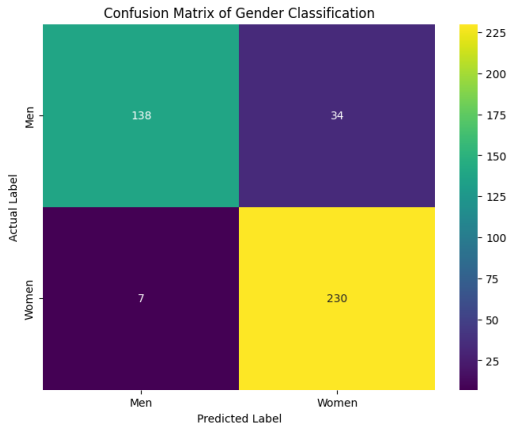

Figure 13. Confusion Matrix of Baseline Model

Due to this feature limitation, the distribution of male features under conditions where the face is covered becomes more similar to the distribution of female features. This shifts the prediction probability and increases the prediction error in the "Men" class. This error is not due to a weak model, but rather to the limited availability of important features.

Based on the results of the experiment in the previous subchapter, when the model did not use $Loss_{PE}$ (baseline model), quite a few predictions were incorrect, but the confidence level was high, as seen in Figure 13. This was most likely caused by overconfidence, meaning that the model was too confident in its class decisions even though they were wrong. However, with the application of entropy regularization, the confidence level for incorrect samples decreased. This means that the model became more cautious in making decisions when the available visual information was not clear enough.

In the context of implementation in specific public spaces, such as changing rooms or public toilets, controlling overconfidence is a very important aspect. The system is not only required to be accurate, but also to be able to indicate the level of uncertainty when visual conditions are ambiguous. Therefore, the results of this analysis confirm that performance improvement is not only measured by accuracy alone, but also by the stability and caution of the model in producing prediction able to indicate

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

the level of uncertainty when visual conditions are ambiguous. Therefore, the results of this analysis confirm that performance improvement is not only measured by accuracy alone, but also by the stability and caution of the model in producing predictions.

Furthermore, this study created a dataset of CCTV-recorded faces from various regions in Indonesia. This represents racial and ethnic diversity nationally, but not enough globally. Therefore, testing with international facial datasets is also necessary to determine the model's generalizability to race or ethnicity globally.

4.2. Analysis of the Impact of Penalty Weight Value on Model Perfomance

All experiments have been conducted by assigning different penalty weight values to determine the most optimal penalty weight. Based on Table 3, the best penalty weight is when delta = 0.5. If the delta value is too small, the penalty given is also small. When the model makes an incorrect prediction but remains confident in its prediction, this loss function only imposes a small penalty. As a result, the model remains overly confident even when its predictions are wrong. On the other hand, if the delta value is too large, the punishment given becomes excessively large as well. This causes the model to be too cautious in making predictions, with the probability values between each class not being too far apart, making it difficult for the model to distinguish whether a face belongs to "Men" or "Women," and the model's performance also decreases. The penalty weight value of 0.5 is a middle ground, the penalty given is appropriate and strong enough to suppress overconfidence without losing decisiveness in predictions. This is what caused its performance to be the best in the entire experiment.

4.3. Model Performance Comparison with Previous Research

This session discusses the comparison between the proposed model, baseline, and previous studies that also discuss face gender classification, in terms of the model, dataset performance results, and limitations of each model discussed in Table 3.

Table 4. Comparison between Previous Studies, Baseline, and the Proposed Model

No	Model	Dataset Used	Accuracy	Precision	Recall	F1-Score
1	Gated Residual Attention Network (GRA_Net) [22]	Adience dataset	0.814	-	-	-
2	MobileNet [21]	LFW dataset	0.929	-	-	-
3	CNN [18]	UTKFace dataset	0.865	-	-	-
4	Baseline (EfficientNet-B0)	Self-constructed from CCTV	0.900	0.910	0.800	0.850
5	Proposed Model (EfficientNet-B0 + penalized Entropy Loss)	Self-constructed from CCTV	**0.930**	**0.930**	**0.930**	**0.930**

Based on Table 4, previous studies on face gender recognition [22][21][18] reported fairly high accuracy rates, as shown in number 2, with the best accuracy of 0.929 achieved using the MobileNet model. However, the weakness of studies [22][21][18] is that they use public datasets taken from a front perspective without any face-obscuring attributes and with relatively stable lighting. These approaches do not fully represent the real challenges that would be encountered when implementing these models with CCTV-based datasets from public spaces.

In this study (No 4 and 5), we used primary datasets collected from CCTV recordings with scenarios designed to resemble field conditions. People captured on CCTV cameras wore various face-

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

obscuring attributes such as glasses and hats, with lighting and shots from various angles. This certainly increased the complexity of classification because the more areas of the face were covered, the less facial information was available to determine gender. In addition, it was found that the proposed method outperformed previous studies in terms of accuracy.

The proposed research is not only superior in terms of accuracy, but also superior in terms of a more realistic approach to implementation in public spaces. By using primary datasets that represent real-world conditions and integrating penalized entropy to control overconfidence, this research offers a more adaptive and relevant solution for CCTV-based security applications compared to previous approaches that were still limited to controlled environments.

This research has significant urgency and scientific impact in the field of computer science, particularly for classification based on biometrics such as facial recognition. This research demonstrates that improving model performance does not only depend on the architecture of the model used, but also on the appropriate loss function corresponding to the task, which is also very important. The use of penalized entropy loss shows that proper overconfidence regulation also affects the model's reliability in classifying based on biometrics. This research is also supported by results that represent that entropy-based regulation can enhance the stability of the model in terms of class prediction probability in conditions where the data experiences occlusion.

5. CONCLUSION

This study proposes a gender classification system based on efficientnetb0 with penalized entropy loss as a specific loss function to handle the issue of overconfidence in model predictions. The dataset was taken from CCTV with a scenario using facial obstruction attributes to resemble real-world conditions. Various experiments have been conducted to determine the best and most optimal penalty weight (delta), which is with a delta value of 0.5 that improves the accuracy of the initial model by 3%. Based on this, if delta is too small, it is not strong enough to control overconfidence. Meanwhile, when the delta value is too high, the model will experience a decrease in performance due to over-regularization, which means the model is too cautious in making decisions.

These results show that Penalized Entropy Loss is an effective solution for the problem of overconfidence in biometric classification models, especially on data with partial occlusion conditions. This research has limitations, as most scenarios were taken under relatively controlled conditions. This system has not yet been tested further for scenarios with very poor lighting, so it could be an opportunity for future research. Scientifically, this research contributes to the development of regulations for confidence calibration in deep learning. In addition, this loss function approach also enhances the robustness of biometric-based models in uncontrolled real-world environments.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest between the authors or with research object in this paper.

ACKNOWLEDGEMENT

All authors would like to express their sincere gratitude to the Directorate of Research and Community Service, for the generous financial support through the Internal Research Grant Scheme under Grant.

REFERENCES

[1] N. C. Weley, F. Y. P. Amboro, and T. D. Seroja, “Behind Closed Lenses: Analyzing the Efficacy of Personal Data Protection Laws in Combatting Hidden Cameras,” *J. Judic. Rev.*, vol. 26, no.

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

1, pp. 89–112, 2024, doi: 10.37253/jjr.v26i1.9158.
[2] S. Herodotou and F. Hao, "Spying on the Spy: Security Analysis of Hidden Cameras," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13983 LNCS, pp. 345–362, 2023, doi: 10.1007/978-3-031-39828-5_19.
[3] W. N. Hartzog and E. Selinger, "Scholarly Commons at Boston University School of Law Privacy Nicks : How the Law Normalizes Surveillance," 2024.
[4] Di. Dao, M. Salman, and Y. Noh, "DeepDeSpy: A Deep Learning-Based Wireless Spy Camera Detection System," *IEEE Access*, vol. 9, pp. 145486–145497, 2021, doi: 10.1109/ACCESS.2021.3121254.
[5] S. A. Sari and W. F. Al Maki, "Masked Face Images Based Gender Classification using Hybrid Bat Algorithm Optimized Bagging," *5th Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2023*, pp. 91–96, 2023, doi: 10.1109/ICAIIC57133.2023.10067008.
[6] S. Adinda and M. Dwi, "Modified Hybrid Pooling on FaceNet Embedding to Enhance a Surveillance-Based Suspect Detection System."
[7] E. Ahmad Khorsheed and Z. Ali Nayef, "Face Recognition Algorithms: A Review," *Acad. J. Nawroz Univ.*, vol. 11, no. 3, pp. 202–207, 2022, doi: 10.25007/ajnu.v11n3a1432.
[8] S. A. Sari and M. D. Sulistiyo, "Suspect Identification Based on Facial Recognition from CCTV Using Hybrid Bat Algorithm," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, pp. 205–210, 2024, doi: 10.1109/EECSI63442.2024.10776490.
[9] W. Ali, W. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, *Classical and modern face recognition approaches: a complete review*, vol. 80, no. 3. Multimedia Tools and Applications, 2021.
[10] S. Sandhya, A. Balasundaram, and A. Shaik, "Deep Learning Based Face Detection and Identification of Criminal Suspects," *Comput. Mater. Contin.*, vol. 74, no. 2, pp. 2331–2343, 2023, doi: 10.32604/cmc.2023.032715.
[11] H. Du, H. Shi, D. Zeng, X. P. Zhang, and T. Mei, "The Elements of End-to-end Deep Face Recognition: A Survey of Recent Advances," *ACM Comput. Surv.*, vol. 54, no. 10, 2022, doi: 10.1145/3507902.
[12] D. Fu *et al.*, "Unsupervised Pre-training for Person Re-identification," pp. 14750–14759.
[13] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, "Deep Learning for Person Re-Identification: A Survey and Outlook," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 2872–2893, 2022, doi: 10.1109/TPAMI.2021.3054775.
[14] K. Mohan, A. Seal, O. Krejcar, and A. Yazidi, "FER-net: facial expression recognition using deep neural net," *Neural Comput. Appl.*, vol. 33, no. 15, pp. 9125–9136, 2021, doi: 10.1007/s00521-020-05676-y.
[15] M. Sajjad *et al.*, "A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines," *Alexandria Eng. J.*, vol. 68, pp. 817–840, 2023, doi: 10.1016/j.aej.2023.01.017.
[16] J. Mao *et al.*, "POSTER++: A simpler and stronger facial expression recognition network," *Pattern Recognit.*, vol. 157, 2025, doi: 10.1016/j.patcog.2024.110951.
[17] A. Garain, B. Ray, P. K. Singh, A. Ahmadian, N. Senu, and R. Sarkar, "GRA_Net: A Deep Learning Model for Classification of Age and Gender from Facial Images," *IEEE Access*, vol. 9, pp. 85672–85689, 2021, doi: 10.1109/ACCESS.2021.3085971.
[18] S. Teja Chavali, C. Tej Kandavalli, T. M. Sugash, and R. Subramani, "Smart Facial Emotion Recognition with Gender and Age Factor Estimation," *Procedia Comput. Sci.*, vol. 218, no. 2022, pp. 113–123, 2022, doi: 10.1016/j.procs.2022.12.407.
[19] E. Kim, D. Bryant, D. Srikanth, and A. Howard, "Age Bias in Emotion Detection: An Analysis of Facial Emotion Recognition Performance on Young, Middle-Aged, and Older Adults," *AIES 2021 - Proc. 2021 AAAI/ACM Conf. AI, Ethics, Soc.*, pp. 638–644, 2021, doi: 10.1145/3461702.3462609.
[20] G. Gingin, M. Dwi, M. Arzaki, and E. Rachmawati, "Classifying Gender Based on Face Images Using Vision Transformer," vol. 8, no. March, pp. 18–25, 2024.
[21] M. Oulad-Kaddour, H. Haddadou, C. C. Vilda, D. Palacios-Alonso, K. Benatchba, and E. Cabello, "Deep Learning-Based Gender Classification by Training With Fake Data," *IEEE*

Vol. 7, No. 3, June 2026, Page. 2812-2826
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5675

Access, vol. 11, no. October, pp. 120766–120779, 2023, doi: 10.1109/ACCESS.2023.3328210.

[22] A. Garain, B. Ray, P. K. Singh, A. Ahmadian, N. Senu, and R. Sarkar, "GRA_Net: A Deep Learning Model for Classification of Age and Gender from Facial Images," *IEEE Access*, vol. 9, pp. 85672–85689, 2021, doi: 10.1109/ACCESS.2021.3085971.

[23] M. J. Al Dujaili, H. T. S. Al Rikabi, N. K. Abed, and I. R. N. Al Rubeei, "Gender Recognition of Human from Face Images Using Multi-Class Support Vector Machine (SVM) Classifiers," *Int. J. Interact. Mob. Technol.*, vol. 17, no. 8, pp. 113–134, 2023, doi: 10.3991/ijim.v17i08.39163.

[24] J. G. Thanikkal, A. K. Dubey, and M. T. Thomas, "An Efficient Mobile Application for Identification of Immunity Boosting Medicinal Plants using Shape Descriptor Algorithm," *Wirel. Pers. Commun.*, vol. 131, no. 2, pp. 1189–1205, 2023, doi: 10.1007/s11277-023-10476-3.

[25] L. Martinelli *et al.*, "Face Masks During the COVID-19 Pandemic: A Simple Protection Tool With Many Meanings," *Front. Public Heal.*, vol. 8, no. January, pp. 1–12, 2021, doi: 10.3389/fpubh.2020.606635.

[26] S. Rab, M. Javaid, A. Haleem, and R. Vaishya, "Face masks are new normal after COVID-19 pandemic," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 6, pp. 1617–1619, 2020, doi: 10.1016/j.dsx.2020.08.021.

[27] Oluwatosin Seyi Oyebanji *et al.*, "Enhancing breast cancer detection accuracy through transfer learning: A case study using efficient net," *World J. Adv. Eng. Technol. Sci.*, vol. 13, no. 1, pp. 285–318, 2024, doi: 10.30574/wjaets.2024.13.1.0415.

[28] M. Behzadpour, B. L. Ortiz, E. Azizi, and K. Wu, "Breast Tumor Classification Using EfficientNet Deep Learning Model," pp. 1–19, 2024, [Online]. Available: http://arxiv.org/abs/2411.17870.

[29] A. O. Ibraheem, "Regularizing cross entropy loss via minimum entropy and K-L divergence," pp. 1–5, 2025, [Online]. Available: http://arxiv.org/abs/2501.13709.

[30] D. Feng *et al.*, "Penalized Entropy: a novel loss function for uncertainty estimation and optimization in medical image classification," *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2022-July, pp. 306–310, 2022, doi: 10.1109/CBMS55023.2022.00061.

[31] W. Kim and Y. Sung, "An Adaptive Entropy-Regularization Framework for Multi-Agent Reinforcement Learning," *Proc. Mach. Learn. Res.*, vol. 202, pp. 16829–16852, 2023.

[32] G. Jagatap, A. Joshi, A. B. Chowdhury, S. Garg, and C. Hegde, "Adversarially Robust Learning via Entropic Regularization," *Front. Artif. Intell.*, vol. 4, pp. 1–19, 2022, doi: 10.3389/frai.2021.780843.

[33] R. Jin, H. Li, J. Pan, W. Ma, and J. Lin, "Face Recognition Based on MTCNN and FaceNet," 2021, [Online]. Available: www.aaai.org.

[34] S. S. Khan, D. Sengupta, A. Ghosh, and A. Chaudhuri, "MTCNN++: A CNN-based face detection algorithm inspired by MTCNN," *Vis. Comput.*, vol. 40, no. 2, pp. 899–917, 2024, doi: 10.1007/s00371-023-02822-0.

[35] S. G. C, K. H. S, S. Shirahatti, and S. R. Bangari, "Face Recognition System for Real Time Applications Using Svm Combined With Facenet and Mtcnn," *Int. J. Electr. Eng. Technol.*, vol. 12, no. 6, pp. 328–335, 2021, doi: 10.34218/IJEET.12.6.2021.031.

[36] P. Khairunnisa, W. E. Putra, W. Yitong, A. Jufrizal, and M. N. A. Makmum, "Convolutional Neural Networks Using EfficientNetB0 Architecture and Hyperparameters on Skin Disease Classification," *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 2, no. 2, pp. 127–137, 2025, doi: 10.57152/predatecs.v2i2.1569.

[37] Q. L. MingxingTan, "EfficientNet: RethinkingModelScalingforConvolutionalNeuralNetworks," *Int. Conf. Mach. Learn. Long Beach, Calif.*, 2019.

[38] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017, [Online]. Available: http://arxiv.org/abs/1704.04861.

[39] H. L. Potgieter, C. Mouton, and M. H. Davel, "Impact of Batch Normalization on Convolutional Network Representations," *Commun. Comput. Inf. Sci.*, vol. 2326 CCIS, pp. 235–252, 2025, doi: 10.1007/978-3-031-78255-8_14.