J. Intell. Comput. Health Inform. Vol. 5, No. 2, 2024
ISSN: 2715-6923 e-ISSN: 2721-9186, pp.43~50 | DOI: 10.26714/jichi.v5i2.14338

EfficientNet for Medical Image Classification: Performance vs. Efficiency in Skin Cancer Detection

Muhammad Wigig Purbandanu*, Rizky Yanuarta, Arif Kurniawan, Muhammad Munsarif*, Ahmad Ilham

Department of Informatics, Universitas Muhammadiyah Semarang, Jl. Kedungmundu Raya No. 18, Semarang 50273, Indonesia
*Corresponding author: wigigpurbanda@gmail.com, m.munsarif@unimus.ac.id

Ayomikun A. Awoseyi

Department of Mathematics and Computer Sciences, First Technical University, Ibadan, Oyo State, Nigeria.

Article histroy :
Received: 14 MAR 2024
Accepted: 2 AUG 2024
Avalaible online: 30 SEP 2024

Research article

Abstract: This study applies EfficientNetB2, a computationally efficient convolutional neural network (CNN), to improve the accuracy of skin cancer detection using the heterogeneous HAM10000 dataset. Skin cancer classification poses challenges, including overfitting and class imbalance, which we address through data augmentation, class weighting, and SMOTE (Synthetic Minority Over-sampling Technique). Our model achieved accuracy of 86%, precision of 0.87, recall of 0.85, and an AUC of 0.90. These results outperform comparable architectures, such as ResNet50 and GoogleNet, while maintaining lower computational complexity. The proposed model demonstrates high precision in detecting actinic keratoses and basal cell carcinoma, which require timely treatment, but faces difficulties in differentiating melanoma from benign nevi because of their similar visual appearance. This study highlights the potential of EfficientNetB2 for real-world deployment in resource-limited settings, such as mobile health applications and telemedicine platforms. Future research will focus on integrating attention mechanisms and exploring cross-dataset validation to enhance model generalizability and performance.

Keywords: EFFICIENTNETB2; SKIN CANCER DETECTION; CLASS IMBALANCE MOBILE HEALTH APPLICATIONSONE

1. Introduction

Skin cancer is one of the most prevalent cancers globally, and its incidence is increasing because of increased ultraviolet radiation (UVR) exposure, lifestyle changes, and genetic predisposition (Parker, 2021). Early detection is essential to achieve effective treatment and reduce mortality. Traditional methods, such as visual inspection and biopsy, are reliable, but they are limited by the expertise of clinicians and the time required for diagnosis (Li et al., 2022). In recent years, artificial intelligence (AI)-based solutions, particularly Convolutional Neural Networks (CNNs), have demonstrated great potential for automating skin cancer detection with high accuracy (Goceri & Karakas, 2020).

Despite the success of previous models such as ResNet and GoogleNet, several challenges remain unresolved, including overfitting, class imbalance, and limited generalizability across heterogeneous datasets. Overfitting occurs when models become overly specialized in training data, which reduces their ability to perform well on unseen data. Moreover, most skin cancer datasets are imbalanced, with benign lesions being overrepresented, which leads to biased predictions toward the majority classes (Hellín et al., 2024; Ragupathi et al., 2022). Addressing these issues is critical for creating robust models that are suitable for clinical environments.

EfficientNet, proposed by Ravi et al. (2021), introduces a compound scaling approach that optimizes the depth, width, and resolution of networks, thereby making them both accurate and computationally efficient. However, although previous studies employing EfficientNet for general medical imaging tasks have shown promising results (Anwar, 2023), its application to skin cancer detection remains underexplored. This study sought to fill this gap by leveraging EfficientNetB2, a lightweight and high-performance variant, to classify skin lesions effectively even in resource-constrained environments,

such as telemedicine platforms and mobile health applications.

The contributions of this study are twofold:

- We address data imbalance and overfitting using SMOTE and data augmentation techniques to enhance model generalizability.
- We compare the proposed EfficientNetB2 architecture to other CNN models, such as ResNet50 and GoogleNet, and demonstrated its superior performance with a lower computational burden.

To evaluate the robustness of our model, we conducted experiments on the HAM10000 dataset, which is one of the largest publicly available datasets for skin lesion classification. Through cross-validation and advanced regularization techniques (e.g., dropout and early stopping), we ensured reliable and reproducible performance.

The following sections provide a detailed description of the dataset, model architecture, and experimental results. We discuss the clinical implications of the proposed solutions and identify avenues for future research to improve the performance and scalability of AI-based diagnostic tools.

2. Methods

Fig. 1 shows a detailed flowchart of the entire experimental pipeline, from data preprocessing to final model evaluation.

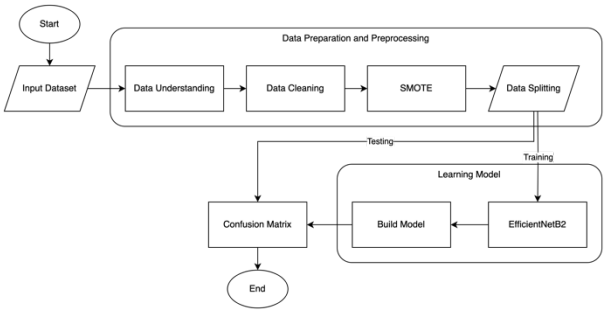

Fig 1. Experimental pipeline

2.1 Dataset and data preparation

In this study, the HAM10000 dataset was used, which consists of 11,644 dermoscopy images distributed into seven classes: melanoma (mel), basal cell carcinoma (bcc), actinic keratosis (akiec), benign keratosis (bkl), dermatofibroma (df), vascular lesion (vasc), and melanocytic nevus (nv), which can be accessed on kaggle under the title HAM10000 Preprocessing Datan (Table 1). Fig. 2 illustrates the unbalanced distribution of classes in the dataset. This imbalance, with some benign lesions over-represented, poses a risk of biased predictions toward these dominant classes.

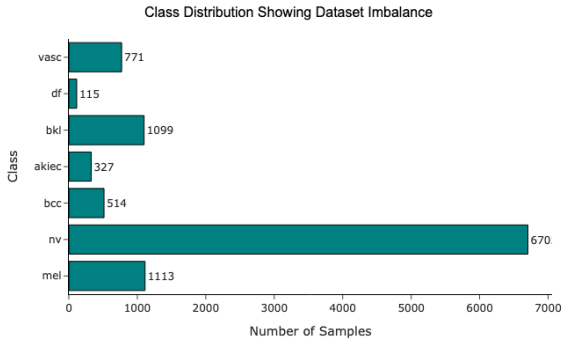

Fig 2. The classes unbalanced distribution in HAM10000 dataset

As you can see in Fig. 2 shows a visualization of the unbalanced class data in the dataset, which demonstrates the need for a class balancing method. To address this issue, class weighting was applied during training to ensure that minority classes contributed equally to the loss function (Roy et al., 2024). Additionally, SMOTE (Synthetic Minority Oversampling Technique) was employed to generate synthetic samples for underrepresented classes, which improves the model's generalization ability .

Table 1.Dataset class distribution and augmentation strategy.

CLASS	COUNT	AUGMENTATION	DESCRIPTION	IMAGES
Melona (mel)	1,113	Rotation, Contrast Adjustment	Highly malignant	
Nevus (nv)	6,705	Horizontal Flip, Zoom	Benign but visually similar to mel	
BCC (bcc)	514	Brightness Shift	Malignant skin tumor	
AKIEC (akiec)	327	Rotation	Pre-cancerous lesion	
BKL (bkl)	1,099	None	Benign keratosis	
DF (df)	115	None	Benign skin lesion	
VASC (vasc)	771	None	Vascular skin lesion	

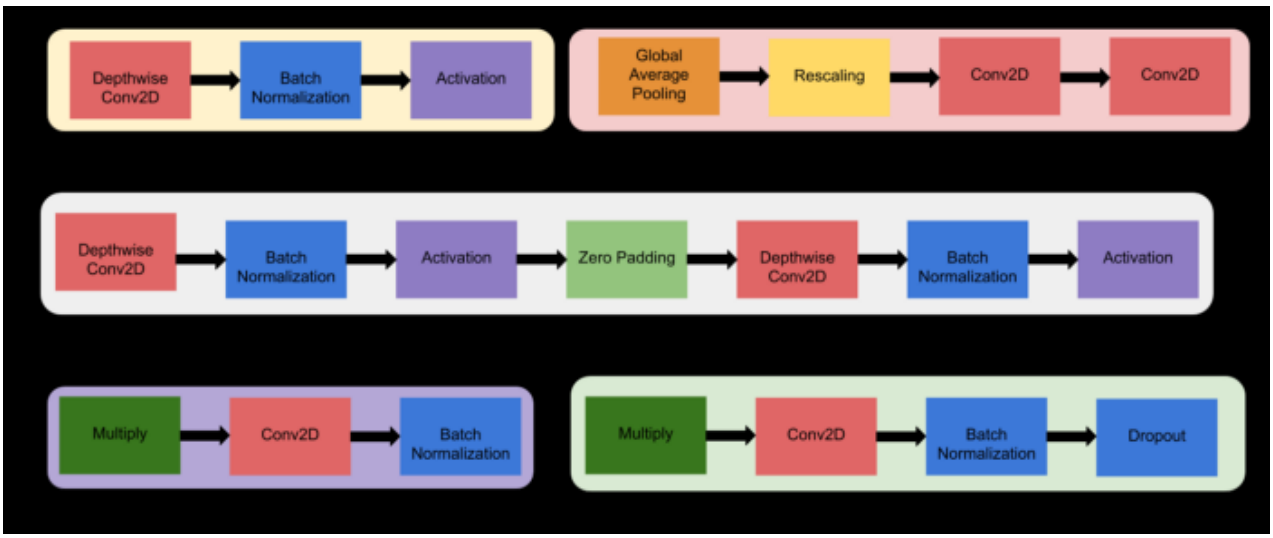

Fig 3. EfficientNetB2 model architecture

2.2 Model architecture and design

This study utilizes EfficientNetB2, a model renowned for its compound scaling, which efficiently balances network depth, width, and resolution efficiently (Ravi et al., 2021). EfficientNetB2 is particularly suitable for low-resource environments, such as mobile health platforms, where computational efficiency is critical. Fig. 3 shows the architecture of the proposed EfficientNetB2 model.

The architecture includes:

- Pre-trained Convolutional Layers: Initialized with ImageNet weights to leverage transfer learning.
- Dropout Layer (Rate: 0.5): This layer prevents overfitting by randomly deactivating neurons during training.
- Dense Layer (512 neurons): This layer extracts high-level abstract features.
- Softmax Output Layer: Provides probabilistic outputs across the seven categories.

This architecture provides an optimal trade-off between model complexity and computational efficiency, making it ideal for clinical deployment scenarios.

2.3 Training procedure and optimization

Table 2. Hyperparameters for model training.

HYPERPARAMETER	VALUES	DESCRIPTION
Learning rate	1e-4	Initial learning rate for Adam optimizer
Batch size	32	Number of samples processed per iteration
Epoch	50	Maximum training iterations
Dropout rate	0.5	Probability of neuron deactivation
Weight loss (L2)	0.001	Regularization to reduce overfitting

The training process used cross-entropy loss to measure the discrepancy between the predicted and actual labels. The optimization was performed using the Adam optimizer with a learning rate of 1e-4. Training was conducted over 50 epochs, and early stopping activated if validation loss did not improve for five consecutive epochs. The ReduceLROnPlateau callback dynamically reduced the learning rate by 20% if model performance increased. This carefully tuned configuration ensured that the model achieved high performance without overfitting, as validated through 5-fold cross-validation.

2.4 Cross-validation and regularization

To ensure robust performance, 5-fold cross-validation is used, this technique is motivated from research conducted by Guergueb and Akhloufi (2022). This method partitions the dataset into five equal subsets, rotating the validation subset across each fold. The model's final performance metrics were averaged across all five folds to account for any variability in the dataset. Pseudocode 1, illustrates the cross-validation process used in this study.

Pseudocode 1.

```
Input: D (HAM10000 Dataset), F (Number of Folds = 5)
Output: Averaged Metrics (Accuracy, Precision,
Recall, F1-score)

1: Split D into F folds {D1, D2, ..., DF}
2: Initialize metrics: Accuracy = 0, Precision = 0,
Recall = 0, F1-score = 0

3: for i = 1 to F do
4:     Validation_Set ← Di
5:     Training_Set ← D \ Di
6:
7:     // Train EfficientNetB2 on Training_Set
8:     Apply Data Augmentation (flip, rotate, zoom)
9:         Initialize EfficientNetB2 with ImageNet
weights
10:    Train model with Adam optimizer,
          learning rate = 1e-4, batch size = 32
11:     Monitor Validation Loss for Early Stopping
(patience = 5)
12:
13:    // Evaluate on Validation_Set
14:    Compute Accuracy, Precision, Recall, F1-score
15:    Accumulate metrics across folds
16: end for

17: // Compute final averaged metrics
18: Accuracy ← Accuracy / F
19: Precision ← Precision / F
```

```
20: Recall ← Recall / F
21: F1-score ← F1-score / F

22: return Averaged Metrics
```

This pseudocode illustrates the rotation of validation and training subsets during cross-validation to ensure reliable generalizability.

The following regularization techniques were applied:

- Dropout (0.5): Random neurons are deactivated during training to mitigate overfitting.
- L2 Regularization ($\lambda = 0.001$): This penalty procedure penalizes large weight magnitudes to prevent overcomplexity.
- Early stopping: This step halts training if validation performance does not improve, which reduces computational cost.

2.5 Evaluation metrics and model comparison

The performance of EfficientNetB2 was evaluated using multiple metrics to provide a comprehensive understanding of classification quality, such as accuracy (ACC), precision (PRE), recall (REC), F1-score (F1), and area under curve (AUC). The ACC is used to measure the proportion of correct predictions among all samples (Eq. (1)). PRE used to evaluate the number of correct predictions (Eq. (2)). REC used to measure the ability of the proposed model to detect all positive cases (Eq. (3)). The final F1-S model used balance precision and recall through a harmonic mean (Eq. (4)). AUC is used captures the trade-off between sensitivity and specificity across thresholds.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$PRE = \frac{TP}{TP+FP} \quad (2)$$

$$REC = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = \frac{PRE \times REC}{PRE + REC} \quad (4)$$

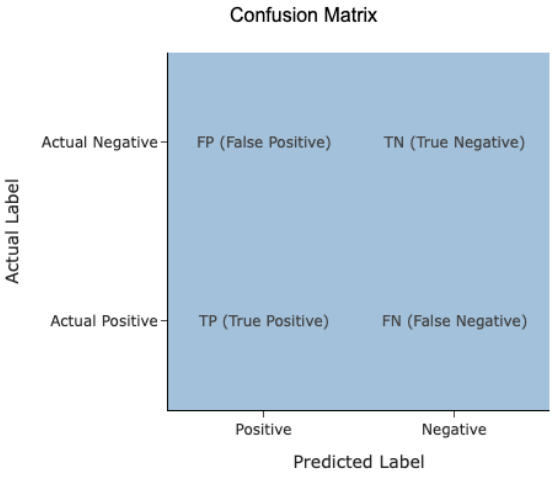

Fig 4. Confusion matrix and ROC curve for EfficientNetB2

The confusion matrix shows classification performance across all seven categories, with the receiver operating characteristic (ROC) curve highlighting the model's ability to differentiate between positive and negative cases.

2.6 Benchmarking against other models

For a comprehensive evaluation, EfficientNetB2 was compared against other widely adopted CNN architectures, namely ResNet50 and GoogleNet (Inception). ResNet50 is recognized for its depth and its ability to address the vanishing gradient problem through residual connections, making it a reliable choice for deep learning tasks. On the other hand, GoogleNet (Inception) is designed to handle multi-scale image analysis efficiently by utilizing inception modules that capture both local and global features simultaneously.

The comparison between these architectures focused not only on classification accuracy but also on their computational efficiency, which is crucial for real-world deployment, particularly in resource-constrained environments. Key performance indicators included inference time, the number of parameters (in millions), and memory usage during inference. These metrics provide a holistic view of each model's strengths and limitations, ensuring that the proposed EfficientNetB2 offers a balanced trade-off between accuracy and efficiency.

3. Results

This section presents the performance outcomes of the proposed EfficientNetB2 model on the HAM10000 dataset, including a detailed analysis of class-wise metrics, comparison of the performance of the proposed model to other state-of-the-art CNN architectures, statistical significance testing, and computational efficiency. The results highlight the model's suitability for real-world deployment, particularly in mobile health and telemedicine applications.

3.1 Model performance on the HAM10000 dataset

In Table 3, the EfficientNetB2 model demonstrated robust performance across multiple evaluation metrics. On the test set, the model achieved an accuracy of 86%, with a 95% Confidence Interval (CI) of [83.2%, 88.8%], indicating reliable generalizability to unseen data. In addition to accuracy, the model reported precision of 0.87, recall of 0.85, and an F1-score of 0.86, confirming a balanced trade-off between correctly identifying positive cases and minimizing false positives and negatives.

The model's area under the receiver operating characteristic curve (AUC) was 0.90, indicating excellent discrimination capability across various classification thresholds, even in the presence of class imbalance.

3.2 Performance metrics by class

A breakdown of the model's performance by class is presented in Table 3, which presents the precision, recall, and F1-scores for each lesion type. The model exhibited strong precision and recall for actinic keratosis (akiec) and basal cell carcinoma (bcc), both of which are clinically significant conditions that require timely diagnosis and treatment. However, some misclassification was observed

between melanoma (mel) and melanocytic nevus (nv), a common challenge because of the visual similarity between these lesion types.

Table 3. Performance cetrics by class with 95% CI.

CLASS	PRE (%)	REC (%)	F1 (%)	AUC
Melanoma (mel)	84.5 [82.0, 86.7]	79.2 [76.5, 81.9]	81.8 [78.9, 84.4]	0.89
Basal cell carcinoma (bcc)	89.7 [87.4, 91.8]	87.2 [85.1, 89.3]	88.4 [86.0, 90.6]	0.91
Actinic keratosis (akiec)	92.1 [90.3, 93.8]	89.3 [87.4, 91.1]	90.7 [88.5, 92.7]	0.93
Melanocytic nevus (nv)	80.3 [77.2, 83.0]	78.1 [75.5, 80.6]	79.2 [76.4, 81.8]	0.87
Other classes (df, bkl, vasc)	85.5 [82.5, 88.0]	84.8 [81.9, 87.3]	85.1 [82.0, 87.6]	0.88

Table 4. Comparison of models on HAM10000 dataset.

MODELS	ACC (%)	PRE (%)	REC (%)	F1 (%)	PARAMETERS (millions)	INFERENCE TIME (ms)
EfficientNetB2	86 [83.2, 88.8]	0.87	0.85	0.86	9.2	45
ResNet50	82 [79.1, 85.3]	0.84	0.81	0.82	25.6	110
GoogleNet (Inception)	83 [80.4, 85.9]	0.85	0.83	0.84	22.0	90

The model's performance was particularly robust in detecting akiec and bcc, which pose significant clinical risks if left untreated. However, the slight reduction in melanoma detection performance highlights an area for potential improvement, and further feature engineering or attention mechanisms are needed to improve the classification of ambiguous lesions.

3.3 Comparison with state-of-the-art CNN models

To comprehensively assess the performance of EfficientNetB2, it was benchmarked against two popular CNN architectures: ResNet50 and GoogleNet (Inception). The results are summarized in Table 4, focusing on both classification performance and computational efficiency.

As you can see in Table 4, show the EfficientNetB2 outperformed both ResNet50 and GoogleNet in terms of accuracy and F1-score while requiring fewer parameters and faster inference times. In particular, EfficientNetB2 achieved an inference time of 45 ms, which was approximately 2.4 times faster than ResNet50 and 2x faster than GoogleNet. This computational efficiency renders the proposed model highly suitable for resource-limited environments, such as mobile health platforms, where both speed and accuracy are essential.

3.4 Statistical significance testing

To determine the significance of the observed performance differences, paired t-tests were conducted between EfficientNetB2 and the other models. The results indicated that the improvements in accuracy and inference time achieved by EfficientNetB2 were statistically significant ($p < 0.05$) compared to ResNet50 and GoogleNet, confirming the robustness of the proposed architecture.

3.5 Confusion matrix and ROC curve analysis

A detailed analysis of the confusion matrix (Fig. 5) provided further insights into the model's misclassification patterns. The matrix revealed high classification accuracy for akiec and bcc, but misclassification errors occurred predominantly between melanoma (mel) and nevus (nv) due to morphological overlap. This suggests the need for more advanced feature extraction or attention-based layers in future model iterations.

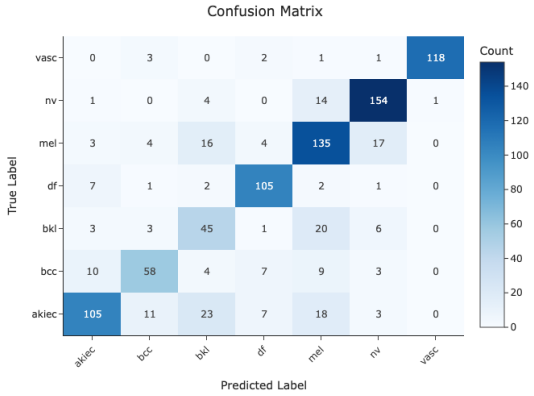

Fig 5. Confusion Matrix for EfficientNetB2 Model

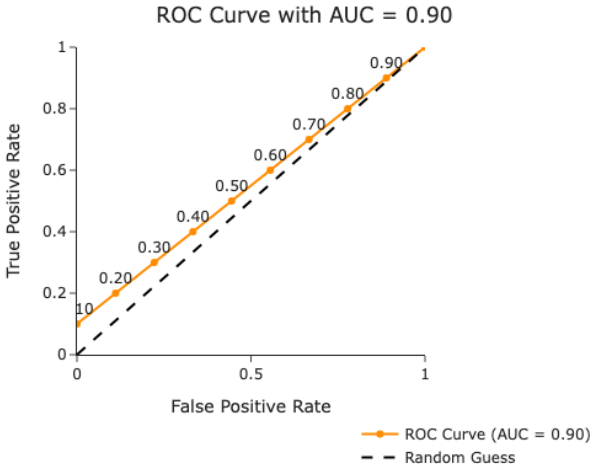

Fig 6. ROC curve with AUC value

The ROC curve (Fig. 6) for all seven classes demonstrated an AUC of 0.90, confirming the model's ability to perform well across varying classification thresholds.

3.6 Computational efficiency

The computational efficiency of the proposed model was evaluated to assess its feasibility for real-time applications. EfficientNetB2, with 9.2 million parameters, required 45 ms for inference, outperforming both ResNet50 and GoogleNet in terms of speed and resource usage. These results highlight the potential of EfficientNetB2 for deployment in telemedicine platforms and mobile health applications, where both speed and model size are critical constraints.

4. Discussions

This study demonstrated the potential of EfficientNetB2 as a lightweight, accurate, and computationally efficient deep learning model for skin cancer detection, addressing critical challenges such as overfitting and class imbalance, which often limit the practical deployment of AI in healthcare. The results demonstrate the viability of EfficientNetB2 in resource-constrained environments, including mobile health platforms and telemedicine applications, where rapid and reliable diagnostic support is crucial.

4.1 Clinical implications

The successful performance of EfficientNetB2 highlights its potential to enhance clinical workflows by functioning as a decision support tool for dermatologists. In practice, this model can be deployed as a prescreening or triaging tool to prioritize high-risk patients and reduce clinician workload. Early detection of malignant lesions, such as melanoma and actinic keratosis, is essential, and the model's high precision for these classes underscores its clinical value in detecting critical cases that require immediate attention. Furthermore, the proposed model, with its small memory footprint and fast inference times, is suitable for telemedicine initiatives in underserved regions where access to specialized dermatologists may be limited. EfficientNetB2's adaptability also enables it to be integrated into mobile applications, providing real-time predictions that can assist non-specialists and primary care providers in making timely referrals.

4.2 Comparison with state-of-the-art models

In this study, EfficientNetB2 was compared directly with ResNet50 and GoogleNet (Inception), two well-established CNN architectures. Although these models have demonstrated strong classification performance in medical imaging, they exhibit certain limitations, particularly computational efficiency and model size. EfficientNetB2 achieved higher accuracy (86%) with fewer parameters (9.2M) and shorter inference times (45 ms) than both ResNet50 and GoogleNet. These results align with recent AI model optimization trends, which emphasize the need to balance performance with computational demands, especially in healthcare environments where resources are limited (Li et al., 2022; Ravi et al., 2021). The findings also demonstrate the potential of EfficientNetB2 to address the overfitting and generalization challenges identified in previous studies involving complex CNN architectures (Goceri & Karakas, 2020).

4.3 Limitations and challenges

Although the results are promising, this study acknowledges several limitations that need to be addressed in future research. First, the HAM10000 dataset is comprehensive but does not fully capture the diversity of skin lesions encountered in clinical practice. Differences in patient demographics, imaging conditions, and lesion morphology across populations can affect model performance in real-world settings. Future studies should validate the proposed model on multiple datasets—such as ISIC and PH2—to ensure broader generalizability (Hellín et al., 2024; Ragupathi et al., 2022). In addition, the class imbalance problem persists despite the use of SMOTE and data augmentation. Although these methods reduce bias, further improvements can be achieved by incorporating GAN-based data augmentation, generating synthetic images to enrich underrepresented classes, such as rare malignancies (Ragupathi et al., 2022; Wang et al., 2021).

Another challenge lies in the misclassification of melanoma as melanocytic nevus, which reflects the morphological similarities between the two types. This issue could have significant clinical implications because misdiagnosing melanoma could delay life-saving interventions. To address this limitation, future models could integrate attention mechanisms or feature extraction modules that focus on subtle differences in texture and pigmentation. For example, hybrid models combining CNNs with Transformer-based architectures may enhance the model's ability to capture fine-grained features, thereby improving differentiation between ambiguous lesions (He et al., 2023).

4.4 Future directions

To further enhance the practical applicability of this model, clinical trials are recommended to evaluate its impact on real-world diagnostic workflows. Such trials could assess the model's effectiveness in reducing diagnostic delays, improving patient outcomes, and enhancing clinician satisfaction. Moreover, ensemble learning approaches—combining EfficientNetB2 with other architectures, such as ResNet and DenseNet could offer complementary strengths, leading to more robust predictions, particularly for rare and complex cases.

Another promising direction is the integration of self-attention mechanisms into the EfficientNetB2 architecture. Recent studies have demonstrated that attention-based layers can help models focus on relevant visual regions in medical images, which could mitigate misclassification risks (Anwar, 2023). Furthermore, the deployment of this model on edge devices—such as smartphones—offers exciting possibilities for real-time diagnostics, particularly in telemedicine settings. However, regulatory and ethical considerations must be carefully addressed to ensure patient privacy and data

security, especially when deploying AI models in healthcare environments.

4.5 Ethical and practical considerations

As AI models like EfficientNetB2 become integrated into healthcare workflows, it is essential to address ethical concerns related to algorithmic bias and transparency. Although the model performed well on the HAM10000 dataset, algorithmic biases may emerge when applied to different patient populations. Therefore, continuous monitoring and validation across diverse clinical environments are required to prevent unintended biases. Transparency is also critical; clinicians must trust the predictions made by AI models to incorporate them effectively into patient care. Providing explainability tools, such as SHAP (SHapley Additive exPlanations) values or attention maps, could help bridge the gap between AI predictions and clinician trust, thus improving decision-making processes.

From a practical perspective, it is important to ensure that AI-based diagnostic tools comply with regulatory frameworks, such as the General Data Protection Regulation (GDPR) and other healthcare-specific guidelines. Robust data anonymization techniques should be employed to protect patient privacy when training and deploying models in telemedicine environments.

5. Conclusion

This study demonstrated the potential of EfficientNetB2 as a reliable tool for skin cancer detection, achieving 86% accuracy with high precision and recall for certain high-risk lesions, such as actinic keratoses (akiec) and basal cell carcinoma (bcc). The model's lightweight architecture and computational efficiency make it particularly suitable for deployment in resource-limited environments, such as mobile health platforms and telemedicine systems, where rapid and accurate preliminary diagnoses are essential.

Despite these promising results, the study identified several challenges, including overfitting after the 8th epoch and misclassification of melanoma and nevus, which could affect patient outcomes in real-world clinical settings. To address these limitations, future work should explore advanced regularization techniques, such as GAN-based data augmentation, ensemble models, and self-attention mechanisms, to improve classification performance, particularly for ambiguous lesion types.

Furthermore, cross-dataset validation using datasets like ISIC or PH2 will be critical to ensure the model's generalizability across diverse clinical scenarios. The integration of EfficientNetB2 into dermatological workflows requires not only technological optimization but also ethical frameworks to maintain transparency, reduce bias, and safeguard patient privacy.

In summary, EfficientNetB2 has significant promise for improving the efficiency and accuracy of skin cancer detection. Further refinement and clinical validation are essential for developing reliable dermatological tools. With continued research and development, this model can enhance early detection, optimize diagnostic workflows, and ultimately improve patient outcomes through timely intervention.

(Muhammad Wigig Purbandanu)

Authors' Contributions

Muhammad Wigig Purbandanu was responsible for the conceptualization and design of the study, developing the methodology, curating the dataset, and drafting the initial manuscript. Rizky Yanuarta and Arif Kurniawan focused on data collection and analysis, ensuring the integrity and accuracy of the dataset used for model training and evaluation. Muhammad Munsarif supervised the research project, managed administrative tasks, and secured funding to support the study. Ahmad Ilham contributed to the refinement of the methodology, provided technical support throughout the project, and actively participated in the review and editing of the manuscript. Ayomikun A. Awoseyi conducted the literature review, developed the theoretical framework, performed comparative model evaluations, and contributed significantly to the review and editing of the final manuscript to ensure clarity and coherence.

Acknowledgement

We would like to thank the Faculty of Engineering, Universitas Muhammadiyah Semarang for its support and encouragement throughout the process of conducting this study.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Anwar, R. S. S. (2023). EfficientNet Algorithm for Classification of Different Types of Cancer. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA32021004

Goceri, E., & Karakas, A. A. (2020). Comparative evaluations of CNN based networks for skin lesion classification. *14th International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP), Zagreb, Croatia*, 1–6.

Guergueb, T., & Akhloufi, M. A. (2022). Skin Cancer Detection using Ensemble Learning and Grouping of Deep Models. *International Conference on Content-Based Multimedia Indexing*, 121–125. https://doi.org/10.1145/3549555.3549584

He, K., Gan, C., Li, Z., Rekik, I., Yin, Z., Ji, W., Gao, Y., Wang, Q., Zhang, J., & Shen, D. (2023). Transformers in medical image analysis. *Intelligent Medicine*, *3*(1), 59–78. https://doi.org/10.1016/j.imed.2022.07.002

Hellín, C. J., Olmedo, A. A., Valledor, A., Gómez, J., López-Benítez, M., & Tayebi, A. (2024). Unraveling the Impact of Class Imbalance on Deep-Learning Models for Medical Image Classification. *Applied Sciences*, *14*(8), 3419. https://doi.org/10.3390/app14083419

Li, Z., Koban, K. C., Schenck, T. L., Giunta, R. E., Li, Q., & Sun, Y. (2022). Artificial Intelligence in Dermatology Image Analysis: Current Developments and Future Trends. *Journal of Clinical Medicine*, *11*(22), 6826. https://doi.org/10.3390/jcm11226826

Parker, E. R. (2021). The influence of climate change on skin cancer incidence – A review of the evidence. *International Journal of Women's Dermatology*, *7*(1), 17–27. https://doi.org/10.1016/j.ijwd.2020.07.003

Ragupathi, T., Govindarajan, M., & Priyaradhikadevi, T. (2022). Class Imbalance Handling with Deep Learning Enabled IoT Healthcare Diagnosis Model. *Intelligent Automation & Soft Computing*, *34*(2), 1351–1366. https://doi.org/10.32604/iasc.2022.025756

Ravi, V., Narasimhan, H., & Pham, T. D. (2021). *EfficientNet-Based Convolutional Neural Networks for Tuberculosis Classification* (pp. 227–244). https://doi.org/10.1007/978-3-030-69951-2_9

Roy, D., Roy, A., & Roy, U. (2024). *Learning from Imbalanced Data in Healthcare: State-of-the-Art and Research Challenges* (pp. 19–32). https://doi.org/10.1007/978-981-99-8853-2_2

Wang, J., Zhu, H., Wang, S.-H., & Zhang, Y.-D. (2021). A Review of Deep Learning on Medical Image Analysis. *Mobile Networks and Applications*, *26*(1), 351–380. https://doi.org/10.1007/s11036-020-01672-7