Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 2, April 2026, pp: 696-711 e-ISSN: 2656-8632

RESEARCH ARTICLE

OPEN ACCESS

CNN-Based Facial Image Analysis for Pediatric Down Syndrome Classification

Yunidar Yunidar[1], Inda Mariana Harahap[2], Melinda Melinda[1], Rosmawinda[1], Nurlida Basir[3] , Aufa Rafiki[1], and Imam Fathur Rahman[1]

[1] Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia
[2] Department of Pediatry, Faculty of Nursing, Universitas Syiah Kuala, Darussalam, Banda Aceh, Indonesia
[3] Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Nilai Negeri Sembilan, Malaysia

Corresponding author: Yunidar Yunidar (e-mail: yunidar@usk.ac.id), **Author(s) Email**: Inda Mariana Harahap (e-mail: indamariana@usk.ac.id), Melinda Melinda (e-mail: melinda@usk.ac.id), Rosmawinda (e-mail: rosmawinda@mhs.usk.ac.id), Nurlida Basir (e-mail: nurlida@usim.edu.my), Aufa Rafiki (e-mail: aufa35@mhs.usk.ac.id), and Imam Fathur Rahman (e-mail: imamfr@mhs.usk.ac.id).

Abstract **Down syndrome (trisomy 21) is a genetic disorder caused by an extra copy of chromosome 21, resulting in distinctive developmental facial characteristics and intellectual delays. Early detection is crucial to enable timely medical intervention. However, conventional diagnostic procedures still rely on clinical observation and genetic testing, which can be invasive and expensive. This study proposes a facial image–based classification system for detecting Down syndrome using a Convolutional Neural Network (CNN) approach. Seven CNN architectures were evaluated, namely EfficientNetB0, MobileNetV2, ResNet34, ShuffleNetV2, AlexNet, VGG19, and InceptionV3, under two training scenarios: with and without early stopping. The dataset consisted of 1,000 facial images of children with and without Down syndrome, split into training, validation, and test sets at 60:20:20. Face detection was performed using the Haar Cascade Classifier, followed by data augmentation techniques including rotation, zoom, translation, horizontal flipping, and Gaussian noise to improve model generalization and reduce overfitting. Experimental results show that the VGG19 architecture achieved the best performance, with an accuracy of 94.5%, precision of 91.59%, recall of 98%, and an F1-score of 94.69%. A one-way ANOVA test yielded an F-value of 0.003 and a p-value of 0.955 (> 0.05), indicating no statistically significant difference between models trained with and without early stopping. Grad-CAM visualization highlighted key facial regions, namely the eyes, nose, and mouth, as the primary contributors to classification, while analysis using 68 facial landmark points revealed distinctive morphological patterns associated with Down syndrome. The integration of CNN models, Grad-CAM visualization, and facial landmark analysis demonstrates a promising, interpretable, and non-invasive approach to supporting early Down syndrome screening using facial images.**

Keywords **Down Syndrome; CNN; Grad-CAM; Facial Landmark; Haar Cascade; Early Stopping; ANOVA**

I. Introduction

Down Syndrome (trisomy 21) is a genetic disorder caused by the presence of all or part of an extra chromosome on chromosome 21. This condition is the most common chromosomal disorder, with a prevalence of approximately 1 in every 700 live births worldwide [1]. Syndromes are also at high risk of various medical disorders that affect almost all organ systems in the body, including the cardiovascular, respiratory, immune, and digestive systems [2]. Based on data from the World Population Review, the highest number of Down Syndrome cases was recorded in Malta, with approximately 989,000 cases, while in Indonesia, there were approximately 32,000 cases [3], [4]. Clinically, Down Syndrome is characterized by a combination of distinctive facial features and developmental delays, such as slower growth and mild to moderate intellectual disability. These genetic changes result in unique facial morphological features, including epicanthic folds, a flat nasal bridge, and hypertelorism (widely spaced eyes), which are important diagnostic indicators. Evolutionary morphological research indicates that both genetic and environmental factors during facial development contribute to the formation of these distinctive characteristics, with relatively slight variation across populations [5]. Although the core facial features associated with Down syndrome are generally consistent across populations, subtle variations in facial morphology may still occur due to ethnic diversity, environmental influences, and demographic differences. Therefore, the availability of diverse,

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

(a) Normal (b) Syndrome

Fig. 1. Sample Facial Image Dataset of Normal and Down Syndrome Children

representative datasets remains important to ensure that deep learning models generalize effectively across different populations.

Although visual observation-based diagnostic methods are non-invasive and readily accessible, their accuracy remains highly dependent on the clinician's expertise. This study demonstrates that an AI-based Down syndrome detection model can achieve 99.1%–99.2% accuracy, providing a more objective and standardized alternative to manual observation [6], [7]. Challenges with conventional methods, such as subjective bias and limited access to chromosome testing in remote areas [3], mean that many cases are not identified until significant developmental delays have already emerged.

Various studies have sought to address these limitations by leveraging facial imagery and artificial intelligence (AI). Studies [8], [9] used fetal ultrasound images to detect Down Syndrome and found that fetal head and facial morphological patterns could be potential visual indicators. Studies [10], [11], [12] used a deep learning-based facial recognition approach to identify various genetic syndromes and demonstrated that facial features can be an effective, non-invasive diagnostic indicator.

However, most previous Down Syndrome research has been limited to using a single CNN architecture and has not considered the influence of training strategies such as early stopping, which can significantly impact model performance and generalization. Furthermore, studies that deeply analyze facial landmark patterns or provide visual interpretations of classification results using Grad-CAM remain scarce, even though both are crucial for ensuring model transparency, interpretability, and reliability in medical contexts.

Based on these gaps, this research contributes in three main ways.

1. This study evaluates and compares seven CNN architectures: EfficientNetB0, MobileNetV2, ResNet34, ShuffleNetV2, AlexNet, VGG19, and InceptionV3 for facial image-based Down syndrome detection. These architectures were selected to represent different generations and design philosophies of convolutional neural networks, ranging from classic deep learning architectures like AlexNet and VGG19, residual networks like ResNet34, and multi-scale architectures like InceptionV3, to lightweight and computationally efficient models like MobileNetV2, ShuffleNetV2, and EfficientNetB0.
2. It evaluates 68 facial landmark points on facial images of individuals with Down Syndrome to identify distinctive morphological patterns that influence classification.

Therefore, this research is expected to provide an integrated approach that not only assesses model classification performance but also visually explains how CNNs recognize typical facial features of Down Syndrome. This approach has the potential to support the development of a non-invasive, objective, and efficient facial-image-based early-diagnosis system, especially for healthcare facilities with limited resources.

II. Material and Methods

A. Database Facial Image

This study utilized a dataset of children's facial images categorized into two classes: normal children and children with Down syndrome. The dataset was obtained from publicly available sources previously used in related studies [12], [13]. Specifically, the dataset referenced in [12] originates from a Kaggle repository developed for Down syndrome detection, containing facial images of children aged approximately 0–15 years, including both individuals with Down syndrome and healthy controls. The dataset was curated to incorporate variations in skin tone, eye color, hair type, and facial appearance, thereby improving diversity in facial characteristics.

For this study, a dataset of 1,000 images (500 per class) was selected for analysis. The data were accessed through publicly available platforms, including Kaggle and Roboflow [14], which provide

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

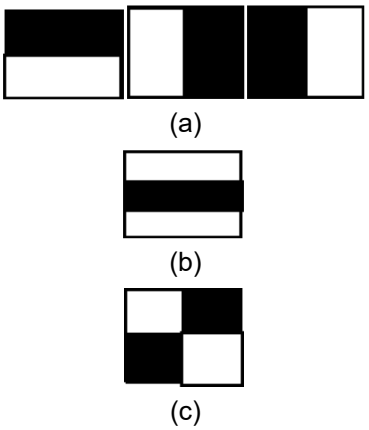

Fig. 2. Haar-like features, (a) Edge feature, (b) Line feature and (c) Four-rectangle feature

datasets specifically prepared for machine learning research. All images were distributed in anonymized form, meaning that no personally identifiable information, such as names, medical records, or geographic identifiers, was included. The dataset contains only facial photographs intended for academic research, ensuring compliance with ethical considerations for pediatric image data. The dataset was split into training, validation, and test sets at a 60:20:20 ratio. This split was chosen to ensure a sufficient number of samples for model training while maintaining separate validation and test sets for reliable performance evaluation. For relatively small datasets such as the 1,000 images used in this study, this proportion provides a balance between maximizing the amount of training data and preserving enough samples for unbiased validation and testing, which is a common practice in deep learning–based image classification studies. An example of the dataset is shown in Fig. 1.

B. Haar Cascade Classifier

The Haar Cascade Classifier is a classic yet effective method for object detection, particularly face detection, in digital images. This algorithm was first introduced by Paul Viola and Michael Jones (2001) [14], [15] through the Rapid Object Detection using a Boosted Cascade of Simple Features approach, which became a milestone in real-time face detection systems. This method uses a combination of Haar-like features, integral image processing, AdaBoost learning, and a cascade classifier to quickly and efficiently detect facial regions [16], [17]. Haar-like features are simple representations of the intensity differences between two or more rectangular areas in an image. This feature mimics the working principle of an edge detector, where light and dark areas of the face, such as around the eyes, nose, and mouth, produce distinctive contrast patterns. As illustrated in Fig. 2 , Haar-like features consist of basic patterns such as edge features, line features, and four-rectangle features that capture essential facial structures. Each Haar-like feature is calculated using Eq. (1) [18], which represents the difference between the sum of pixel intensities in the bright and dark rectangular regions. The integral image representation in Eq. (2) [19] enables efficient computation, while the cascade classifier defined in Eq. (3) [19] is used for face detection

$$f(x) = \sum(I_{white}) - \sum(I_{black}) \quad (1)$$

where I_{white} and I_{black}) indicate the sum of pixel intensities in the light and dark areas, respectively. To accelerate feature computation, the algorithm uses an integral image representation, defined as:

$$II(x,y) = \sum_{x' \le x, y' \le y} I(x',y') \quad (2)$$

where $II(x,y)$represents the integral image value at position $(x,y)$and $I(x',y')$is the pixel intensity at location (x',y'). This representation allows rapid computation of rectangular region sums regardless of their size. During detection, a sliding window scans the entire image to evaluate candidate regions. The cascade classifier then applies a sequence of increasingly complex classifiers to determine whether a region corresponds to a face. Mathematically, the detected facial region can be represented as:

$$R = H\,(I) \quad (3)$$

where I represents the input image and H denotes the Haar Cascade detection function that outputs the detected facial region R. In this study, face detection was implemented using a pre-trained Haar Cascade classifier provided by the OpenCV library, specifically the haarcascade_frontalface_default.xml model. This classifier has been widely used for frontal face detection and was trained on a large dataset of positive and negative facial samples using the Viola–Jones framework. The pre-trained model allows efficient detection of frontal facial regions without requiring additional training. During preprocessing, each input image was scanned using a sliding detection window across multiple scales. Regions detected as faces were then cropped and resized to 224 × 224 pixels to ensure consistent input dimensions for the CNN models used in the classification stage.

C. Augmentation

To improve model generalization and reduce the risk of overfitting, data augmentation was applied exclusively to the training dataset. Data augmentation is a widely used technique in deep learning that artificially increases the diversity of training data by generating new image variations from existing samples [20], [21]. In this study, five augmentation techniques were implemented: (1) zoom, with a scaling factor of 0.1 to simulate variations in the distance between the face and the camera; (2) random rotation, within a range of

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

±15° to represent variations in head orientation; (3) image translation, shifting the image by 2% of its original dimensions both horizontally and vertically; (4) horizontal flipping, which reflects natural symmetry variations in facial appearance; and (5) Gaussian noise addition, with a sigma value of 25 to improve the model's robustness to image noise. These augmentation operations were applied during training, before the images were fed into the CNN models.

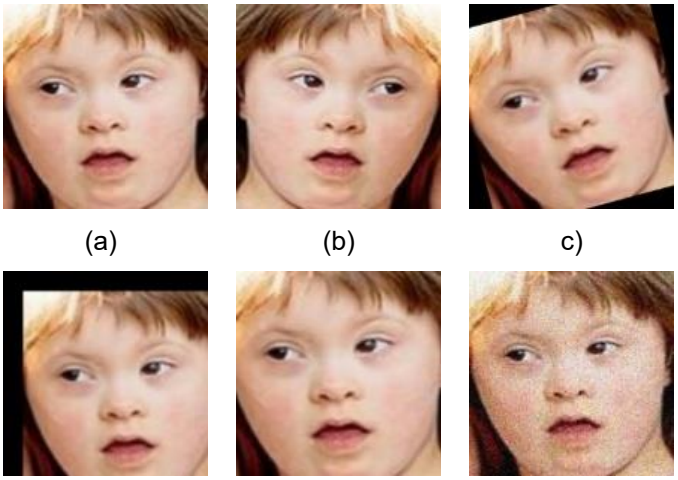

Fig. 3. Examples of Image Augmentation Techniques Applied to a Down Syndrome Dataset: (a) Original, (b) Flipping, (c) Rotation, (d) Translation, (e) Zoom, and (f) Noise

Through these transformations, the number of training samples increased from 600 original images to 3,600 augmented images. This augmentation strategy helps the model learn more robust facial representations that are invariant to variations in pose, scale, and noise. Examples of augmented images are shown in Fig. 3. Data augmentation has been proven to significantly improve recall and convergence performance in facial expression-based ASD classification tasks, particularly in medical-oriented deep learning evaluations [22],[23]. This augmentation process increases the amount of training data and enriches the distribution of children's facial variations. These enhancements play a crucial role in improving the performance and generalization capability of Convolutional Neural Network (CNN) models [24], [25]. By applying augmentation techniques such as zooming, rotation, translation, horizontal flipping, and the addition of Gaussian noise, the model becomes more robust to variations in image conditions, including differences in facial pose and the presence of visual noise [26].

D. Classification Scheme

Fig. 4 illustrates the overall architecture and workflow of the proposed Down syndrome classification system. The process begins with inputting a facial image dataset, followed by face detection using the Haar Cascade Classifier method to identify facial regions. The detected faces are then cropped to Regions of Interest (ROIs) and resized to a uniform 224 × 224 pixels. The dataset is then split into training, validation, and test sets in a 60:20:20 ratio. Data augmentation techniques are applied to the training data to increase data diversity and improve model generalization. The augmentation process uses five techniques: rotation, zoom, translation, horizontal flipping, and the addition of Gaussian noise, resulting in a total of 3,600 augmented training images. The augmented images are then used to train several Convolutional Neural Network (CNN) architectures. Model performance is then evaluated using standard classification metrics: accuracy, precision, recall, and F1-score. In the final stage, model interpretation is performed using Grad-CAM visualizations on the last convolutional layer of the trained architecture, as well as the detection of 68 facial landmark points to identify the facial areas that contribute most to the classification results. In this study, several CNN architectures were evaluated, including EfficientNetB0, MobileNetV2, ResNet34, ShuffleNetV2, AlexNet, VGG19, and InceptionV3. These models were selected to represent different design characteristics in modern convolutional neural networks. AlexNet is one of the earliest deep CNN architectures to demonstrate strong performance in image classification tasks [27]. VGG19 is a deeper network with stacked 3×3 convolutional layers that enable detailed feature extraction from facial images [28]. ResNet34 incorporates residual connections that facilitate deeper network training and improve feature learning [29]. InceptionV3 employs multi-scale convolutional modules that allow the network to capture visual patterns at different spatial resolutions [30]. Meanwhile, EfficientNetB0 [31], MobileNetV2 [32], and ShuffleNetV2 [33] are lightweight architectures designed for computational efficiency and optimized parameter utilization, making them suitable for deployment in resource-constrained environments. By evaluating architectures with varying depths, feature extraction mechanisms, and computational efficiencies, this study aims to provide a comprehensive comparison of CNN models for facial image–based Down syndrome detection.

Table 1. Hyperparameter Training

Epoch	Learning Rate	Optimizer	Batch Size
100	0.0001	Stochastic Gradient Descent (SGD)	32

During CNN model training, several key parameters affect performance: epoch, learning rate, batch size, optimizer, and loss function. Epoch

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

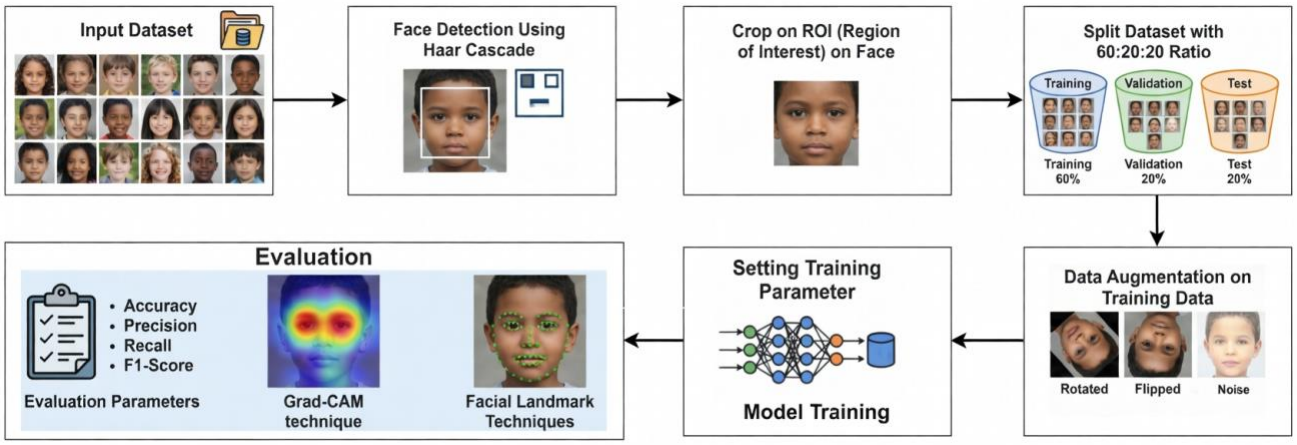

Fig. 4. Simulation of Deep Learning System for Automated Down Syndrome Detection

represents a complete cycle in which the model processes the entire dataset through forward and backward propagation [34]. Learning rate controls the amount of weight updates during each iteration [35], while batch size defines the number of data samples processed simultaneously. The optimizer determines how the model updates its weight based on the calculated loss [36]. This study uses a Stochastic Gradient Descent (SGD) optimization algorithm with Binary Cross-Entropy (BCE) as a loss function, which is suitable for binary classification tasks [36]. Binary Cross-Entropy (BCE), also known as Log Loss, is a standard loss function used in binary classification problems. The BCE loss value for each sample is defined as follows [37]:

$$L\,(y_i, \hat{p}_i) = [y_i\, log(\hat{p}_i) + (1 - y_i)\, log(1 - \hat{p}_i)] \quad (4)$$

In model training practice, what is minimized is the average loss value of all the data in the dataset, which is formulated as:

$$\mathcal{L}_{\text{BCE}}(\theta) = \frac{1}{n}\sum_{i=1}^{n} L\,(y_i, \hat{p}_i) \quad (5)$$

by stating the total number of samples and θ is a parameter of the CNN model. To iteratively update model parameters based on the loss values, the Stochastic Gradient Descent (SGD) optimization algorithm is used. The weight update process is performed using the following equation [38]:

$$w_t + 1 = w_t - \eta\,.\nabla\mathcal{L}_{\text{BCE}}\,(w_t) \quad (6)$$

where $w_t$denotes the weight at the t-th iteration, η is the learning rate, and $\nabla\mathcal{L}_{\text{BCE}}(w_t)$ represents the gradient of the loss function with respect to the weights.

Table 1 presents the training hyperparameters applied in this study. The model used 100 epochs, meaning it processed the entire training dataset 100 times in a single training cycle. The learning rate was set to 0.0001 to enable gradual, stable weight updates throughout training. The model processed 32 images per iteration as the batch size, which determined the number of samples used before updating the weights.

In addition to the standard training configuration, an early stopping mechanism was applied in one of the training scenarios to prevent overfitting and reduce unnecessary training iterations. Early stopping monitors the validation loss during training and stops the training process when no improvement is observed for a predefined number of epochs. In this study, validation loss was used as the monitored metric, with a patience value of 10 epochs and a minimum delta of 0.0001 to define a significant improvement. If the validation loss did not decrease by at least this threshold within 10 consecutive epochs, the training process was automatically terminated. This strategy helps prevent overfitting while improving training efficiency.

D. Evaluation Metrics

To evaluate the performance of the proposed CNN models, four standard classification metrics were employed, namely accuracy, precision, recall, and F1-score [32]. These metrics are computed based on the confusion matrix components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [39]. Accuracy measures the proportion of correctly classified samples among all samples and is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Precision represents the proportion of correctly predicted positive samples among all predicted positive samples and is expressed as:

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

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

Recall, also known as sensitivity, measures the proportion of correctly predicted positive samples among all actual positive samples and is calculated as:

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

The F1-score is the harmonic mean of precision and recall, which provides a balanced measure between the two, and is defined as:

$$F1\ Score = \frac{2\ (Precision\ .\ Recall)}{Precision + Recall} \quad (10)$$

Overall, the use of multiple evaluation metrics enables a more rigorous and reliable assessment of model performance, particularly in medical screening. While accuracy provides a general overview of classification correctness, it may be insufficient when class distributions are imbalanced. In this regard, precision and recall become more critical, as they reflect the model's ability to minimize false positives and false negatives, respectively. Notably, in medical applications such as Down syndrome screening, a high recall is essential to reduce the risk of missed diagnoses, which could delay early intervention. The F1-score further complements this analysis by providing a balanced measure between precision and recall. Therefore, the combined use of these metrics ensures a comprehensive evaluation framework that not only measures performance but also supports the reliability and clinical relevance of the proposed CNN-based classification system.

Additionally, beyond these commonly used metrics, it is important to consider the model's consistency, stability, and generalization across different data splits, architectures, and training strategies. Variations in training, validation, and test data can significantly influence performance, particularly with relatively limited and heterogeneous datasets such as pediatric facial images. Therefore, evaluating how models behave under different experimental settings, such as with and without early stopping, or across lightweight and deep architectures, provides deeper insight into their robustness. A reliable model is not only one that achieves high accuracy, but also one that consistently maintains high recall and a balanced F1-score across various conditions, indicating its ability to detect positive cases while minimizing both false positives and false negatives. This is especially crucial in medical screening tasks, where inconsistent performance could lead to unreliable decision-making. Furthermore, consistent evaluation results strengthen confidence that the model is not overfitting to specific data patterns, but instead learning meaningful and generalizable features. This ensures the model is reliable for real world clinical applications.

III. Result

A. Training and Validation

The training and validation loss curves for the seven CNN architectures, with and without early stopping, are presented in fig. 5. In general, different training behaviors depend on their complexity. For efficientnetb0, MobileNetV2, ResNet34, VGG19, AlextNet, and InceptionV3, the validation loss remains consistently higher than the training loss, indicating a tendency toward overfitting. In contrast, ShuffleNetV2 shows a more stable pattern, with both training and validation losses decreasing at a similar rate, suggesting better generalization. Early stopping helps reduce overfitting across several models. For example, training stops at epoch 48 for EfficientNetB0, 31 for MobileNetV2, 23 for ResNet34, 26 for VGG19, 30 for AlexNet, and 20 for InceptionV3. These earlier stopping points indicate that the models reach their optimal performance before completing the full training

Table 2. Performance Comparison of CNN Models for Down Syndrome Facial Image Classification

Architecture	Accuracy	Precision	Recall	F1 Score
EfficietNetB0	92%	90,38%	94%	92,16%
EfficietNetB0 Early Stopping	90%	86,36%	95%	90,48%
MobileNetV2	83,5%	81,31%	87%	84,06%
MobileNetV2 Early Stopping	84%	84%	84%	84%
ResNet34	94%	90,74%	98%	92,43%
ResNet34 Early Stopping	93,5%	93,07%	94%	93,53%
ShuffleNetV2	75%	70,83%	85%	77,27%
ShuffleNetV2 Early Stopping	77%	72,88%	86%	78,9%
VGG19	94,5%	91,59%	98%	94,69%
VGG19 Early Stopping	93,5%	89,18%	99%	93,83%
AlexNet	94%	94,89%	93%	93,93%
AlexNet Early Stopping	94%	92,31%	96%	94,12%
InceptionV3	94%	90,74%	98%	92,43%
InceptionV3 Early Stopping	91,5%	91,92%	91%	91,46%

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

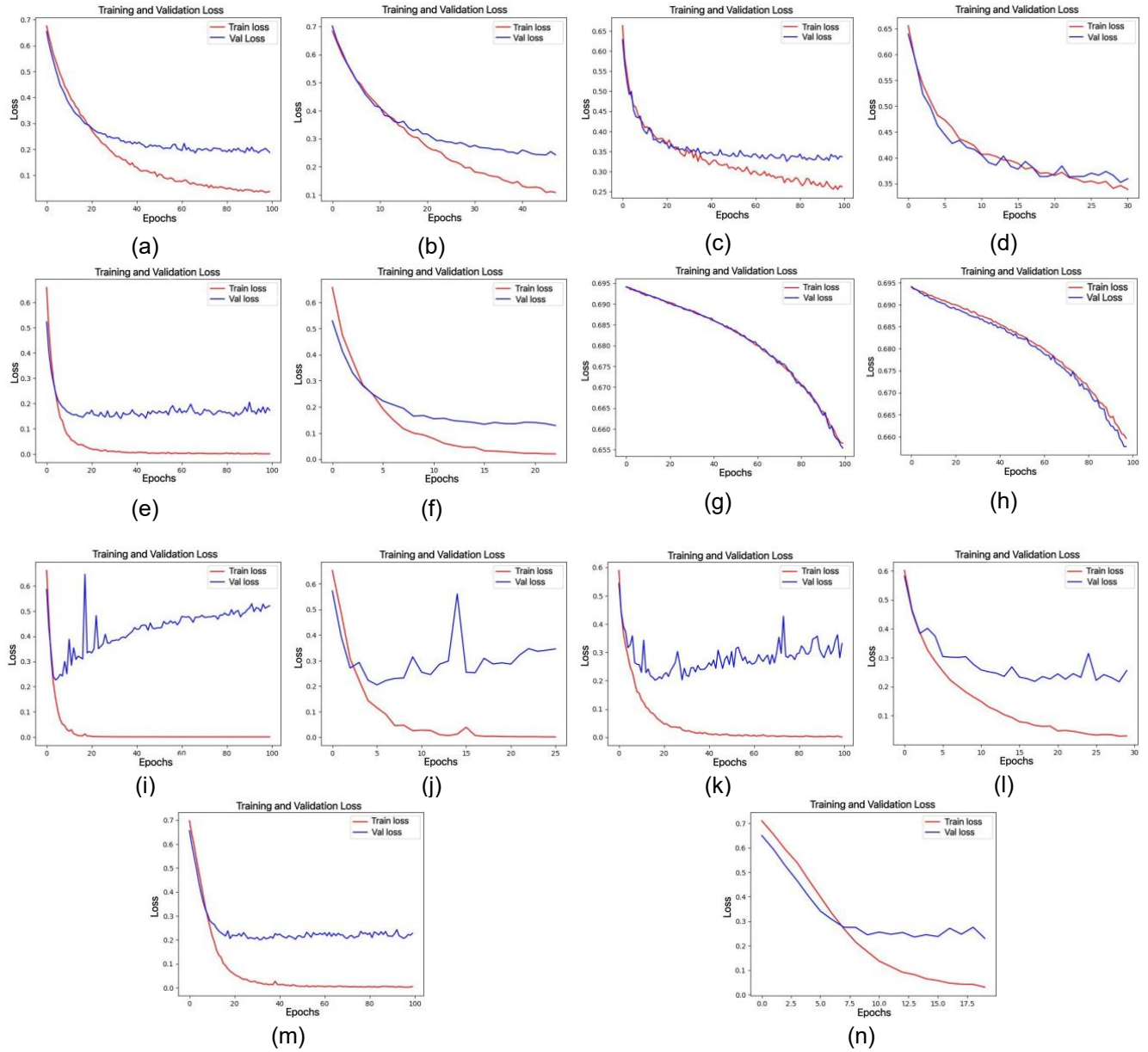

Fig. 5 Training and Validation Loss Graphs with Early Stopping Across CNN Models: (a) EfficietNetB0, (b) EfficietNetB0 Early Stopping, (c) MobileNetV2, (d) MobileNetV2 Early Stopping, (e) ResNet34, (f) ResNet34 Early Stopping, (g) ShuffleNetV2, (h) ShuffleNetV2 Early Stopping, (i) VGG19, (j) VGG19 Early Stopping, (k) AlexNet, (l) AlexNet Early Stopping, (m) InceptionV3, and (n) InceptionV3 Early Stopping

process. Meanwhile, ShuffleNetV2 behaves differently, where early stopping is triggered at epoch 98, which is much closer to the maximum training limit. This suggests that the model maintains stable learning without significant overfitting throughout training.

Overall, implementing early stopping with a patience value of 10 and a maximum of 200 epochs effectively prevents unnecessary training iterations in models prone to overfitting, while still allowing stable models such as ShuffleNetV2 to converge properly. In addition, this strategy helps optimize computational efficiency by reducing training time without significantly compromising model performance. It also improves generalization by stopping training at an optimal point before the model begins to memorize noise in the training data. As a result, the model achieves a better balance between accuracy and generalization.

B. Testing

The testing phase evaluates the performance of each CNN architecture using quantitative metrics derived from the confusion matrix (Fig. 6) and summarized in Table 2, including accuracy, precision, recall, and F1-score. These metrics enable direct comparison of the architectures and form the basis for further analysis

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

and discussion. Based on Table 2, VGG19 achieves the highest performance among all models, with an accuracy of 94.5%, precision of 91.59%, recall of 98%, and F1-score of 94.69%. Compared to other architectures, this model appears better at capturing distinguishing facial characteristics between the normal and Down syndrome classes. Similar observations have also been reported in previous studies, where deeper architectures tend to learn more discriminative features [26]. This indicates that increasing network depth enhances feature extraction, leading to more robust classification performance in facial image analysis.

ResNet34, AlexNet, and InceptionV3 achieve the same accuracy (94%), though their detailed metrics differ. ResNet34 and InceptionV3 both achieve a recall of 98%, indicating strong sensitivity, while AlexNet achieves the highest precision (94.89%), suggesting fewer false positives. This difference reflects a typical trade-off in classification models, where improving sensitivity may reduce precision, and vice versa [40].

A different pattern is observed in lightweight architectures. EfficientNetB0 achieves 92% accuracy,

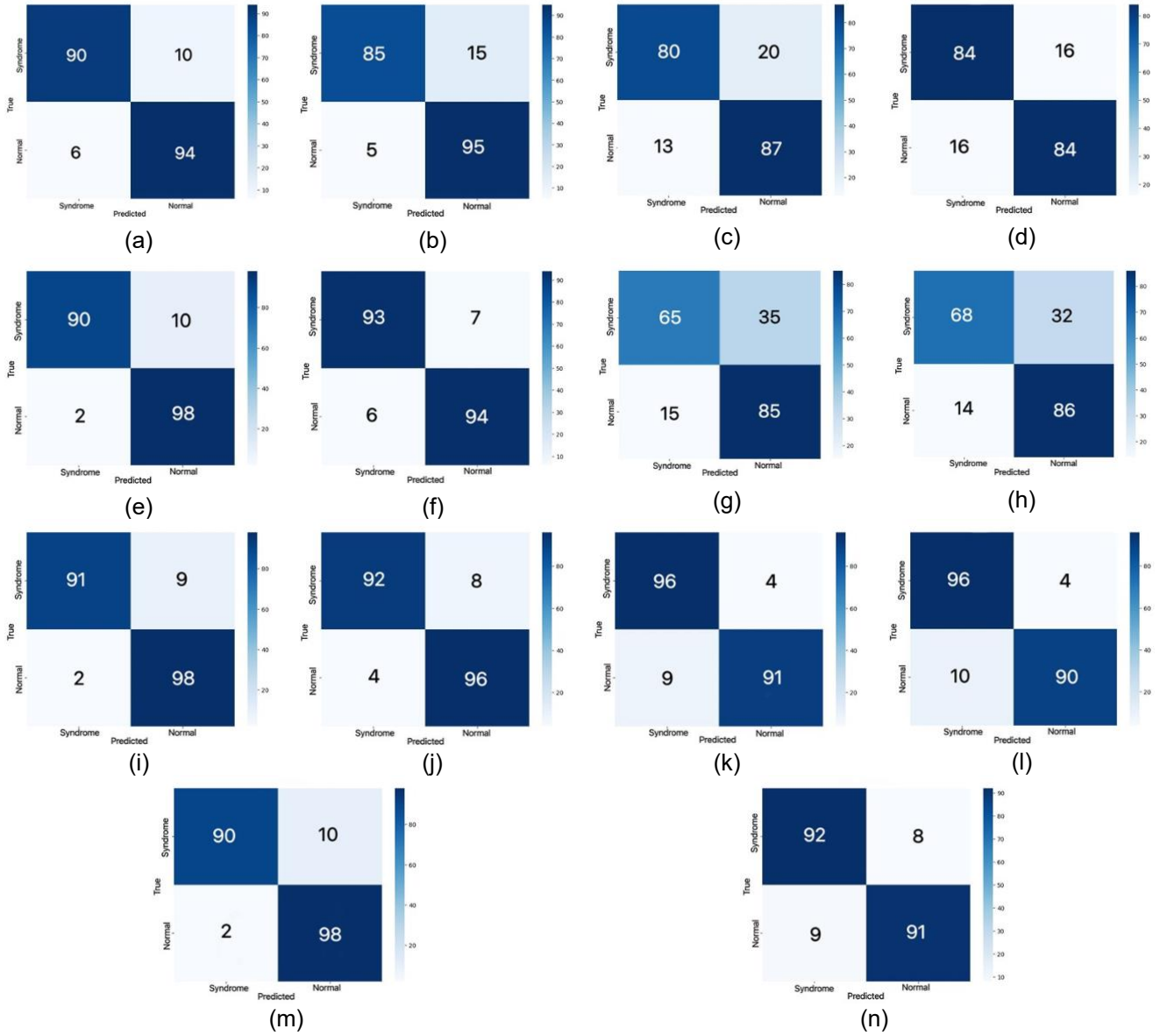

Fig. 6 Confusion Matrix Comparison of CNN Architectures: (a) EfficietNetB0, (b) EfficietNetB0 Early Stopping, (c) MobileNetV2, (d) MobileNetV2 Early Stopping, (e) ResNet34, (f) ResNet34 Early Stopping, (g) ShuffleNetV2, (h) ShuffleNetV2 Early Stopping, (i) VGG19, (j) VGG19 Early Stopping, (k) AlexNet, (l) AlexNet Early Stopping, (m) InceptionV3, and (n) InceptionV3 Early Stopping

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

while MobileNetV2 drops to 83.5%. ShuffleNetV2 performs worst, with 75% accuracy, 70.83% precision, 85% recall, and an F1-score of 77.27%. Although its training process appears stable, the lower scores indicate that the model struggles to represent more complex facial patterns. This aligns with previous findings that lightweight models often prioritize efficiency over feature representation capacity [34]. Looking more closely at recall values, models such as VGG19, ResNet34, and InceptionV3 (all 98%) show a strong ability to detect positive cases. In contrast, MobileNetV2 (87%) and ShuffleNetV2 (85%) are more likely to miss some cases. In medical screening, this difference becomes important because undetected cases can lead to delayed diagnosis [40]. The F1-score further supports this observation. VGG19 achieves 94.69%, followed by AlexNet (93.93%) and ResNet34 (92.43%), indicating relatively balanced precision and recall. ShuffleNetV2, on the other hand, remains significantly lower (77.27%), reinforcing its weaker overall performance.

The impact of early stopping is not uniform across models. For deeper architectures, such as VGG19, accuracy decreases slightly from 94.5% to 93.5%, while EfficientNetB0 drops from 92% to 90%. In contrast, MobileNetV2 shows a small improvement (83.5% to 84%), and ShuffleNetV2 increases from 75% to 77%. This suggests that early stopping may help simpler models avoid overfitting, but can limit deeper networks before reaching optimal performance [35]. From the confusion matrix in Fig. 6, misclassifications are relatively limited in high-performing models such as VGG19 and ResNet34. Most errors occur when facial characteristics are less distinct, suggesting that data variability also affects model performance. In general, the results point to the importance of model architecture. Deeper networks, particularly VGG19, consistently achieve higher accuracy and recall. Although slight overfitting can be observed (Fig. 5), the models still perform well on unseen data, suggesting that their generalization capability remains acceptable. The testing phase evaluates the performance of each CNN architecture using quantitative metrics derived from the confusion matrix (Fig. 6) and summarized in Table 2, including accuracy, precision, recall, and F1-score.

C. Evaluation Grad-CAM

This study applied the Gradient-weighted Class Activation Mapping (Grad-CAM) model interpretability technique to identify facial regions that contribute to classification decisions between the normal and Down syndrome classes. The best-performing model, VGG19, achieved an accuracy of 94.5%, precision of 91.59%, recall of 98%, and F1-score of 94.69%, as shown in Table 2. The visualization results in Fig. 7 show that the model consistently focuses on specific facial areas, particularly the eyes, nose, and mouth. These regions appear as high-activation areas (red–yellow), indicating their strong contribution to the classification output. Interestingly, these regions align with clinically recognized facial characteristics of Down syndrome, such as hypertelorism (widened interpupillary distance), a flattened nasal bridge, and a distinctive mouth structure. In terms of quantitative performance, the recall value of 98% suggests that the model is highly sensitive in detecting Down syndrome cases, meaning that only a small number of positive cases are missed. This aspect is especially important in medical screening, where false negatives can delay diagnosis and treatment. The statistical analysis further supports these findings. The one-way ANOVA test yielded an F-value of 0.003 and a p-value of 0.955 (Table 4), indicating no significant difference between models trained with and without early stopping. This implies that the observed performance differences are more likely to be influenced by the choice of CNN architecture than by the training stopping strategy.

Table 3. One-Way ANOVA Results Comparing CNN Models with and without Early Stopping

Source of Variation	SS	df	MS	F	P-value	Fcrit
Between Groups	0.1875	1	0.1875	0.003278	0.955473	4.964603
Within Groups	572.0416667	10	57.20417			
Total	572.2291667	11				

Table 4. Quantitative Facial Morphology Analysis Using 68-Point Landmark Measurements

Class	Forehead width (cm)	Mouth width (cm)	Nose height (cm)	Face height (cm)
Normal	4.69 - 5,78	2,53 - 5,37	2,04 - 4,31	5,14 - 9,48
Syndrome	4.79 - 5,78	2,85 - 5,36	2,16 - 3,90	5,66 - 9,51

Overall, the agreement between Grad-CAM visualization and quantitative metrics suggests that the model does not rely on arbitrary features. Instead, it focuses on facial regions that are both visually and clinically meaningful. This enhances the model's interpretability and supports its potential use in practical medical applications. Furthermore, this alignment between visual explanation and numerical performance increases trust in the model's decision-

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

making process, which is essential in sensitive domains such as healthcare. It also indicates that the model has successfully learned relevant feature representations that are consistent with established medical knowledge, thereby strengthening its potential for real-world clinical adoption.

D. Evaluation Using Facial Landmark 68

This study applied a 68-point facial landmark detection method to analyze morphological differences between normal children and children with Down syndrome. The detection was performed using a pre-trained dlib model, resulting in 68 key points distributed across major facial regions, including the eyes, nose, mouth, and facial contours, as illustrated in Fig. 8.

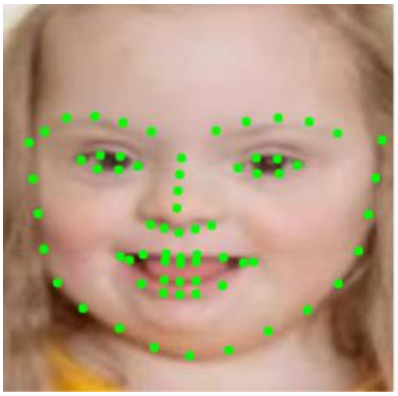

Fig. 8. Facial Landmark Detection Example Showing 68-Keypoint Localization

Quantitative analysis based on Euclidean distance measurements between landmark points (Table 3) shows clear differences between the two groups. In terms of mouth width, children with Down syndrome exhibit values ranging from 2.85 to 5.36 cm, compared with 2.53 to 5.37 cm in normal children. For nose height, the range in Down syndrome cases is 2.16-3.90 cm, which is generally lower than that of normal children (2.04-4.31 cm). Similarly, face height in children with Down syndrome ranges from 5.66 to 9.51 cm, slightly higher than the range observed in normal children (5.14-9.48 cm). These numerical differences indicate distinct facial morphology patterns. Specifically, children with Down syndrome tend to have a relatively wider mouth, a shorter nasal structure, and a facial proportion that appears more vertically compact. These findings are consistent with known clinical characteristics, such as hypertelorism, a flattened nasal bridge, and a distinctive facial structure. In addition to distance measurements, the spatial distribution of landmark points also shows a noticeable pattern. Landmark points in Down syndrome faces appear more concentrated in the central region of the face, suggesting a denser facial structure around the eyes, nose, and mouth. This observation is consistent with the Grad-CAM results (Fig. 7), which show that the CNN model places greater attention on these same regions.

The agreement between geometric measurements (Table 3) and visual attention (Grad-CAM) suggests that the model is not only accurate but also relies on meaningful facial features. This combination of quantitative analysis and visual interpretation strengthens the reliability of the proposed system. Furthermore, it enhances model interpretability, making the system more suitable for practical use in medical image analysis. Moreover, these findings highlight the potential of integrating geometric feature analysis with deep learning to improve both model transparency and diagnostic confidence. This integrated approach can serve as a robust foundation for developing clinically relevant and explainable AI systems in pediatric healthcare applications.

IV. Discussion

A. CNN Architecture Performance

VGG19, characterized by its deeper architecture and stacked 3×3 convolutional layers, achieved the best performance among all evaluated models, with an accuracy of 94.5%, precision of 91.59%, recall of 98%, and an F1-score of 94.69% (Table 2). The high recall value (98%) indicates that the model is highly sensitive in detecting Down syndrome cases, meaning that only a small number of positive cases are missed. This is particularly important in medical screening, where minimizing false negatives is critical for early diagnosis and intervention. Other architectures, including ResNet34, AlexNet, and InceptionV3, also achieved high accuracy values of 94%, although their performance differs in terms of precision and recall. For instance, AlexNet achieved the highest precision (94.89%), indicating fewer false positive predictions, while ResNet34 and InceptionV3 reached recall values of 98%, demonstrating comparable sensitivity to VGG19. In contrast, EfficientNetB0 achieved 92% accuracy, while MobileNetV2 achieved 83.5%, suggesting limitations in capturing more complex facial features.

ShuffleNetV2 exhibited the lowest performance, with an accuracy of 75%, precision of 70.83%, recall of 85%, and F1-score of 77.27%. Despite showing stable generalization, as indicated by the close alignment between the training and validation loss curves in Fig. 5, its lower accuracy and precision suggest that the model lacks sufficient representational capacity for this classification task. This highlights a trade-off between computational efficiency and classification performance. The effect of early stopping appears to be relatively limited across models. For example, VGG19 accuracy decreased slightly from 94.5% to 93.5%, EfficientNetB0 from 92% to 90%, and InceptionV3 from 94% to 91.5%. These small

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

differences are consistent with the ANOVA results (Table 4), which show a p-value of 0.955, indicating no statistically significant difference between models trained with and without early stopping. This suggests that performance variation is primarily influenced by the choice of CNN architecture rather than the training stopping strategy. Overall, these findings indicate that the representational capacity of the CNN architecture plays a more dominant role in determining performance. Deeper architectures, such as VGG19, are more effective in learning complex facial patterns, resulting in higher accuracy and recall. This observation is consistent with previous studies, which found that VGG19 achieved higher accuracy than lightweight models such as ShuffleNet, due to its stronger feature extraction capabilities [26].

B. Model Interpretability and Validation

The main contribution of this study lies in integrating quantitative performance evaluation with interpretability techniques to improve model transparency. Based on the results in Table 2, the best-performing model (VGG19) achieved an accuracy of 94.5%, a precision of 91.59%, a recall of 98%, and an F1-score of 94.69%. These values indicate that the model not only performs well in terms of accuracy but is also highly sensitive in detecting Down syndrome cases. The Grad-CAM visualization in Fig. 7 shows that the model consistently focuses on key facial regions, particularly the eyes, nose, and mouth. These regions appear as high-activation areas, indicating their strong contribution to classification. The model's focus on these areas is consistent with known clinical characteristics of Down syndrome, suggesting that the classification decisions are based on meaningful facial features rather than irrelevant patterns.

Further validation is provided by the facial landmark analysis (Table 3), which quantifies morphological differences between normal and Down syndrome subjects. For example, the mouth width in Down syndrome cases ranges from 2.85 to 5.36 cm, compared to 2.53 to 5.37 cm in normal cases. Similarly, the nose height is generally lower in Down syndrome cases (2.16–3.90 cm) than in normal subjects (2.04–4.31 cm). These measurable differences support the Grad-CAM findings and confirm that the model captures anatomically relevant variations. The relationship between interpretability and performance is evident in the model's high recall (98%), indicating that the features it uses are effective at identifying Down syndrome cases. In other words, the regions highlighted by Grad-CAM are not only visually prominent but also contribute directly to improved classification performance.

Overall, the agreement between Grad-CAM visualization (Fig. 7) and landmark-based measurements (Table 3) suggests that the model relies on clinically meaningful facial features. This enhances the interpretability and reliability of the system, making it better suited for medical screening applications that require both accuracy and transparency.

C. Implications, Limitations, and Future Work

The proposed system demonstrates strong potential as a non-invasive, fast, and cost-effective tool for preliminary screening, particularly in settings where access to genetic testing, such as karyotyping, is limited. This is supported by the experimental results (Table 2), which show that the best-performing model achieved 94.5% accuracy and 98% recall. The high recall value indicates that the system can identify most Down syndrome cases, which is essential for early detection and timely intervention in medical practice. Despite these promising results, several limitations should be acknowledged. First, the dataset used in this study consists of 1,000 facial images, which is relatively limited for training deep learning models such as VGG19. A dataset of this size may restrict the model's ability to generalize well to unseen data, especially when dealing with real-world variability.

Second, the dataset was collected from publicly available sources, which introduces variations in pose, lighting conditions, age distribution, and ethnic representation. These factors may affect model performance and introduce bias. For instance, although the model achieves a high overall accuracy (94.5%), this performance may not be consistent across all demographic groups. From an ethical perspective, the use of non-representative data can reduce the reliability of AI-based screening systems when applied to broader populations. If certain groups are underrepresented in the training data, the model may produce less accurate predictions for those populations. This highlights the importance of developing more diverse and balanced datasets in future research.

Future work should focus on expanding the dataset to include a larger, more diverse population and on improving robustness to variations in real-world conditions. In addition, integrating multimodal data, such as combining facial images with clinical or genetic information, may further improve classification performance. Exploring more advanced architectures or hybrid approaches could also enhance both accuracy and interpretability. Overall, while the proposed system achieves high performance (accuracy 94.5%, recall 98%), it should be considered as a supporting tool rather than a replacement for clinical diagnosis, ensuring that final decisions remain under professional medical supervision.

V. Conclusion

This study developed a deep learning–based facial image classification system for Down syndrome

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

detection using multiple CNN architectures. The results show that model architecture has a greater impact on performance than training strategy adjustments, with VGG19 achieving the best results (accuracy 94.5%, recall 98%, precision 91.59%, F1-score 94.69%). The high recall is especially important in medical screening to minimize missed cases. In contrast, ShuffleNetV2 demonstrated the most stable training behavior and good generalization but achieved lower accuracy (75%), highlighting the trade-off between model complexity and predictive performance. Beyond classification accuracy, this study emphasizes model interpretability and clinical relevance. Grad-CAM visualization confirmed that the CNN focuses on medically meaningful facial regions such as the eyes, nose, and mouth, while 68-point facial landmark analysis quantitatively validated morphological differences associated with Down syndrome. The agreement between deep learning attention maps and geometric facial measurements strengthens the reliability and transparency of the proposed system. Overall, the approach shows strong potential as a non-invasive, fast, and cost-effective preliminary screening tool, with future work needed to expand dataset diversity, perform external validation, and explore more advanced architectures and multimodal integration.

Acknowledgment

The authors would like to thank the Institute for Research and Community Service (LPPM) of Syiah Kuala University for their continued support and provision of research facilities that enabled the successful completion of this study. The Associate Professor Research Grant funded this research for Fiscal Year 2025, Syiah Kuala University, under Grant Number 492/UN11.L1/PG.01.03/14316-PTNBH/2025. Thank you for your attention. Sincerely.

Funding

This research was funded by the Institute for Research and Community Service, Syiah Kuala University (LPPM USK) with PNBP funding source through the Senior Lecturer Research Grant Scheme for the 2025 Fiscal Year.

Data Availability

This study uses a dataset of public facial images resulting from research described in [12]. The dataset can be accessed through the following repository: https://universe.roboflow.com/projectsqbm9c/detection-of-down-syndrome. All data used in this study are anonymized and intended solely for research purposes.

Author Contribution

Yunidar conceptualized the study, designed the overall research framework, supervised the implementation of the deep learning models, and led the interpretation of results and manuscript preparation. Inda Mariana Harahap contributed expertise in pediatric and child health perspectives, ensured clinical relevance of the study, and participated in manuscript review and validation. Melinda contributed to the development of multimedia and signal processing components, assisted in data analysis, and supported the writing and technical refinement of the manuscript. Nurlida Basir provided expert input in artificial intelligence methodology and signal processing, validated the experimental design, and contributed through critical revision of the manuscript. Rosmawinda conducted data collection, dataset preparation, model implementation, and assisted in performance evaluation and result documentation. Aufa Rafiki contributed to dataset management, experimental design, facial landmark analysis, and drafting the results and discussion sections. Imam Fathur Rahman contributed to CNN model implementation, Grad-CAM visualization, and statistical processing. All authors reviewed and approved the final version of the manuscript and agreed to be accountable for all aspects of the work, ensuring its accuracy and integrity.

Declarations

Ethical Approval

This study did not require formal approval from an ethics committee as it did not involve direct interaction with human participants. All facial images used in this study were obtained from a fully anonymized facial image dataset originally published by K. Rezaee (2025), ensuring that all ethical considerations regarding data collection have been taken into account by the original authors [11].

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] L. R. Chapman, I. V. P. Ramnarine, D. Zemke, A. Majid, and S. M. Bell, “Gene Expression Studies in Down Syndrome: What Do They Tell Us about Disease Phenotypes?,” *Int. J. Mol. Sci.*, vol. 25, no. 5, 2024, doi: 10.3390/ijms25052968.

[2] S. E. Antonarakis *et al.*, “Down syndrome,” *Nat. Rev. Dis. Prim.*, vol. 6, no. 1, pp. 1–43, 2021, doi: 10.1038/s41572-019-0143-7.

[3] Word Population Review, “Down Syndrome

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

Rates by Country 2025." Accessed: Sep. 13, 2025. [Online]. Available: https://worldpopulationreview.com/country-rankings/down-syndrome-rates-by-country

[4] T. I. W. Agustini Utari, Ferdy Kurniawan Cayami, Tithasiri Audi Rahardjo, Selvia Eva Sabatini, Vynda Ulvyana, "Critical Issue In The Identification Of Down Syndrome and Its Problems in Central Java, Indonesia: The Fact Of Needing Health Care And Better Management," *Intractable Rare Dis. Res.*, vol. 13, no. 2, pp. 121–125, 2024, doi: 10.5582/irdr.2023.01103.

[5] K. R. Nattariya Vorravanpreecha, Thanayoot Lertboonnum, Rungrote Rodjanadit, Pak Sriplienchan, "Studying Down syndrome recognition probabilities in Thai children with de-identified computer-aided facial analysis," *Am. J. Med. Genet.*, vol. 176, no. 9, 2018, doi: https://doi.org/10.1002/ajmg.a.40483.

[6] M. A. Shaikh and H. S. Al-Rawashdeh, "Deep Learning-Enabled Interpretable Down Syndrome Detection Model," *J. Disabil. Res.*, vol. 4, no. 3, pp. 1–12, 2025, doi: 10.57197/JDR-2025-0011.

[7] A. Mumuni and F. Mumuni, "Data augmentation with automated machine learning: approaches and performance comparison with classical data augmentation methods," *Knowl. Inf. Syst.*, vol. 67, no. 5, pp. 4035–4085, 2025, doi: 10.1007/s10115-025-02349-x.

[8] S. P. Arjunan and M. C. Thomas, "A Review of Ultrasound Imaging Techniques for the Detection of Down Syndrome," *Irbm*, vol. 41, no. 2, pp. 115–123, 2020, doi: 10.1016/j.irbm.2019.10.004.

[9] Vincy Devi VK and Rajesh R, "Down Syndrome Identification and Classification Using Facial Features With Neural Network," *Glob. J. Eng. Technol. Adv.*, vol. 12, no. 1, pp. 001–011, 2022, doi: 10.30574/gjeta.2022.12.1.0090.

[10] A. Raza, K. Munir, M. S. Almutairi, and R. Sehar, "Novel Transfer Learning Based Deep Features for Diagnosis of Down Syndrome in Children Using Facial Images," *IEEE Access*, vol. 12, no. February, pp. 16386–16396, 2024, doi: 10.1109/ACCESS.2024.3359235.

[11] N. Network, "Automatic Identification of Down Syndrome Using Facial Images with Deep Convolutional Neural Network," *Diagnostics*, vol. 10, no. 7, p. 487, 2020, doi: https://doi.org/10.3390/diagnostics10070487.

[12] K. Rezaee, "Machine learning and facial recognition for down syndrome detection : A comprehensive review," *Comput. Hum. Behav. Reports*, vol. 17, no. September 2024, p. 100600, 2025, doi: 10.1016/j.chbr.2025.100600.

[13] "Detection of Down Syndrome Computer Vision Project." [Online]. Available: https://universe.roboflow.com/projects-qbm9c/detection-of-down-syndrome

[14] Dherya Bengani and Prof. Vasudha Bah, "Face Detection using Viola Jones with Haar Cascade," *Int. J. Mod. Trends Sci. Technol.*, vol. 6, no. 11, pp. 131–134, 2020.

[15] M. K. Kumar *et al.*, "A Hybrid Model for Face Detection Using HAAR Cascade Classifier and Single Shot Multi-Box Detectors Based on Open CV," *Int. Res. J. Multidiscip. Scope*, vol. 5, no. 1, pp. 650–660, 2024, doi: 10.47857/irjms.2024.v05i01.0304.

[16] A. H. Alyousef, "Implementing Face Detector using Viola-Jones Method," *SSRG Int. J. Electr. Electron. Eng.*, vol. 10, no. 7, pp. 140–147, 2023, doi: 10.14445/23488379/IJEEE-V10I7P113.

[17] N. A. Mohd Ariffin, U. A. Gimba, and A. Musa, "Face Detection based on Haar Cascade and Convolution Neural Network (CNN)," *J. Adv. Res. Comput. Appl.*, vol. 38, no. 1, pp. 1–11, 2025, doi: 10.37934/arca.38.1.111.

[18] A. Jauhari, D. R. Anamisa, and Y. D. P. Negara, "Detection system of facial patterns with masks in new normal based on the Viola Jones method," *J. Phys. Conf. Ser.*, vol. 1836, no. 1, 2021, doi: 10.1088/1742-6596/1836/1/012035.

[19] C.-H. Choi, "Face Detection Using Haar Cascade Classifiers Based on Vertical Component Calibration," *Human-centric Comput. Inf. Sci.*, vol. 12, no. 11, p. 18, 2022, doi: https://doi.org/10.22967/HCIS.2022.12.011.

[20] H. Hassan *et al.*, "Review and classification of AI-enabled COVID-19 CT imaging models based on computer vision tasks," *Comput. Biol. Med.*, vol. 141, no. December 2021, p. 105123, 2022, doi: 10.1016/j.compbiomed.2021.105123.

[21] R. Kim and E. White, "Convolutional Neural Network for Data Augmentation Convolutional Neural network for Data Augmentation," *World J. Adv. Eng. Technol. Sci.*, vol. 13, no. 2, pp. 870–886, 2025, doi: https://doi.org/10.30574/wjaets.2024.13.2.0528.

[22] M. Melinda, N. A. C. Andryani, Y. Nurdin, V. Khariyunnisa, Y. Yulita, and I. K. A. Enriko, "Deep Learning Performance Analysis for Facial Expression Based Autism Spectrum Disorder Identification," *Radioelectron. Comput. Syst.*, vol. 4225, no. 2(110), pp. 30–40, 2024, doi: 10.32620/reks.2024.2.03.

[23] T. Kumar, R. Brennan, A. Mileo, and M. Bendechache, "Image Data Augmentation Approaches: A Comprehensive Survey and Future Directions," *IEEE Access*, vol. 12, pp. 187536–187571, 2024, doi: 10.1109/ACCESS.2024.3470122.

[24] A. Mumuni and F. Mumuni, "Data Augmentation:

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

A Comprehensive Survey of Modern Approaches," *Elseveir*, vol. 16, 2022, doi: 10.1016/j.array.2022.100258.
[25] F. Fahmi, M. Melinda, P. D. Purnamasari, E. Elizar, and A. Rafiki, "Recognition of EEG Features in Autism Disorder Using SWT and Fisher Linear Discriminant Analysis," *Diagnostics*, vol. 15, no. 18, pp. 1–16, 2025, doi: 10.3390/diagnostics15182291.
[26] M. Melinda, M. Oktiana, Y. Nurdin, I. Pujiati, M. Irhamsyah, and N. Basir, "Performance of ShuffleNet and VGG-19 Architectural Classification Models for Face Recognition in Autistic Children," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 13, no. 2, pp. 674–680, 2023, doi: 10.18517/ijaseit.13.2.18274.
[27] W. Tang, J. Sun, S. Wang, and Y. Zhang, "Review of AlexNet for Medical Image Classification," *EAI Endorsed Trans. e-Learning*, vol. 9, 2023, doi: 10.4108/eetel.4389.
[28] M. S. & A. M. Monika Bansal, Munish Kumar, "Transfer learning for image classification using VGG19: Caltech-101 image data set," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 3609–3620, 2023, doi: https://doi.org/10.1007/s12652-021-03488-z.
[29] L. Ma and Z. Long, "A Face Recognition Method Using ResNet34 and RetinaFace," *Acad. J. Comput. Inf. Sci.*, vol. 6, no. 10, pp. 18–23, 2023, doi: 10.25236/ajcis.2023.061003.
[30] W. Tang, "Review of Image Classification Algorithms Based on Graph Convolutional Networks," *Remote Sens.*, vol. 13, no. 22, pp. 1–51, 2021, doi: https://doi.org/10.3390/rs13224712.
[31] A. Zhou *et al.*, "Multi-head attention-based two-stream EfficientNet for action recognition," *Multimed. Syst.*, vol. 29, no. 2, pp. 487–498, 2023, doi: 10.1007/s00530-022-00961-3.
[32] S. Duhan, P. Gulia, N. S. Gill, T. A. Oliveira, and P. Kumar, "Evaluation of lightweight and efficient deep learning models for plant disease classification IN Evaluation of Lightweight and Efficient Deep Learning Models for Plant Disease Classification AR IN," *Discov. Internet Things*, 2026, doi: https://doi.org/10.1007/s43926-026-00310-0.
[33] A. Bauravindah and D. H. Fudholi, "Lightweight Models for Real-Time Steganalysis: A Comparison of MobileNet, ShuffleNet, and EfficientNet," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 158, 2024, doi: 10.29207/resti.v8i6.6091.
[34] F. Mehmood, S. Ahmad, and T. K. Whangbo, "An Efficient Optimization Technique for Training Deep Neural Networks," *Mathematics*, vol. 11, no. 6, p. 22, 2023, doi: 10.3390/math11061360.
[35] W. Cheng, R. Pu, and B. Wang, "AMC: Adaptive Learning Rate Adjustment Based on Model Complexity," *Mathematics*, vol. 13, no. 4, 2025, doi: 10.3390/math13040650.
[36] Y. Dong, "Practice and Optimization of Deep Learning Model Training," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 1, pp. 48–58, 2025, doi: 10.62051/ijcsit.v5n1.05.
[37] J. T. D. C. J. Romero, "A comprehensive survey of loss functions and metrics in deep learning," *Artif Intell Rev*, vol. 58, no. 195, 2025, doi: ttps://doi.org/10.1007/s10462-025-11198-7.
[38] A. Jentzen and A. Riekert, "A proof of convergence for stochastic gradient descent in the training of artificial neural networks with ReLU activation for constant target functions," *Zeitschrift fur Angew. Math. und Phys.*, vol. 73, no. 5, 2022, doi: 10.1007/s00033-022-01716-w.
[39] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020, doi: 10.1186/s12864-019-6413-7.
[40] B. Saif, "A Comprehensive Survey on AlexNet improvements and fusion techniques," *Fusion Pract. Appl.*, vol. 17, no. 2, pp. 123–146, 2025, doi: 10.54216/FPA.170210.

Author Biography

Yunidar was born in Banda Aceh, Aceh, on June 29th, 1974. She has been a lecturer at the Faculty of Engineering, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, since March 2000. After completing her undergraduate education in Physics at Universitas Syiah Kuala, Aceh, Indonesia, in 1997, she then obtained a master of Engineering (MT) degree in Optoelectrotechnics and Laser Applications from the University of Indonesia, Jakarta, Indonesia, in 2000. After which she has taken a doctoral degree program in electrical and computer engineering at Syiah Kuala University and graduated in 2025. She is also a member of IEEE. Her research interests include the implementation of biomedical engineering and sensors used in biomedical applications, including multimedia. She can be contacted at email: yunidar@usk.ac.id.

Inda Mariana Harahap, S.Kep., Ns., MNS, is a Lecturer (Lektor) in the Department of Pediatric Nursing at the Faculty of Nursing, Universitas Syiah Kuala, Indonesia. She specializes in pediatric nursing and child health, with research interests including child development, mental health in adolescents, and community-based health

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

promotion. Her recent research work includes examining the relationship between academic resilience and levels of anxiety, academic stress, and depression among high school students in Banda Aceh. She is actively engaged in teaching, research, and community service programs that support improving child and adolescent well-being. She is registered under the Indonesian SINTA ID: 6716419. Email: indamariana@usk.ac.id.

Melinda was born in Bireuen, Aceh, on June 10, 1979. She received a B. Eng degree from the Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, in 2002. She completed her master's degree at the Faculty of Electrical Department, University of Southampton, United Kingdom, with a concentration in the field study of Radio Frequency Communication Systems in 2009. She completed her Doctoral degree at the Department of Electrical Engineering, Engineering Faculty of Universitas Indonesia in February 2018. She has been with the Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala since 2002. She is also a member of IEEE. Her research interests include multimedia signal processing and fluctuation processing. She can be contacted at email: melinda@usk.ac.id.

Dr. Nurlida Basir is an Associate Professor at the Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM). She began her academic career at USIM in 2002 and has since been actively involved in teaching, research, and educational leadership. She holds a Diploma, a Bachelor's, and a Master's degree in Computer Science from Universiti Teknologi Malaysia (UTM), and earned her Ph.D. in Computer Science from the University of Southampton, United Kingdom. Her research interests span software engineering, cybersecurity, malware detection, signal processing, and artificial intelligence. Her research has been extensively published in prominent academic journals and conference proceedings. Alongside her research, she is a dedicated educator, mentoring both undergraduate and postgraduate students in computer science. She is a member of the Institute of Electrical and Electronics Engineers (IEEE), reflecting her active participation in the global academic and research community. She can be contacted at nurlida@usim.edu.my.

Rosmawinda was born in Pekanbaru on May 17, 2001. She is currently a Master's student in Electrical Engineering at Syiah Kuala University, with interests in multimedia technology and biomedicine. During her undergraduate studies, she focused on multimedia technology, particularly facial recognition research, which served as the foundation for her master's-level research. As a student in the class of 2025, she actively participates in lectures. She continues to deepen her understanding of image processing, deep learning, and the application of intelligent technology in electronics. Her commitment is reflected in her ongoing efforts to combine theory with practice, resulting in relevant and innovative research. Dedicated to scientific development and applied skills, she is determined to make a significant contribution to technological advancement, particularly in the development of biometric systems and artificial intelligence. She can be reached via email: rosmawinda@mhs.usk.ac.id.

Aufa Rafiki was born on April 20, 2003, in Banda Aceh. He received his Bachelor's degree in 2025 from the Department of Electrical and Computer Engineering, Universitas Syiah Kuala, where he focused on multimedia technology and the analysis of EEG signals. He is currently pursuing a Master's degree in Electrical Engineering at Universitas Syiah Kuala, with research interests centred on biomedical signal processing and deep learning for EEG-based applications. During his undergraduate studies, he served as a teaching assistant and programming laboratory assistant, gaining experience in both teaching and practical implementation. He is committed to strengthening his expertise in theory and applied research to contribute to technological developments in his field. He can be contacted at aufa35@mhs.usk.ac.id.

Imam Fathur Rahman was born on April 23, 2003, in Bireuen. He is currently pursuing a Master's degree in Electrical Engineering at Universitas Syiah Kuala, with a concentration in Biomedical Engineering. He completed his undergraduate studies at the Department of Electrical and Computer Engineering, Universitas Syiah Kuala, focusing on multimedia telecommunications engineering, with research specialization in EEG signal analysis. He actively engages in academic activities and continuously enhances his expertise in the field. Beyond his studies, he has gained experience as a teaching assistant and a digital signal processing

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523

laboratory assistant during his seventh semester. As part of the 2021 cohort, he strives to refine his skills and broaden his practical experience. His academic journey reflects a strong commitment to integrating theory and practice, equipping him with the knowledge to contribute to future technological advancements. He can be contacted at: imamfr@mhs.usk.ac.id.

Manuscript received 2 February 2026; Revised 10 April 2026; Accepted 15 April 2025; Available online 23 April 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1523