Available online at http://iorajournal.org/indx.php/orics /index

Operations Research: International Conference Series

e-ISSN: 2722-0974
p-ISSN: 2723-1739

Vol. 6, No. 2, pp. 117-122, 2025

Implementing EfficientNetB0 for Facial Recognition in Children with Down Syndrome

Dede Irman Pirdaus[1,*], Muhammad Bintang Eighista Dwiputra[2], Moch Panji Agung Saputra[3]

[1]*Informatics Study Program, Faculty of Computer Science, University of Informatics and Business, Bandung, Indonesia*
[2]*Computer Science Study Program, Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia, Bandung, Indonesia*
[3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, Indonesia*

**Corresponding author email: dedeirmanpirdaus@gmail.com*

Abstract

Early detection of Down Syndrome in children is crucial to provide more appropriate medical and educational interventions. This study aims to build and evaluate a deep learning-based classification model using the EfficientNetB0 architecture to distinguish facial images of children with Down Syndrome and healthy children. The dataset used consists of two classes (Down Syndrome and healthy), which have gone through an augmentation process to increase data diversity and prevent overfitting. The model was trained using the Adam algorithm with a learning rate of 0.0001 and a sparse categorical crossentropy loss function for 10 epochs. The training results showed that the model achieved a validation accuracy of 93.94%, with the lowest validation loss value of 0.2390. Further evaluation was carried out using a confusion matrix, which showed that the model was able to properly classify 312 out of 333 Down Syndrome images and 309 out of 330 healthy children images, resulting in an overall accuracy of 94%. In addition, the precision, recall, and f1-score values for both classes were in the range of 0.94, indicating a balanced and strong performance. Visual analysis of the misclassified images indicates that some misclassifications occur on healthy children's faces with certain expressions, angles, or lighting conditions that resemble Down syndrome. Conversely, some children with Down syndrome are also predicted as healthy when their facial features are not too prominent or similar to normal children under certain lighting conditions. This shows that despite the high performance of the model, sensitivity to facial feature variations remains a challenge.

Keywords: Down Syndrome, face classification, EfficientNetB0, deep learning, model evaluation, medical image

1. Introduction

Down Syndrome is a genetic disorder caused by the presence of an extra copy of chromosome 21 (trisomy 21), which affects a child's physical, intellectual, and facial growth (Akhtar and Bokhari, 2018). Typical features often seen in children with Down Syndrome include a flat face, slanted eyes that point upwards, a small nose, and a short neck. Early identification of these facial characteristics is essential to support early diagnosis, appropriate intervention, and optimal therapy planning, especially in areas with limited access to genetic testing and specialist medical services (Kaur and Patel, 2025).

The development of computer vision and deep learning technology in recent years has provided significant progress in the fields of facial recognition and image classification (Memari, 2023). Convolutional Neural Network (CNN) models have been proven to be able to extract spatial features from images efficiently, and have been widely used in various medical applications, including disease detection from X-ray images, MRI, and facial abnormality identification. Transfer learning techniques allow the use of pretrained models to improve accuracy on small datasets with shorter training times (Li et al., 2023).

Several previous studies have tried to apply CNN to detect Down Syndrome from facial images. Pranatha et al. (2020) used the ResNet34 architecture with a transfer learning approach and achieved an accuracy of 87.9% in distinguishing the faces of children with and without Down Syndrome. Another study by Evansyah and Kusuma (2025) compared the performance of the VGG16 and VGG19 models in classifying the faces of European Down Syndrome children, using a dataset of 1,570 images. They reported that VGG16 was superior with an accuracy of 94%, compared to VGG19 at 90%.

Despite showing promising results, these approaches generally use more complex models and do not focus on computational efficiency. EfficientNet is a modern CNN architecture designed to achieve a balance between accuracy and computational efficiency through compound scaling techniques (Tan and Le, 2019). Among its variants, EfficientNetB0 is a lightweight version that still has superior performance even with a relatively small number of parameters. This architecture is well suited for medical classification applications in resource-constrained environments (Ozsari et al., 2024).

Based on the literature review that has been conducted, no previous studies have been found that specifically implement the EfficientNetB0 architecture for the task of classifying children's faces with Down Syndrome. The absence of this research indicates that there is a research gap that is still open in the utilization of EfficientNet-based deep learning models. Therefore, this study aims to apply and investigate the performance of the EfficientNetB0 architecture in detecting and classifying children's faces with Down Syndrome.

2. Methodology

2.1. Data collection

The facial image data of children with and without Down Syndrome in this study were obtained through the Roboflow platform, which provides an open image annotation dataset for various computer vision-based object detection and classification needs. This dataset consists of 2,317 image files that have been categorized into two classes, namely Down Syndrome and Non-Down Syndrome (Aisha, 2025).

2.2. Data Preprocessing

Preprocessing is a crucial stage in digital image processing to ensure that the data entering the model has a uniform structure and format. In this study, several preprocessing stages were carried out before the model training process, as explained below:

1) Image Resizing
   All images in the dataset were resized to 224 × 224 pixels, adjusting to the standard input of the EfficientNetB0 architecture. This process aims to equalize the input dimensions and reduce the computational burden (Alhijaj and Khudeyer, 2023).
2) Dataset Division
   The process of dividing the dataset into three parts, namely training data (training set), validation (validation set), and testing (testing set). Of the total dataset obtained from the Roboflow platform, 2317 images were used, consisting of two classes: children with Down Syndrome and non-Down Syndrome children. A total of 1854 images were allocated for training, while the rest were used for validation and testing.
3) Augmentation
   Augmentation is performed in real-time during the training process and includes transformations such as horizontal flips, random rotations, random zooms, and contrast changes. These augmentation techniques significantly increase the diversity of the training data without having to manually add new data, thus helping the model learn more general features (Shorten and Khoshgoftaar, 2019).
4) Prefetching
   Prefetching allows the system to load the next batch of data in parallel while the model is still processing the current batch. This improves training efficiency and reduces waiting time in the data pipeline (Varsaci and Busonera, 2025). The entire dataset is then converted to a batch format with a size of 32 images per batch.

2.3. Model Architecture

The model used in this study is based on the EfficientNetB0 architecture, which is a modern convolutional model designed with a compound scaling approach to balance the depth, width, and resolution of the network efficiently.

Table 1: Architectural structure of the EfficientNetB0 model used

No	Layer (type)	Output Shape	Parameter	Information
1	EfficientNetB0 (backbone)	(None, 1280)	4.049.571	Pretrained feature extractor (ImageNet)
2	Flatten	(None, 1280)	0	Flattening the feature tensor
3	Dense (512 units, ReLU)	(None, 512)	655.872	Fully-connected layer
4	Dense (2 units, Softmax)	(None, 2)	1.026	Final classification layer

In its implementation, the EfficientNetB0 model is used as a feature extractor with pretrained weights from ImageNet. The top layer of the model is removed, the last 20 layers of EfficientNetB0 are unfrozen. The complete structure of the model can be summarized in Table 1.

2.4. Model Training

After the architecture development process is complete, the EfficientNetB0 model is compiled using the Adam optimization algorithm and a learning rate of 0.0001. The loss function used is sparse categorical crossentropy, because the target labels are encoded in integer form (not one-hot encoding) and the number of classes is two. As an evaluation metric, accuracy is used, which calculates the percentage of correct predictions against the total data. The model is trained for 10 epochs using previously processed and augmented training data.

2.5. Model Evaluation

Model Evaluation One of the main approaches in evaluating the performance of a classification model is the use of a confusion matrix. A confusion matrix is a tabular representation that shows the number of correct and incorrect predictions of a classification model, based on the actual labels and predicted labels. In practice, a confusion matrix is used to calculate accuracy, precision, recall and F1-score.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

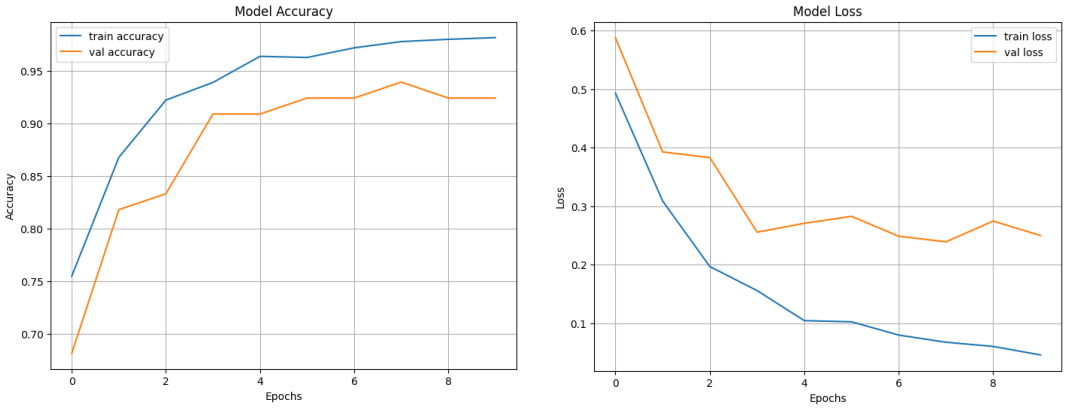
$$F1 - Score = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

where TP (True Positive), (TN) True Negative, (FP) False Positive and (FN) False Negative.

3. Results and Discussion

3.1. Performance during the model training process

The results during the model performance training process in Figure 1 (a) show the accuracy graph during training, while Figure 2 (b) shows the loss graph.



(a) (b)

Figure 1: Graph during model training

In the first epoch, the model achieved a training accuracy of 66.06% and a validation accuracy of 68.18%, with loss values of 0.5936 and 0.5878, respectively. As the number of epochs increases, the model accuracy increases significantly. In the 4th epoch, the model has achieved stable performance acceleration, marked by a validation accuracy of 90.91% and a decrease in validation loss to 0.2554. Performance continues to increase until it reaches the highest validation accuracy of 93.94% in the 8th epoch, with the lowest validation loss of 0.2390. Although there is a slight fluctuation in the loss value after that, the accuracy remains stable above 92% until the end of training.

3.2. Model Evaluation

Model performance evaluation was conducted using test data with 663 images that were evenly divided between Down Syndrome and Healthy classes. The model performance evaluation metrics will be presented in Table 1:

Table 2: Model performance evaluation

Label	Precision	Recall	F1-Score	Support
Down	0.94	0.94	0.94	333
Healthy	0.94	0.94	0.94	330
Accuracy			0.94	663
Macro Avg	0.94	0.94	0.94	663
Weighted Avg	0.94	0.94	0.94	663

Table 2 shows a very good and balanced classification performance for both classes, namely Down Syndrome and Healthy. The precision value achieved for each class is 0.94. This shows that of all model predictions for a class, 94% of them are correct. In other words, the model rarely makes false positive prediction errors. The F1-score value of 0.94 for both Down Syndrome and Healthy classes reflects the balance between precision and recall. The overall average accuracy of the model on all test data reaches 94%, which is a competitive result and illustrates the reliable performance of the model. The macro and weighted average metrics, which are also 0.94, indicate that the distribution of model performance is not only stable across classes but also considers the proportion of data in each class in the overall evaluation. Figure 2 shows the confusion matrix of the model prediction results on the test data.

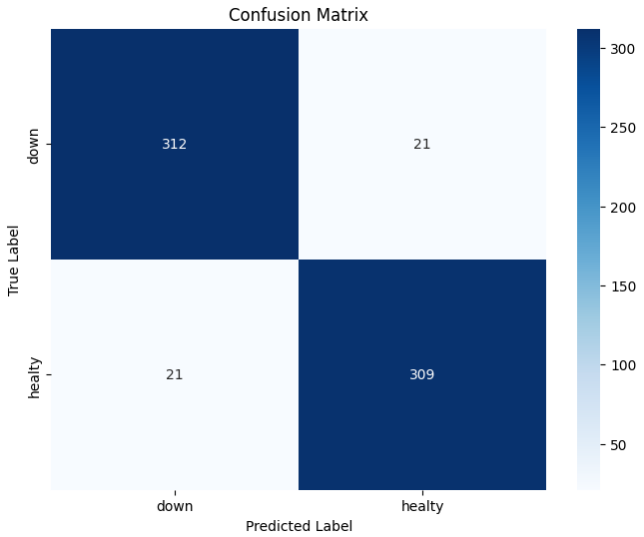


Figure 2: Confusion matrix

In the Down Syndrome class classification, 312 images were successfully predicted correctly as Down Syndrome (True Positive), while 21 other images were incorrectly predicted as Healthy (False Negative). For the Healthy class, 309 images were correctly predicted (True Negative), and 21 images were incorrectly classified as Down Syndrome (False Positive). This shows that the model also has a good level of specificity, with the ability to effectively distinguish healthy individuals from those with Down Syndrome. The model produced a total of 621 correct predictions from 663 images, which is equivalent to 94% accuracy. The distribution of model errors looks balanced in both classes, with the same number of errors (21) in each class, so there is no tendency for bias towards one class.

3.3. Visual Analysis of True and False Predictions

In this section we will explain how the model predicts true and false based on the input images which will be shown in figures 3 and 4.

Figure 3: Example of correct prediction

Some of the children with Down syndrome in these images display facial expressions or angles that resemble those of children without Down syndrome, such as smiling, holding their head high, or making strong direct eye contact. Typical visual characteristics of Down syndrome, such as slanted eyes, small noses, or certain facial shapes, are not as prominent in some of these images, either due to lighting, pose, or image resolution. The model may have difficulty detecting variations in the facial expressions or attributes of children with Down syndrome that are more subtle or do not fit the general patterns it learned during training.

Figure 4: Examples of incorrect predictions

Some children display facial expressions such as open mouths, blank stares, or slightly tilted heads that coincidentally resemble typical patterns of children with Down syndrome from the model's perspective. The model may recognize certain features such as eye shape or facial structure that resemble images of children with Down syndrome from the training data, resulting in incorrect classifications. Both figures show that despite the model's high accuracy (94%), there are still some incorrect predictions, especially in the case of faces with neutral or unclear expressions. This suggests that the model may be relying more on superficial or specific visual features than understanding the overall context of the face, and it is therefore important to consider data augmentation, increasing the diversity of training images, or even model interpretability approaches such as Grad-CAM in future development.

4. Conclussion

This study successfully developed and implemented a deep learning-based classification model using the EfficientNetB0 architecture to recognize the faces of children with Down Syndrome. By utilizing the facial image dataset from Roboflow, preprocessing processes such as resizing, augmentation, and prefetching were carried out to improve the quality of the training data. The model was trained for 10 epochs with the Adam algorithm and the sparse categorical crossentropy loss function. The training results showed very good performance, with the highest validation accuracy of 93.94% and the lowest loss value of 0.2390. Evaluation using test data showed an overall accuracy of 94%, as well as balanced precision, recall, and f1-score values at 0.94 for both classes (Down Syndrome and healthy). Visual analysis of incorrect predictions showed that facial expressions, shooting angles, and lighting could affect the classification results.

References

Aisha. (2025). Down Syndrome - Computer Vision Project. Roboflow Universe. https://universe.roboflow.com/aisha-bbzt4/down-syndrome-mfku9

Akhtar, F., & Bokhari, S. R. A. (2018). Down syndrome.

Alhijaj, J. A., & Khudeyer, R. S. (2023). Integration of efficientnetb0 and machine learning for fingerprint classification. *Informatica*, *47*(5).

Evansyah, E. B., & Aditya, C. S. K. (2025). Comparison of VGG16 and VGG19 Models in the Classification of Down Syndrome in the European Region with Transfer Learning. *INOVTEK Polbeng-Seri Informatika*, *10*(2), 922-933.

Kaur, K., & Patel, B. (2025). Retinoblastoma. *StatPearls*.

Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. *Frontiers in public health*, *11*, 1273253.

Memari, M. (2023). Advances in computer vision and image processing for pattern recognition: a comprehensive review. *International Journal of Engineering and Applied Sciences*, *11*(05), 2896-2901.

Ozsari, S., Kumru, E., Ekinci, F., Akata, I., Guzel, M. S., Acici, K., ... & Asuroglu, T. (2024). Deep Learning-Based Classification of Macrofungi: Comparative Analysis of Advanced Models for Accurate Fungi Identification. *Sensors*, *24*(22), 7189.

Pranatha, M. D. A., Setiawan, G. H., & Maricar, M. A. (2024). Utilization of ResNet Architecture and Transfer Learning Method in the Classification of Faces of Individuals with Down Syndrome. *Journal of Applied Informatics and Computing*, *8*(2), 434-442.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, *6*(1), 1-48.

Versaci, F., & Busonera, G. (2025). Hiding Latencies in Network-Based Image Loading for Deep Learning. *arXiv preprint arXiv:2503.22643*.