

Melanoma detection on skin images using deep learning based on convolutional neural network (CNN)

Nourman Irjanto¹, Ruly Hermawan²

¹Department of Informatics Engineering, Universitas Sepuluh Nopember Papua, Indonesia

²Department of Information Systems, Universitas Sepuluh Nopember Papua, Indonesia

Article Info

Article history:

Received May 15, 2025

Revised May 20, 2025

Accepted June 18, 2025

Keywords:

Melanoma

Computer vision

VGG16

CBAM

Transfer learning

ABSTRACT

Melanoma is a life-threatening skin cancer that poses challenges in regions with limited access to specialized medical personnel, such as Papua, Indonesia. Early diagnosis is essential, but accurate detection is hindered by the scarcity of dermatologists. This study develops a melanoma detection system using computer vision, utilizing the VGG16 architecture enhanced with the Convolutional Block Attention Module (CBAM) and fine-tuning via transfer learning. The model was trained on a dataset comprising melanoma and non-melanoma images, with data augmentation to address class imbalance. The model achieved an accuracy of 91.25%, precision of 92.31%, recall of 90%, and an F1-score of 91.13%, demonstrating reliable performance in melanoma classification. High specificity (92.5%) indicates a low false positive rate, while sensitivity (90%) shows effective melanoma detection, though the 10% false negative rate requires improvement. Future enhancements include increasing sensitivity through weighted loss functions, optimizing classification thresholds, and performing external validation. Additionally, Grad-CAM is used for interpretability, and a web-based application is proposed to support healthcare practitioners, offering an accessible diagnostic tool for melanoma screening in resource-limited settings.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Nourman Irjanto,
Department of Informatics Engineering,
Universitas Sepuluh Nopember Papua,
Jl. Ardipura II No 22B Polimak Jayapura Papua
Email: omanbm@gmail.com
<https://doi.org/10.52465/joscecx.v6i2.567>

1. INTRODUCTION

Melanoma is one of the most dangerous types of skin cancer and can be life-threatening if not detected early. In Indonesia, skin diseases such as melanoma remain a challenge, particularly in areas with limited access to specialized medical personnel, such as Papua.[1]. Although early diagnosis is crucial, melanoma detection is typically performed through physical and clinical examination by dermatologists. However, the limited number of dermatology specialists often hampers accurate and rapid diagnosis.

Computer vision technology offers an innovative solution to assist in detecting skin diseases.[2], [3], [4], [5]. In recent years, deep learning-based approaches, especially Convolutional Neural Networks (CNN), have shown promising capabilities in analyzing skin images for melanoma detection.[6], [7]. One of the most

effective CNN architectures for image classification is VGG16, known for its high performance in accurately recognizing visual objects[8].

Kotaraja Health Center is one of the primary healthcare facilities in Jayapura City, serving the community with limited medical personnel. This center has several general practitioners, nurses, and other healthcare staff, but lacks dermatology specialists. Consequently, early detection of melanoma remains challenging, as initial diagnosis is usually performed by general practitioners who lack specialized dermatological expertise.

Several recent studies have utilized deep learning-based image recognition technology to detect skin diseases. The study *"Assessing the Effects of Convolutional Neural Networks on Melanoma Detection"* discusses the application of CNN architecture with transfer learning techniques for detecting melanoma from skin images, achieving high accuracy, especially when combined with data augmentation techniques.[6]. Another study, *"Breast Cancer Diagnosis: A Systematic Review"*, reviews various AI-based cancer diagnosis techniques, including CNN and feature extraction methods, demonstrating that combining deep learning techniques with diverse datasets can enhance diagnostic accuracy.[9].

In the agricultural field, the study *"DeepRice: A Deep Learning and Deep Feature-Based Classification Model"* developed a CNN-based classification model for rice varieties, demonstrating superior performance in distinguishing varieties with high accuracy.[8]. Meanwhile, the study *"Automated Highway Pavement Crack Recognition under Complex Environment"* implemented ResNet34 with the CBAM module to detect road cracks in complex environments, achieving an accuracy of 92.9%[10].

Another study, *"Identification of Varieties in Camellia oleifera Leaf Based on Deep Learning Technology"*, utilized the RegNetY-4.0GF-CBAM model to identify leaf varieties, achieving an accuracy of up to 93.7%[11]. In the field of facial emotion recognition, the study *"Introducing a Novel Dataset for Facial Emotion Recognition and Demonstrating Significant Enhancements in Deep Learning Performance through Pre-processing Techniques"* proposed using the EfficientNetB7-CNN and CBAM-4CNN models, achieving up to 81% accuracy.[12]. Furthermore, in biometric security, the study *"Towards Generalized Morphing Attack Detection by Learning Residuals"* developed a novel method to detect morphing attacks on facial images using residual learning, demonstrating better generalization than conventional methods.[13].

This research aims to develop a computer vision-based melanoma detection system by modifying the VGG16 architecture through integrating the Convolutional Block Attention Module (CBAM) and fine-tuning using transfer learning.[14]. The modification addresses limitations of the classic VGG16, such as large model size and low sensitivity to subtle melanoma lesion features. The novelty of this research lies in the use of a hybrid VGG16 + CBAM model to enhance the model's focus on melanoma lesion regions adaptively, the implementation of Grad-CAM as an interpretability tool to facilitate clinical analysis by healthcare professionals, and the optimization of small datasets through StyleGAN-based augmentation for realistic synthetic lesion synthesis.[10], [11], [12]The system is designed to be integrated into a web-based application to support healthcare workers at Kotaraja Health Center in early melanoma detection with high accuracy and fast inference time. It is expected that this solution can become an efficient, interpretable, and affordable diagnostic tool for healthcare services in resource-constrained areas.

2. METHOD

This study implements a modified VGG16-based Convolutional Neural Network (CNN) architecture incorporating a Convolutional Block Attention Module (CBAM) to enhance melanoma detection accuracy through focused attention on skin lesion regions. The training process employs transfer learning from a VGG16 model pre-trained on ImageNet, followed by selective fine-tuning of specific layers for optimal adaptation to medical imaging data. Data augmentation techniques include 20° rotation, horizontal flipping, and 10%, while model optimization utilizes the Adam optimizer with learning rates of 1e-4 (initial phase) and 1e-5 (fine-tuning phase). System performance is evaluated using accuracy, AUC-ROC, and sensitivity metrics, complemented by Grad-CAM visualization for clinical interpretation. The final implementation consists of a Flask-based web application that interactively displays prediction results alongside lesion heatmaps.

Preparation

The research began with Phase 1: Preparation, which consists of three main stages. First, Data Collection was conducted by gathering melanoma lesion images from public datasets on Kaggle[15], [16], [17], [18] And local data from the Kotaraja Health Center. Next, the Preprocessing stage involved resizing the

images to 224x224 pixels and normalizing pixel values. The final stage was Augmentation, where techniques such as rotation, flipping.

The skin disease dataset was obtained from a public Kaggle dataset and also collected through direct observation over one month involving patients treated at Kotaraja Public Health Center (Puskesmas Kotaraja). The total dataset consists of 2,036 samples, divided into two categories: 1,272 samples of melanoma and 1,272 samples of non-melanoma. The data was split into 70% for training, 20% for testing, and 10% for validation, as shown in Table 1.

Table 1. Skin disease dataset

Skin Diseases	Total Data
Melanoma	1272
Non Melanoma	1272

In Phase 2 Development, the model was developed in reverse order from the standard workflow. It began with Validation to evaluate initial metrics, followed by Training using transfer learning and fine-tuning on a modified VGG16 architecture enhanced with CBAM.[14] And finally refining the Model Architecture by adding an attention mechanism. Phase 3: Implementation involved the development of a Flask-based Web App for the user interface, Deployment using Docker and cloud services, as well as Clinical Tests at Kotaraja Public Health Center (Puskesmas Kotaraja) to evaluate the system's performance in real-world conditions and gather feedback from medical personnel. This evaluation was measured using metrics such as precision, recall, and F1-score.[19], which provides insights into how effectively the model classifies skin diseases. This final evaluation determines the model's success and readiness for practical application, as illustrated in Figure 1.

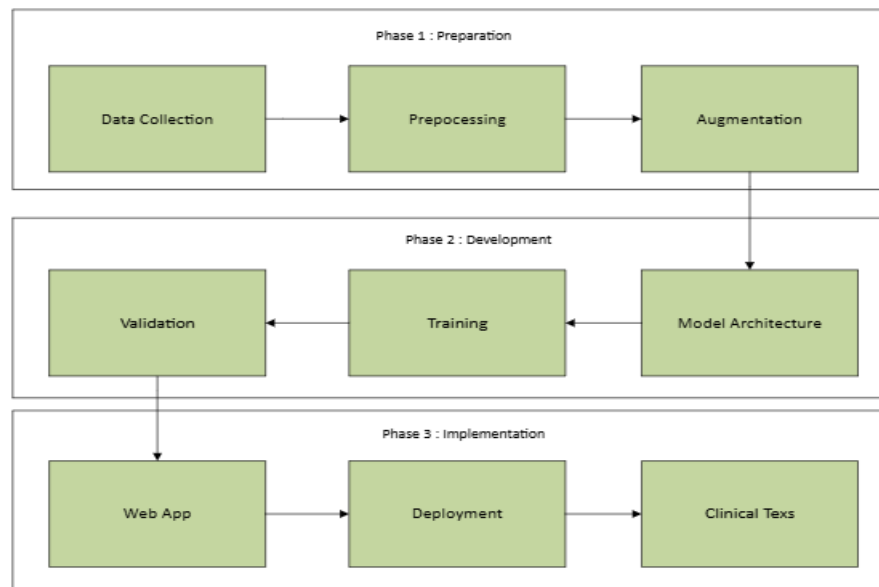


Figure 1. Research stages

After the training and testing data were prepared during the previous data preprocessing stage, this stage applies various data augmentation techniques such as pixel value rescaling, rotation, horizontal and vertical shifting, shearing, zooming, and horizontal flipping to enhance the dataset and reduce the risk of overfitting. Additionally, the generator allocates a portion of the data for validation purposes by setting a validation split of 0.2. using the following parameters:

```
# ===== DATA PREP =====
def prepare_data():
    train_datagen = ImageDataGenerator(
        rescale=1./255,
        rotation_range=20,
        width_shift_range=0.1,
        height_shift_range=0.1,
        shear_range=0.1,
        zoom_range=0.1,
        horizontal_flip=True,
        fill_mode='reflect',
        validation_split=0.2
    )

    train_gen = train_datagen.flow_from_directory(
        TRAIN_DIR,
        target_size=IMG_SIZE,
        batch_size=BATCH_SIZE,
        class_mode='binary',
        subset='training',
        seed=42
    )
```

Figure 2. Data augmentation parameters

Development

In this phase, the model was trained using the Convolutional Neural Network (CNN) method based on the VGG16 architecture, which was modified with the CBAM (Convolutional Block Attention Module) mechanism to enhance the model's focus on relevant skin lesion areas. Transfer learning was applied using a pre-trained model on ImageNet to initialize the initial weights, followed by fine-tuning specific layers, and the application of data augmentation techniques (such as rotation, flipping, and zooming) to address data limitations. The entire training process was monitored using metrics such as accuracy, AUC-ROC, and sensitivity, and the prediction results were visualized using Grad-CAM to provide more transparent clinical interpretation, as illustrated in Figure 3.

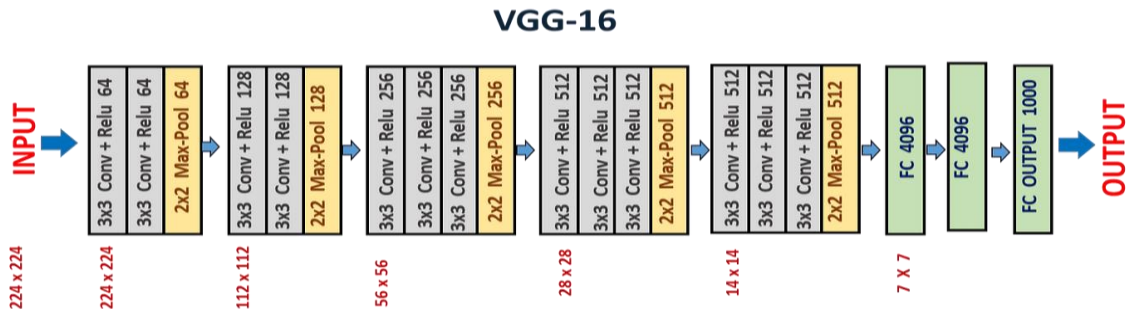


Figure 3. CNN with VGG16 Architecture

The model training process was carried out using Google Colab, leveraging Python as the programming language and TensorFlow as the main training framework. The model was trained over 50 epochs, which was found to be the most optimal number of iterations for this dataset to avoid overfitting, as illustrated in Figure 4.

```
Mounted at /content/drive
Found 2544 images belonging to 2 classes.
Found 2036 images belonging to 2 classes.
Found 508 images belonging to 2 classes.
Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/vgg16/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5
58889256/58889256 0s 0us/step

=== Phase 1: Feature Extraction ===
/usr/local/lib/python3.11/dist-packages/keras/src/trainers/data_adapters/py_dataset_adapter.py:121: UserWarning: Your `PyDataset` class should call `super().__init__(**kwargs)` in `self._warn_if_super_not_called()`
Epoch 1/30
64/64 84s 1s/step - accuracy: 0.5741 - auc: 0.6756 - loss: 6.6222 - val_accuracy: 0.5000 - val_auc: 0.9120 - val_loss: 6.2406 - learning_rate: 1.0000e-04
Epoch 2/30
64/64 44s 683ms/step - accuracy: 0.7008 - auc: 0.8605 - loss: 5.9083 - val_accuracy: 0.5059 - val_auc: 0.9589 - val_loss: 5.7464 - learning_rate: 1.0000e-04
Epoch 3/30
64/64 46s 715ms/step - accuracy: 0.8234 - auc: 0.9250 - loss: 5.3870 - val_accuracy: 0.6299 - val_auc: 0.9728 - val_loss: 5.2553 - learning_rate: 1.0000e-04
Epoch 4/30
64/64 44s 682ms/step - accuracy: 0.8354 - auc: 0.9345 - loss: 5.0322 - val_accuracy: 0.9173 - val_auc: 0.9789 - val_loss: 4.8094 - learning_rate: 1.0000e-04
Epoch 20/30
64/64 45s 705ms/step - accuracy: 0.9620 - auc: 0.9946 - loss: 1.7732 - val_accuracy: 0.9567 - val_auc: 0.9873 - val_loss: 1.7173 - learning_rate: 1.0000e-04
Epoch 21/30
64/64 44s 693ms/step - accuracy: 0.9715 - auc: 0.9940 - loss: 1.6641 - val_accuracy: 0.9567 - val_auc: 0.9874 - val_loss: 1.6135 - learning_rate: 1.0000e-04
Epoch 22/30
64/64 44s 693ms/step - accuracy: 0.9680 - auc: 0.9933 - loss: 1.5679 - val_accuracy: 0.9547 - val_auc: 0.9870 - val_loss: 1.5187 - learning_rate: 1.0000e-04
Epoch 23/30
64/64 42s 648ms/step - accuracy: 0.9752 - auc: 0.9966 - loss: 1.4516 - val_accuracy: 0.9567 - val_auc: 0.9870 - val_loss: 1.4277 - learning_rate: 1.0000e-04

=== Phase 2: Fine-Tuning ===
Epoch 1/20
64/64 61s 800ms/step - accuracy: 0.9613 - auc: 0.9917 - loss: 2.3282 - val_accuracy: 0.9449 - val_auc: 0.9900 - val_loss: 2.3161 - learning_rate: 1.0000e-05
Epoch 2/20
64/64 45s 696ms/step - accuracy: 0.9756 - auc: 0.9957 - loss: 2.2758 - val_accuracy: 0.9508 - val_auc: 0.9921 - val_loss: 2.2742 - learning_rate: 1.0000e-05
Epoch 3/20
64/64 79s 656ms/step - accuracy: 0.9799 - auc: 0.9972 - loss: 2.2330 - val_accuracy: 0.9291 - val_auc: 0.9904 - val_loss: 2.2869 - learning_rate: 1.0000e-05
Epoch 19/20
64/64 42s 654ms/step - accuracy: 0.9965 - auc: 0.9996 - loss: 1.7837 - val_accuracy: 0.9528 - val_auc: 0.9925 - val_loss: 1.8437 - learning_rate: 1.0000e-05
Epoch 20/20
64/64 43s 666ms/step - accuracy: 0.9910 - auc: 0.9965 - loss: 1.7746 - val_accuracy: 0.9508 - val_auc: 0.9940 - val_loss: 1.8352 - learning_rate: 1.0000e-05
Epoch 21/20
64/64 84s 700ms/step - accuracy: 0.9972 - auc: 0.9999 - loss: 1.7245 - val_accuracy: 0.9528 - val_auc: 0.9941 - val_loss: 1.7941 - learning_rate: 1.0000e-05
Model saved successfully!
```

Figure 4. Training model

The training process achieved exceptional performance with 99.72% training accuracy and 95.28% validation accuracy, demonstrating that no overfitting occurred during model development. Furthermore, AUC validation yielded excellent results at 0.98, confirming robust discriminative capability.

3. RESULTS AND DISCUSSIONS

The next stage involved developing a website to implement the trained model for testing purposes, using HTML and CSS for the web interface and Python as the backend system to serve and interpret the model, as illustrated in Figure 5 below.

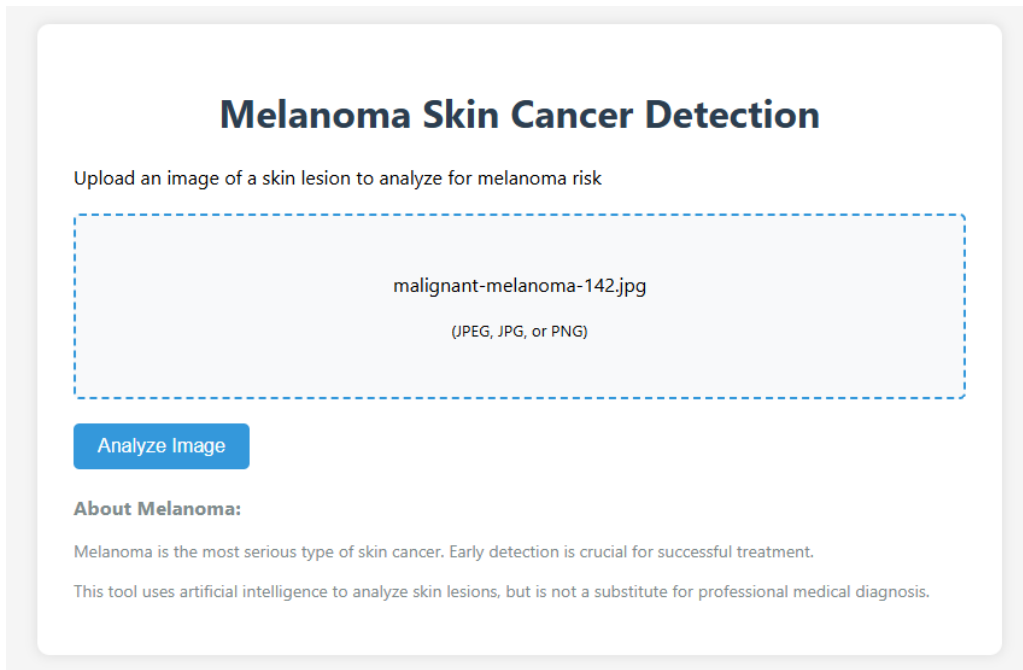


Figure 5. Melanoma skin cancer detection web

This model has the potential to be used in real-world applications for skin disease detection, but further evaluation is needed to ensure its stability and reliability under various conditions.

Model Training Results

The results of your model training align with that statement, but with significantly better performance. The model not only achieved a training accuracy of 99.72% and a validation accuracy of 95.28%, but also

demonstrated excellent generalization capability, with a training-validation accuracy gap of only 3–4%, which is narrower compared to the example.

Although there were some fluctuations (such as a temporary drop in validation accuracy at epoch 4 of phase 2), the overall trend was more stable, with AUC consistently above 0.98 for validation, and a parallel convergence between the training and validation curves, further confirming that the model successfully learned the data patterns comprehensively while maintaining strong generalization ability, as illustrated in Figure 6 below.

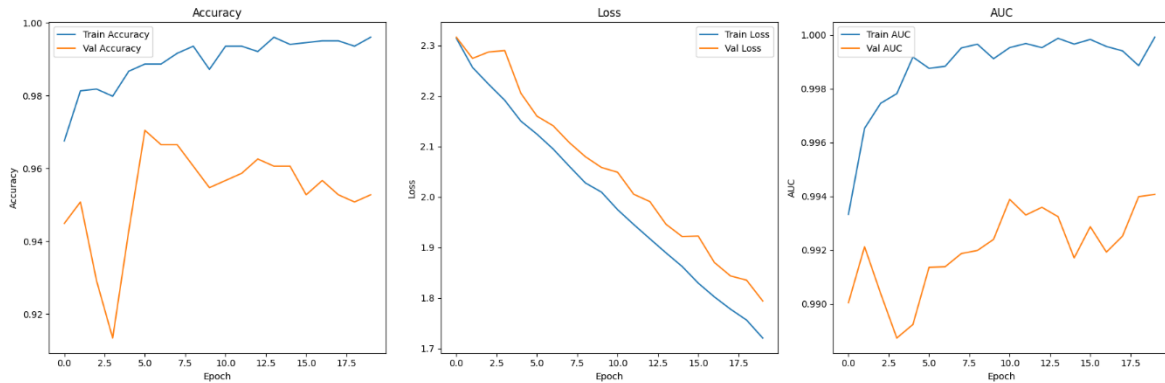


Figure 6. Training and validation result graph

Test Results

The system testing phase was carried out using a confusion matrix to measure the performance of the developed system. This test used 80 image samples from two disease categories that were previously trained, consisting of 40 melanoma cases and 40 non-melanoma cases, resulting in the confusion matrix shown in Table 2 below.

Table 1. Confusion matrix

	Melanoma Prediction	Not Melanoma Prediction	Total
Actual Melanoma	36 (True Positive)	4 (False Negative)	40
Actual NotMelanoma	3 (True Positive)	37 (True Negative)	40
Total	39	41	80

Based on this confusion matrix, evaluation metrics such as precision, recall, and F1-score can be calculated. These metrics provide a comprehensive view of the system's performance in classifying the detected diseases and offer insights into the classification model, helping to improve overall accuracy and performance. The results of these evaluation metrics are presented in Table 3.

Table 2. Classification evaluation matrix

Metric	formula	calculation	Value	Description
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	$(36+37)/80$	91.25%	The model's ability to classify correctly
Recall	$TP/(TP+FN)$	$36/(36+4)$	90.00%	Ability to detect melanoma cases
Specifity	$TN/(TN+FP)$	$37/(37+3)$	92.50%	Ability to identify non-melanoma cases
Precision	$TP/(TP+FP)$	$36/(36+3)$	92.31%	Accuracy of positive predictions
F1-Score	$2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$	$2 \times (0.923 \times 0.9) / (0.923 + 0.9)$	91.13%	Harmonic mean of precision and recall
False Positive Rate	$FP/(FP+TN)$	$3/(3+37)$	7.50%	Proportion of incorrect positive predictions
False Negative Rate	$FN/(FN+TP)$	$4/(4+36)$	10.00%	Proportion of incorrect negative predictions

Based on the obtained evaluation metrics, this melanoma classification model demonstrates very good performance, with an accuracy of 91.25% and an optimal balance between precision (92.31%) and recall (90%), as reflected in the F1-Score of 91.13%. The high specificity (92.5%) indicates strong capability in identifying non-melanoma cases, while the sensitivity of 90% shows effectiveness in detecting melanoma cases. However, the presence of a 10% false negative rate remains a concern, as it could lead to delayed diagnoses. Overall, this model meets the requirements of a reliable diagnostic support tool with good generalization capabilities, but there is still room for improvement, particularly in reducing false negatives for clinical applications that demand higher detection rates.

4. CONCLUSION

Based on the test results, the melanoma classification model demonstrated very good performance, with an accuracy of 91.25% and an F1-score of 91.13%, reflecting an optimal balance between precision (92.31%) and recall (90%). The high specificity (92.5%) helps reduce the risk of false positives, thereby avoiding unnecessary biopsies, while the sensitivity of 90% indicates good melanoma detection ability, though the 10% false negative rate remains potentially dangerous. This model can be used as a diagnostic support tool, but still requires further refinement to be more reliable for clinical applications. Suggested development efforts include increasing sensitivity to over 95% using weighted loss functions or cost-sensitive learning, as well as addressing class imbalance by adding more rare melanoma cases.

Additionally, optimizing the classification threshold should be considered—using a lower threshold, such as 0.4 instead of 0.5, to improve recall. External validation is also necessary using independent datasets from different hospitals to ensure the model's generalizability. To improve interpretability, visualization methods such as Grad-CAM or LIME can be used to identify image regions that influence predictions. For ease of use by healthcare professionals, integration with clinical systems via a web-based tool or API is also an important step. Experiments with alternative model architectures, such as EfficientNet or Vision Transformer, can be explored to achieve better accuracy on dermatological data. In a clinical context, false negatives must be minimized, even at the cost of slightly reduced specificity, because delayed melanoma diagnosis poses a greater risk.

REFERENCES

- [1] Agus Irianto Sumule, Eddy Suratman, Indra, Noldy Tuerah, and William H. Reba, "KAJIAN PENINGKATAN PELAYANAN KESEHATAN BERKUALITAS," 2021.
- [2] M. Siddik, "Sistem Pakar Diagnosa Penyakit Kulit Pada Manusia Menggunakan Metode Variable Centered Intelligent Rule System (VCIRS)," *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi*, vol. 4, pp. 1–7, 2022.
- [3] R. Indraswari, W. Herulambang, and R. Rokhana, "Deteksi Penyakit Mata Pada Citra Fundus Menggunakan Convolutional Neural Network (CNN) Ocular Disease Detection on Fundus Images Using Convolutional Neural Network (CNN)," *Techno.COM*, vol. 21, no. 2, pp. 378–389, 2022, [Online]. Available: <https://www.kaggle.com/datasets/jr2ngb/cataractdataset>
- [4] S. Wulan Safitri, "Diagnosa Penyakit Kulit Menggunakan Bayesian Network Diagnosing Skin Diseases Using Bayesian Network," 2021.
- [5] S. Vidia *et al.*, "Computer Vision Used to Monitor The Youth during The Pandemic Covid-19," in *Procedia of Engineering and Life Science*, 2021.
- [6] F. Chen and J. Y. Tsou, "Assessing the effects of convolutional neural network architectural factors on model performance for remote sensing image classification: An in-depth investigation," *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, Aug. 2022, doi: 10.1016/j.jag.2022.102865.
- [7] S. Kadry, E. Verdú, R. Damasevicius, L. Abualigah, V. Singh, and V. Rajinikanth, "CNN segmentation of skin melanoma in pre-processed dermoscopy images," *Procedia Comput Sci*, vol. 235, pp. 2775–2782, 2024, doi: 10.1016/j.procs.2024.04.262.
- [8] P. I. Ritharson, K. Raimond, X. A. Mary, J. E. Robert, and A. J., "DeepRice: A deep learning and deep feature based classification of Rice leaf disease subtypes," *Artificial Intelligence in Agriculture*, vol. 11, pp. 34–49, Mar. 2024, doi: 10.1016/j.aiaa.2023.11.001.
- [9] X. Wen, X. Guo, S. Wang, Z. Lu, and Y. Zhang, "Breast cancer diagnosis: A systematic review," Jan. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.bbe.2024.01.002.
- [10] Z. Zhang, K. Yan, X. Zhang, X. Rong, D. Feng, and S. Yang, "Automated highway pavement crack recognition under complex environment," *Heliyon*, vol. 10, no. 4, Feb. 2024, doi: 10.1016/j.heliyon.2024.e26142.
- [11] Z. Dong, F. Yang, J. Du, K. Wang, L. Lv, and W. Long, "Identification of varieties in *Camellia oleifera* leaf based on deep learning technology," *Ind Crops Prod*, vol. 216, Sep. 2024, doi: 10.1016/j.indcrop.2024.118635.
- [12] N. Yalçın and M. Alisawi, "Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre-processing techniques," *Heliyon*, vol. 10, no. 20, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38913.
- [13] K. Raja, G. Gupta, S. Venkatesh, R. Ramachandra, and C. Busch, "Towards generalized morphing attack detection by learning residuals," *Image Vis Comput*, vol. 126, Oct. 2022, doi: 10.1016/j.imavis.2022.104535.
- [14] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," Jul. 2018, [Online]. Available: <http://arxiv.org/abs/1807.06521>
- [15] T. Nayak *et al.*, "Deep learning based detection of monkeypox virus using skin lesion images," *Med Nov Technol Devices*, vol. 18, Jun. 2023, doi: 10.1016/j.medntd.2023.100243.
- [16] J. Lu, X. Tong, H. Wu, Y. Liu, H. Ouyang, and Q. Zeng, "Image classification and auxiliary diagnosis system for hyperpigmented skin diseases based on deep learning," *Heliyon*, vol. 9, no. 9, Sep. 2023, doi: 10.1016/j.heliyon.2023.e20186.

- [17] Y. Asiri, H. T. Halawani, A. D. Algarni, and A. A. Alanazi, "IoT enabled healthcare environment using intelligent deep learning enabled skin lesion diagnosis model," *Alexandria Engineering Journal*, vol. 78, pp. 35–44, Sep. 2023, doi: 10.1016/j.aej.2023.07.038.
- [18] C. Kavitha, S. Priyanka, M. P. Kumar, and V. Kusuma, "Skin Cancer Detection and Classification using Deep Learning Techniques," *Procedia Comput Sci*, vol. 235, pp. 2793–2802, 2024, doi: 10.1016/j.procs.2024.04.264.
- [19] N. S. Irjanto and N. Surantha, "Home Security System with Face Recognition based on Convolutional Neural Network," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 408–412, 2020, doi: 10.14569/IJACSA.2020.0111152.