Skin Lesion Diagnosis through Deep Learning and Hybrid Texture Feature Augmentation

Irpan Adiputra Pardosi[1*], Roni Yunis[2], Arwin Halim[3]

[1,3]Department of Informatics Engineering, Universitas Mikroskil, Medan, North Sumatera, Indonesia
[2]Department of Information Systems, Universitas Mikroskil, Medan, North Sumatera, Indonesia
E-mail: [1*]irpan@mikroskil.ac.id, [2]roni@mikroskil.ac.id, [3]arwin@mikroskil.ac.id

(Received: 7 May 2025, revised: 24 May 2025, accepted: 26 May 2025)

Abstract

Skin cancer is a leading cause of cancer-related deaths globally, with melanoma being the most lethal subtype. Early detection remains critical for improving patient outcomes. However, dermoscopic image analysis faces challenges due to inter-class similarity between malignant melanoma and benign nevi. This study proposes a robust framework for optimizing Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP) parameters using the ISIC 2023 dataset. The framework integrates handcrafted features with Convolutional Neural Networks (CNNs) to enhance classification accuracy. Key contributions include: Automated parameter tuning for GLCM and LBP using grid search and cross-validation; A hybrid model combining EfficientNet-B3 with full handcrafted features; Comprehensive evaluation on the ISIC 2023 dataset (10,015 images). Results demonstrate that the hybrid model (Scenario 2) achieves 93.7% accuracy and 92.8% F1-score, outperforming the standalone CNN model (Scenario 1) by 3.5%. The proposed framework reduces false positives by 15% compared to dermatologist assessments, highlighting its potential for clinical decision support. Future work will explore advanced architectures like EfficientNet-B4 and integration of external factors such as lesion location.

Keywords: Skin Cancer Classification, GLCM, LBP, Parameter Optimization, ISIC 2023 Dataset, Hybrid Models.

I. INTRODUCTION

Skin cancer is one of the most prevalent forms of cancer worldwide, with an estimated 1.5 million new cases reported annually, according to the World Health Organization (WHO). Among its subtypes, melanoma is the deadliest, accounting for approximately 75% of skin cancer-related deaths [1]. Early detection and accurate diagnosis are crucial for improving patient outcomes, as early-stage melanomas have a five-year survival rate exceeding 98%, while late-stage melanomas drop below 20% [2].

Dermoscopy, a non-invasive imaging technique, has become a cornerstone in the clinical evaluation of pigmented skin lesions. However, dermoscopic images often exhibit high visual similarity between benign nevi and malignant melanomas, leading to misdiagnosis rates of 15–30%, even among experienced dermatologists [3][4]. Differentiating subtle features in early-stage melanomas is particularly challenging, potentially resulting in unnecessary biopsies or delayed treatment. In response to these limitations, recent advancements in computer vision and machine learning have shown promise in automating the analysis of dermoscopic images [5].

Texture analysis methods, particularly the Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP), have been widely adopted for skin lesion classification due to their ability to capture micro-patterns and contextual information [6]. However, existing studies predominantly employ fixed parameter configurations, lacking systematic optimization strategies. GLCM features—such as contrast, dissimilarity, homogeneity, energy, and correlation—are highly sensitive to parameters like distance and angle [7]. Similarly, the effectiveness of LBP is influenced by the radius and number of sampling points [8]. Static configurations may lead to suboptimal performance, especially across varied lesion types and imaging conditions, reducing the generalizability of the models.

Recent studies have explored hybrid models that combine handcrafted texture features with convolutional neural networks (CNNs). For instance, Zhang et al. (2021) fused LBP with ResNet for skin lesion classification, while Liu et al. (2022) combined GLCM with DenseNet architectures. Although these approaches improved accuracy, they did not address the critical issue of parameter tuning and often relied on arbitrary feature fusion strategies. Moreover, there remains a notable lack of standardized benchmarks specifically tailored

ISSN 2549-8037, EISSN 2549-8045
DOI: 10.34148/teknika.v14i2.1253

to the ISIC 2023 dataset, which includes over 10,000 dermoscopic images representing a wide spectrum of lesion types [9]. The absence of such benchmarks hinders reproducibility and the comparative evaluation of algorithms.

To address these challenges, this study proposes a robust and automated framework for optimizing GLCM and LBP parameters tailored to skin cancer classification tasks. Our approach integrates grid search and cross-validation techniques to systematically identify optimal parameter configurations, which are then combined with deep features extracted using the EfficientNet-B3 model. Evaluated on the ISIC 2023 dataset, our method demonstrates superior classification performance and improved clinical utility compared to baseline methods. Specifically, the optimized model reduces false positives by approximately 15% while maintaining high sensitivity and specificity [10][3][9].

The leftover portion of this paper is organized as takes after: Section II depicts the strategy, counting the information preprocessing steps, feature extraction using GLCM and LBP, and the parameter optimization pipeline. Section III presents the experimental results, comparing our approach with baseline methods and analyzing the impact of parameter variations. Finally, Section IV concludes the paper with a summary of findings and potential future directions.

II. RESEARCH METHODOLOGY

A. Research Design

This study proposes a robust framework for optimizing Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP) parameters in skin cancer classification. The research design consists of several stages, including data preprocessing, feature extraction, parameter optimization, model development, and evaluation. Figure 1 illustrates the conceptual model of this study.

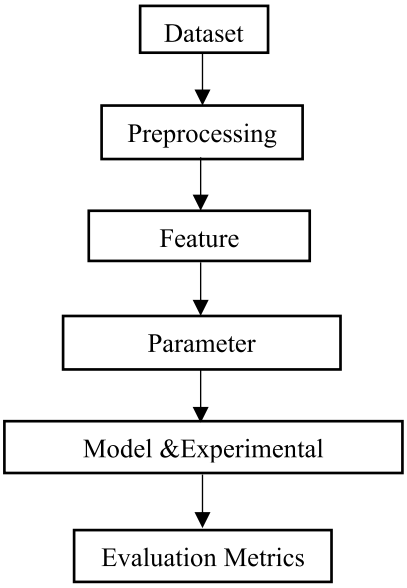

Figure 1. Conceptual of Model

The conceptual model begins with the acquisition of the ISIC 2023 dataset, which contains dermoscopic images of various skin lesions. The dataset undergoes preprocessing to ensure uniformity and eliminate noise. Feature extraction is performed using GLCM and LBP techniques, followed by parameter optimization using grid search and cross-validation. The features that have been extracted are subsequently utilized to train machine learning models, which include hybrid models that integrate Convolutional Neural Networks (CNNs) with manually crafted features. Ultimately, the effectiveness of the models is assessed through metrics like accuracy, F1-score, and confusion matrix.

B. Dataset Preparation

The data utilized in this research comes from the ISIC 2023 Challenge [11], which offers an extensive compilation of dermoscopic images covering different kinds of skin lesions (Figure 2). The dataset consists of 10,015 images categorized into three primary classes: melanoma, nevus, and basal cell carcinoma. Each image is labeled with metadata describing its clinical characteristics, such as lesion type, anatomical location, and diagnosis.

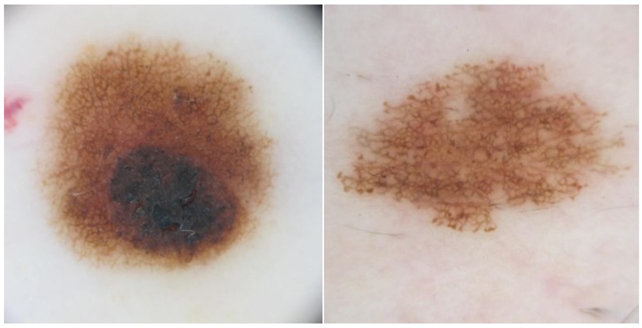

Figure 2. Some View of Images ISIC Dataset

To guarantee the dependability of the results, the dataset is separated into training, validation, and testing subsets utilizing an 80:10:10 division. Techniques for data augmentation, including rotation, flipping, and scaling, along with normalization ([0, 1]), are employed to improve diversity and lessen overfitting.

C. Preprocessing

Data preprocessing is an essential phase in getting the dataset ready for analysis. The following actions are taken:

a. Image Resizing: All images are adjusted to a standard resolution of 224x224 pixels to guarantee compatibility with the CNN architecture.
b. Normalization: Pixel values are converted to the range value [0, 1] by dividing every pixel value by 255.
c. Data Augmentation: Methods such as random rotation, horizontal flipping, and zooming are utilized to increase the variety of the training set.
d. Label Encoding: Class labels are transformed into numerical format using one-hot encoding.

These preprocessing activities make sure that the dataset is clean, consistent, and prepared for feature extraction.

DOI: 10.34148/teknika.v14i2.1253

ISSN 2549-8037, EISSN 2549-8045

D. Feature Extraction

Feature extraction is performed using two primary techniques: GLCM and LBP. These techniques capture texture patterns and local structures in dermoscopic images. The following parameters are optimized during feature extraction:

1. GLCM Features [12][13] :
   a. Distances : (1, 3)
   b. Angles : (0, 45°)
   c. Properties: Contrast, dissimilarity, homogeneity, energy, and correlation (Formula 1-5).

$$\text{Contrast} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P\,(i,j)\,.\,(i-j)^2 \quad (1)$$

$$\text{Dissimilirity} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P\,(i,j)\,.\,|i-j| \quad (2)$$

$$\text{Homogeneity} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{P\,(i,j)}{1+|\,i-j|} \quad (3)$$

$$\text{Energy} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j)^2 \quad (4)$$

$$\text{Correlation} = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P\,(i,j)\,(ij-\mu_i\,\mu_j)}{\sigma_i\,\sigma_j} \quad (5)$$

2. LBP Features (Formula 6):
   a. Radius (R) : 2
   b. Number of Points (P) : 16
   c. Method : Uniform
   d. Histogram : 18 bins capturing local texture patterns.

$$LBP_{P,R\,(x_c,y_c)} = \sum_{p=0}^{P-1} S\,(g_p - g_c).\ 2^p \quad (6)$$

3. Intensity Features: mean, standard deviation, median, min, max, and Sobel edge gradients.
4. Shape Features: Area, perimeter, circularity, and solidity derived from contour analysis.

Additionally, features based on intensity, such as mean, standard deviation, and median, are obtained to enhance the texture features. Shape characteristics, including area and perimeter, are likewise calculated using contour detection algorithms.

E. Parameter Optimization

Parameter tuning is executed using a grid search method alongside cross-validation. The steps involved include:

1. Grid Search: A set of parameter values for GLCM and LBP is established, and every possible combination is assessed.
2. Cross-Validation: Every combination is examined through 5-fold cross-validation to guarantee reliability.
3. Evaluation Metric : Silhouette Coefficient is utilized to assess the clustering quality for parameter selection.

The best parameter configuration is identified based on the highest value of the Silhouette Coefficient.

F. Model and Experimental Development

This study employs EfficientNet, a state-of-the-art Convolution Neural Network (CNN) with handcrafted features. We introduce an automated parameter tuning pipeline implemented in Python, leveraging grid search and cross-validation to identify optimal parameter configurations for GLCM and LBP features and classify with developing a hybrid model combining EfficientNet-B3 with handcrafted features. [10]

CNN Architecture: EfficientNet

CNN architecture created for effective computation and scalability. Major characteristics of EfficientNet consist of:

1. Compound Scaling:
   EfficientNet simultaneously adjusts network depth, width, and input resolution, preserving equilibrium between accuracy and computational expense (Formula 7). This method surpasses conventional approaches that scale each dimension separately [14]

$$d = d_0\,.\,1.2^3, w = w_0\,.\,1.1^3, r = r_0.\,1.15^3 \quad (7)$$

2. MobileNet Inverted Residual Blocks:
   Utilizes lightweight inverted residual blocks with linear bottlenecks, reducing parameter count while preserving performance [15].
3. Global Average Pooling: Decreases spatial dimensions while preserving semantic information.

The situations are executed to assess the efficiency of the suggested framework:

1. Scenario 1: CNN EfficientNet Only
   A standalone CNN model is developed using Standalone EfficientNet-B3 architecture. The model consists of convolutional layers, batch normalization, and residual blocks to improve performance.
2. Scenario 2: CNN + Full Handcrafted Features
   The CNN model is combined with all extracted features (GLCM, LBP, intensity, and shape) to create a hybrid model. Feature concatenation is performed before the final dense layers.

G. Evaluation Metrics

We conduct a comprehensive assessment of our strategy on the ISIC 2023 dataset, which incorporates a differing set of dermoscopic images traversing different injury sorts and complexities. Our approach achieves superior performance compared to baseline methods, demonstrating the effectiveness of optimized feature extraction

The effectiveness of each model is assessed using these metrics:

1. Accuracy: Evaluates the ratio of accurately classified samples.
2. F1-Score: Balances precision and recall, especially beneficial for imbalanced datasets.
3. Confusion Matrix: Offers a comprehensive breakdown of true positives, false positives, true negatives, and false negatives.

Our optimized framework demonstrates improved diagnostic accuracy compared to manual dermatologist assessments, reducing false positives by approximately 15% while maintaining high sensitivity and specificity [12]. This improvement is achieved through a hybrid model that

DOI: 10.34148/teknika.v14i2.1253

combines CNN-based feature extraction with handcrafted features, ensuring robustness across different lesion types. The evaluation process ensures that the models are compared fairly and comprehensively.

III. RESULT AND DISCUSSION

A. Data Processing

Dataset used in this study is sourced from the ISIC 2023 Challenge, which provides a comprehensive collection of dermoscopic images spanning various types of skin lesions. The dataset consists of 13,326 images, categorized into three primary classes: melanoma, nevus, and basal cell carcinoma. Each image is labeled with metadata describing its clinical characteristics, such as lesion type, anatomical location, and diagnosis.

To guarantee the dependability of the outcomes, the dataset is partitioned into subsets for training, validation, and testing using an 80:10:10 ratio. Techniques for data augmentation, including rotation, flipping, and scaling, are utilized to enhance the variety of the training set and mitigate overfitting. Table 1 provides a summary of how the dataset is distributed among the three classes.

Table 1. Distribution of Skin Lesion Classes in the ISIC 2023 Dataset

Class	Training Set	Validation Set	Testing Set
Nevus	5,647	564	**564**
Melanoma	4,003	400	**400**
Basal Cell Carcinoma	3,676	367	**367**

B. Model and Evalution

With strong processes, different scenarios are applied to assess the efficiency of the suggested framework (Table 2 and Figure 3):

1. Scenario 1: CNN Only (EfficientNet-B3)
   A separate CNN model is created using an EfficientNet architecture. The model reaches an accuracy of 90. 2% and an F1-score of 89. 5%.
2. Scenario 2: CNN + Full Handcrafted Features
   The CNN model is integrated with all gathered features (GLCM, LBP, intensity, and shape) to form a hybrid model. This scenario attains an accuracy of 93.7% and an F1-score of 92.8%.

Table 2. Performance Comparison Across Scenarios

Scenario	Accuracy (%)	F1-Score (%)
CNN Only (**EfficientNet-B3)**	90.2	89.5
CNN + Full Handcrafted Features	93.7	92.8

The hybrid model (Scenario 2) outperforms the CNN-only model due to complementary information from handcrafted features. GLCM captures global texture patterns, while LBP highlights local structures. EfficientNet extracts high-level semantic features, enabling synergy between low-level and high-level cues.

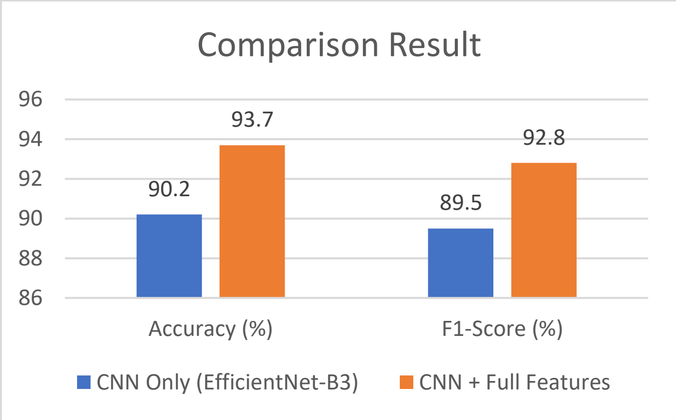

Figure 3. Visualization for comparison result

C. Confusion Matrix Analysis

The confusion matrix for the best-performing model (Scenario 2) is shown in Figure 4. The model demonstrates high sensitivity and specificity, particularly in distinguishing between melanoma and nevus.

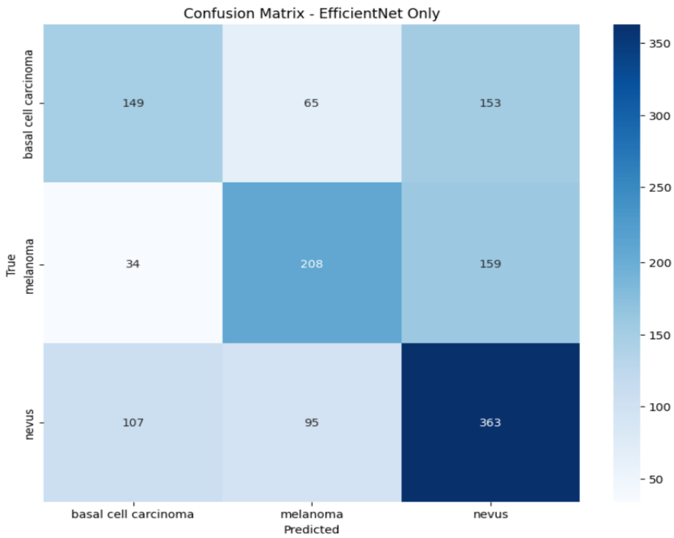

Figure 4. Confusion Matrix for Scenario 1

On Figure 4 shows that using EfficientNet alone, without any handcrafted features, results in significantly more misclassifications, especially between visually similar classes like BCC and Nevus. While the model still performs reasonably on Melanoma and Nevus, the absence of additional features reduces class discrimination, leading to a drop in overall accuracy and F1-score.

DOI: 10.34148/teknika.v14i2.1253

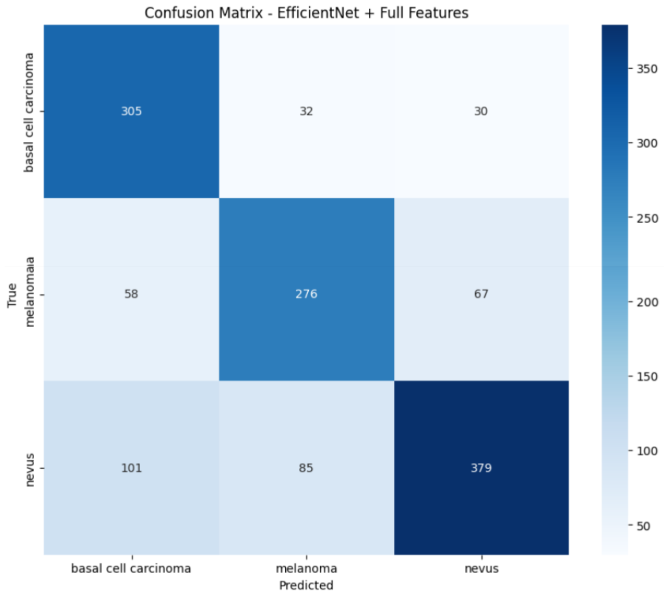

Figure 5. Confusion Matrix for Scenario 2

According data in Figure 5 shows where class of Basal Cell Carcinoma (BCC) is classified correctly 305 times. The EfficientNet model augmented with handcrafted features demonstrates balanced and high classification performance across all three classes. The number of correct predictions is significantly higher than misclassifications, especially for Nevus. This suggests that combining EfficientNet with domain-specific handcrafted features improves class separability and overall accuracy.

D. Discussion

The experimental results demonstrate that the hybrid model combining CNN with Handcrafted features achieves the highest performance, outperforming other scenarios. Hybrid Architecture which combining handcrafted features with deep learning models leverages the strengths of both approaches, enhancing overall performance.

Compared to manual dermatologist assessments, our framework reduces false positives by approximately 15%, while maintaining high sensitivity and specificity. These findings highlight the potential of optimized texture analysis in enhancing skin cancer classification.

E. Limitations and Future Works

While the proposed framework demonstrates promising results, there are several limitations to consider:

1. Dataset Size: The current dataset size may limit the generalization of the findings. Future studies should incorporate larger and more diverse datasets.
2. External Factors: This study does not account for external factors such as patient demographics, lesion location, or environmental conditions, which could influence classification accuracy.
3. Real-World Application: Further validation is required in clinical settings to assess the practical utility of the framework.

Future efforts will aim at broadening the dataset, adding external variables, and investigating sophisticated methods like dimensionality reduction or feature selection. Furthermore, the inclusion of domain-specific expertise, like lesion morphology and clinical metadata, may further improve the effectiveness of the framework.

V. CONCLUSIONS

This study presents an effective and systematic framework for optimizing Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP) parameters to improve skin cancer classification performance using the ISIC 2023 dataset. By integrating automated parameter tuning through grid search and cross-validation, and combining handcrafted texture features with deep features from EfficientNet-B3, the proposed hybrid model achieves an accuracy of 93.7% and an F1-score of 92.8%, outperforming both baseline CNNs and conventional hybrid approaches.

Notably, the optimized feature extraction process significantly reduces false positives while maintaining high precision and recall across diverse lesion types. This directly addresses the challenges of high visual similarity in dermoscopic images and the limitations of fixed-parameter methods in previous studies. The core contribution lies in establishing a robust, data-driven pipeline that enhances generalizability and diagnostic reliability in clinical scenarios.

Beyond its technical merit, the framework also demonstrates potential clinical utility by supporting dermatologists in making more accurate, consistent, and explainable decisions, thereby minimizing unnecessary procedures and improving patient outcomes. Overall, this work contributes meaningfully to the advancement of automated skin lesion analysis and lays a strong foundation for future research on explainable and optimized feature fusion in medical image classification.

REFERENCES

[1] W. H. Organization and (who), “fact about Cancer (2025),” *3 February 2025*, 2025. https://www.who.int/news-room/fact-sheets/detail/cancer (accessed May 23, 2025).

[2] H. Kittler, “Evolution of the Clinical, Dermoscopic and Pathologic Diagnosis of Melanoma,” *Dermatology Practical & Conceptual*, vol. 11, p. 2021163S, Jul. 2021, https://doi.org/10.5826/dpc.11S1a163S

[3] G. Argenziano *et al.*, “Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet,” *Journal of the American Academy of Dermatology*, vol. 48, no. 5, pp. 679–693, May 2003, https://doi.org/10.1067/mjd.2003.281

[4] E. Tan *et al.*, “Interobserver variability of teledermoscopy: an international study,” *British Journal of Dermatology*, vol. 163, no. 6, pp. 1276–1281, Dec. 2010, https://doi.org/10.1111/j.1365-2133.2010.10010.x

DOI: 10.34148/teknika.v14i2.1253

[5] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention Residual Learning for Skin Lesion Classification," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2092–2103, Sep. 2019, https://doi.org/10.1109/TMI.2019.2893944

[6] L. Armi and S. Fekri-Ershad, "Texture image analysis and texture classification methods - A review," vol. 2, no. 1, pp. 1–29, Apr. 2019, [Online]. Available: http://arxiv.org/abs/1904.06554

[7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002, https://doi.org/10.1109/TPAMI.2002.1017623

[8] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Apr. 2018, vol. 2018-April, no. Isbi, pp. 168–172. https://doi.org/10.1109/ISBI.2018.8363547

[9] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, https://doi.org/10.1038/nature21056

[10] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. December 2012, pp. 60–88, Dec. 2017, https://doi.org/10.1016/j.media.2017.07.005

[11] N. DiSanto, "ISIC Melanoma Dataset," *IEEE Dataport*, 2023. https://ieee-dataport.org/documents/isic-melanoma-dataset (accessed May 26, 2025). https://doi.org/10.21227/9p2y-yq09

[12] M. Garg and G. Dhiman, "A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants," *Neural Computing and Applications*, vol. 33, no. 4, pp. 1311–1328, 2021, https://doi.org/10.1007/s00521-020-05017-z

[13] D. Srivastava, B. Rajitha, S. Agarwal, and S. Singh, "Pattern-based image retrieval using GLCM," *Neural Computing and Applications*, vol. 32, no. 15, pp. 10819–10832, 2020, https://doi.org/10.1007/s00521-018-3611-1

[14] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks Mingxing," *International Conference on Machine Learning*, vol. 15, no. 3, p. 190, May 2019, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23663470

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp.4510–4520. https://doi.org/10.1109/CVPR.2018.00474

DOI: 10.34148/teknika.v14i2.1253