Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2026, 2026:79

https://doi.org/10.28919/cmbn/9990

ISSN: 2052-2541

COST-SENSITIVE DENSENET-121 WITH PROGRESSIVE CURRICULUM AUGMENTATION AND DYNAMIC THRESHOLD OPTIMIZATION FOR IMBALANCED SKIN CANCER CLASSIFICATION

PUSPITA KARTIKASARI[1,*], RUKUN SANTOSO[1], NOVRI SUHERMI[2], RIZWAN ARISANDI[3], PARLINDUNGAN SITUMORANG[1]

[1]Department of Statistics, Universitas Diponegoro, Semarang 50275, Indonesia

[2]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

[3]Department of Computer Science, Universitas Bina Nusantara, Jakarta 11480, Indonesia

Abstract: Skin cancer is one of the leading causes of cancer-related mortality worldwide, where early melanoma detection substantially improves patient survival. However, automated dermoscopic image classification remains challenging because malignant lesions are severely underrepresented, causing deep learning models to become biased toward benign cases. This study proposes a clinically oriented deep learning framework for imbalanced skin cancer classification using DenseNet-121 on a rigorously curated subset of the ISIC 2024 dataset. Prior to model development, patient-level integrity validation and duplicate inspection were performed, followed by patient-level data partitioning to eliminate data leakage. A total of 12,000 dermoscopic images were curated and divided into 8,296 training, 1,904 validation, and 1,800 testing images. The proposed framework integrates curriculum-based progressive augmentation, cost-sensitive optimization using weighted binary cross-entropy, MixUp and CutMix regularization, and dynamic

*Corresponding author

E-mail address: puspitakartikasari@live.undip.ac.id

Received May 17, 2026

threshold optimization to prioritize malignant lesion detection. Experimental results demonstrate that the proposed model achieved a test area under the receiver operating characteristic curve (AUC) of 0.8959 and a malignant recall (sensitivity) of 92.68%, correctly identifying 38 of 41 malignant lesions while missing only three cases. Although the clinically optimized decision threshold reduced precision because of increased false positives, it substantially decreased false negatives, making the framework well suited for preliminary melanoma screening. These findings demonstrate that combining rigorous patient-level curation, progressive augmentation, cost-sensitive learning, and clinically driven threshold optimization provides an effective strategy for improving malignant lesion detection in highly imbalanced dermoscopic datasets.

Keywords: Skin lesion classification; melanoma detection; deep learning; DenseNet-121; cost-sensitive learning; curriculum augmentation; class imbalance; dynamic thresholding.

2020 AMS Subject Classification: 68T07.

1. INTRODUCTION

Skin cancer remains one of the most predominant and aggressive malignancies worldwide. It generally originates from the abnormal proliferation of melanocytes or keratinocytes. While genetic predisposition plays a role, prolonged ultraviolet (UV) radiation exposure from sunlight and tanning beds acts as the primary catalyst [1, 2]. In clinical settings, practitioners divide these tumors into two main categories: localized benign lesions and highly invasive malignant forms like melanoma [3]. The global health burden is staggering. Recent GLOBOCAN 2022 statistics show that melanoma alone drove roughly 331,722 new diagnoses and 58,667 deaths, with non-melanoma cases exceeding 1.2 million [3]. Melanoma is particularly lethal once it metastasizes. Yet, there is a clear clinical consensus: catching the disease in its initial stages pushes the survival rate to 97% or even 99% [1, 4]. Timely and accurate diagnosis is absolutely critical.

Today, visual investigation by a dermatologist acts as the frontline defense against skin cancer. Doctors rely heavily on dermoscopy, followed by histological biopsies for confirmation. These manual methods represent the current gold standard. They are, however, incredibly labor-intensive and completely bound to the subjective experience of the examining clinician [5]. Dermoscopy

certainly improves visualization. Still, distinguishing an early-stage melanoma from a harmless benign mole is notoriously difficult [6]. Lesions often share subtle visual traits. Compounding this issue are physical image artifacts such as dense hair occlusions, air bubbles, or uneven illumination which obscure clear analysis. Because of these hurdles, real-world diagnostic accuracy using standard dermoscopy rarely breaks the 80% mark [4, 6]. With skin cancer cases surging globally and a stark shortage of specialized dermatologists, the medical field urgently needs objective, scalable diagnostic tools.

To fill this gap, researchers have turned to computer-aided diagnosis (CAD) systems. Deep learning (DL) models Convolutional Neural Networks (CNNs) in particular have become the dominant approach for medical image analysis. Unlike older classical computer vision models, CNNs do not require tedious manual feature engineering [5, 7]. They learn and extract composite, hierarchical features straight from raw pixel data. Over the past few years, these DL architectures have proven highly effective across a range of diseases. In dermatology, automated models offer a way to rapidly stratify patient risk. They provide a reliable second opinion that can streamline clinical workflows and reduce the heavy reliance on manual screening.

Building a robust DL classifier for skin cancer, however, is not a straightforward task. Clinical datasets present severe challenges. The most prominent issue is extreme class imbalance. In real-world screening scenarios, benign lesions vastly outnumber malignant ones. When trained on heavily skewed data, standard classifiers inevitably develop a strong bias toward the majority class. They memorize the benign features, overfit, and fail to recognize the rare but deadly minority class [2, 4]. Researchers usually apply data augmentation to fight this bias. By applying geometric transformations or color modifications, they artificially expand the minority class distribution so the model can learn to generalize rather than just memorize [4].

This manuscript introduces an advanced deep learning framework explicitly designed to tackle binary skin cancer classification on the highly imbalanced ISIC 2024 dataset. We address the data imbalance through a combination of complementary strategies. A curriculum-based progressive augmentation schedule is employed to gradually expose the model to increasingly challenging

variations as training matures, allowing the network to first learn basic lesion morphology before confronting more severe visual artifacts. This is coupled with batch-level regularisation through MixUp and CutMix to further improve generalisation. To extract intricate multi-level features from these challenging dermoscopic images, the model relies on the DenseNet-121 architecture, with network parameters optimised using the Adam algorithm. Most importantly, this study prioritises patient safety over raw accuracy. We introduce a dynamic thresholding mechanism explicitly tuned to maximise diagnostic sensitivity (recall) for the malignant class. Through this combination of careful data handling and deep learning, the proposed model functions as a highly sensitive preliminary triage tool, actively suppressing false negatives to support early and safe melanoma detection in clinical practice.

2. Data and Methods

2.1 Dataset Description, Integrity Validation, and Patient-Level Curation

Automated classification of skin lesions has evolved rapidly with the widespread adoption of deep convolutional neural networks (CNNs). Early and contemporary studies consistently report that CNN-based approaches outperform classical hand-crafted feature pipelines because CNNs learn hierarchical representations directly from raw pixels [4, 6]. Nevertheless, the reliability of CNN-based medical image classification strongly depends on dataset quality, integrity, and proper curation procedures. Analyses of ISIC releases have identified duplicated images, downsampled copies, and inconsistent patient-level splits that can lead to data leakage and overly optimistic performance estimates if not carefully addressed [12]. Consequently, patient-level deduplication and rigorous train-validation-test partitioning are increasingly regarded as mandatory for credible evaluation [12, 19].

This study utilizes the ISIC 2024 dataset, a leading public repository for dermatoscopic image analysis containing over 400,000 high-quality images accompanied by gold-standard metadata, including binary labels (benign/malignant) and unique patient identifiers [12, 16]. The dataset is publicly available for download at https://challenge.isic-archive.com/data/#2024. Raw medical

image datasets frequently suffer from duplicate images, severe class imbalance, and leakage risks that may bias experimental findings [12, 17]. Cassidy et al. emphasized that failure to remove duplicates and patient overlap can substantially inflate reported model performance [12].

To ensure dataset integrity, a comprehensive metadata validation procedure was conducted prior to model development. The dataset contained 401,062 total records and included all required metadata fields (target, isic_id, and patient_id). No missing values were identified in these key attributes, and no duplicated image identifiers were detected based on the unique isic_id field. However, repeated patient identifiers were frequently observed, indicating that multiple lesion images originated from the same patient.

Following the recommendation that unique lesion diversity contributes more effectively to model generalization than simply increasing image quantity [12], a curated heterogeneous subset of approximately 12,000 images was constructed. The sampling strategy prioritized all malignant lesions by selecting one image per patient to maximize inter-patient lesion variability. Benign images were subsequently added by prioritizing unique patients first and then incorporating additional images only when necessary to achieve the target dataset size. This strategy was designed to mitigate the severe imbalance commonly found in dermoscopic datasets, where benign lesions overwhelmingly dominate malignant cases [12, 18]. The final curated dataset therefore consists of 12,000 images, including 259 malignant samples from unique patients and 11,741 benign samples, corresponding to 2.16% and 97.84% of the dataset respectively.

To prevent data leakage, the curated dataset was partitioned at the patient level using GroupShuffleSplit based on patient_id. First, the data were split into a training set containing 70% of the patients and a temporary set containing the remaining 30%. Due to the constraint that all images of a patient must stay together, the exact sizes were 8,296 training images (69.1%) and 3,704 images in the temporary set (30.9%). The temporary set was subsequently divided into validation and testing sets, yielding 1,904 validation images (15.9%) and 1,800 testing images (15.0%). Consequently, all images from a single patient remain exclusively within one subset. This patient-centric partitioning ensures unbiased evaluation and improves the reproducibility and

clinical credibility of the proposed framework [12, 19].

2.2 Preprocessing and Curriculum-Based Augmentation

Data preprocessing and augmentation play critical roles in medical image classification because dermoscopic datasets often contain limited samples, imaging artifacts, illumination variability, and class imbalance [20, 21]. The proposed preprocessing pipeline combines standard enhancement techniques with progressive curriculum-based augmentation designed specifically for dermoscopic analysis.

2.2.1 Image Preprocessing

All images were resized to 224 × 224 pixels, which is a standard input size for CNN-based image classification [21]. To reduce illumination inconsistency across imaging devices, the Shades of Gray color constancy algorithm was applied to normalize image color distributions [21, 22]. Subsequently, Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to improve local contrast while minimizing excessive noise amplification [20, 23, 24].

Hair occlusion is a common artifact in dermoscopic imaging and may obscure diagnostically important lesion regions [12, 17]. To improve robustness against this issue, a custom synthetic hair augmentation module was implemented by randomly overlaying hair-like curves onto training images. This approach encourages the model to learn invariant lesion representations despite partial visual obstruction [12, 17].

2.2.2 Curriculum-Based Progressive Augmentation

Instead of applying all augmentations simultaneously throughout training, this study adopts a curriculum-based augmentation schedule inspired by progressive learning strategies [20]. The same augmentation pipeline is applied to all classes, while the augmentation complexity gradually increases across training epochs using the Albumentations framework [25].

During the first 20% of training epochs, only mild augmentations were activated, including horizontal and vertical flips with probability 0.5 and color jitter with brightness and contrast variation of ±0.1 applied at probability 0.3. These transformations encourage basic spatial invariance while preserving important low-level lesion morphology. After 20% of training, two

domain-specific transformations were introduced:

1. Dermoscopy style simulation, which randomly shifts image color temperature toward warmer or cooler illumination conditions to mimic variations between dermoscopic devices [25].
2. FFT amplitude jitter, which perturbs frequency-spectrum amplitudes while preserving phase information, thereby increasing texture variability without destroying lesion structure.

Both transformations were applied independently with probability 0.3. After 50% of training epochs, elastic deformation was additionally activated using $\alpha = 20$ and $\sigma = 50$ with probability 0.3. Delaying this stronger geometric transformation allows the network to first stabilize its understanding of lesion morphology before learning more challenging deformations. Throughout training, image resolution was progressively scaled from 179 pixels at the earliest epochs to 269 pixels at later epochs using bicubic interpolation. Black symmetric padding was then added to produce a final 224 × 224 image. Finally, all images were normalized using ImageNet channel statistics with $\mu = [0.485, 0.456, 0.406]$ and $\sigma = [0.229, 0.224, 0.225]$.

The complete curriculum augmentation schedule is summarized in Table 1.

Table 1. Curriculum augmentation schedule.

Training Progress	Image Resolution	Active transforms	Probability
0–20%	179 – 192 px	HorizontalFlip, VerticalFlip ColorJitter (brightness=0.1, contrast=0.1)	0.5 (each)
20–50%	192 – 202 px	+ Dermoscopy-style simulation + FFT Amplitude Jitter (limit=0.15)	0.3
50-100%	202 – 224 px	+ ElasticTransform (α=20, σ=50)	0.3

2.3 DenseNet-121 Architecture

CNN architecture selection significantly affects medical image classification performance [7, 19]. Among modern CNN architectures, DenseNet has demonstrated strong effectiveness because of its dense connectivity mechanism, where each layer receives feature maps from all preceding layers and forwards its outputs to all subsequent layers [27]. Dense connectivity improves information flow, promotes feature reuse, and reduces vanishing gradients while maintaining relatively low parameter counts [27–29]. These properties are particularly beneficial for medical imaging tasks involving limited datasets and subtle inter-class visual differences [21, 28].

Skin lesion images exhibit high intra-class variability and low inter-class separability [12, 29]. DenseNet facilitates the simultaneous extraction of low-level texture information and higher-level semantic lesion representations through multi-level feature reuse [27, 29]. Previous studies successfully applied DenseNet variants for melanoma detection and reported strong performance in dermoscopic classification tasks [7, 30]. This study employs DenseNet-121 implemented through the timm library [19]. The network consists of four dense blocks with growth rate k=32separated by transition layers using convolution and pooling operations. Several modifications were introduced for binary lesion classification:

- Initialization with imagenet pretrained weights,
- Replacement of the original 1000-class output head with a single-neuron sigmoid classifier,
- Compatibility with 224 × 224 rgb input images.

The major architecture specifications are summarized in Table 2.

Table 2. Key specifications of the modified DenseNet-121 architecture.

Component	Description
Base Architecture	DenseNet-121 (timm implementation)
Pre-training	ImageNet
Input Size	224 × 224 × 3
Dense Blocks	4 blocks, growth rate k=32
Transition Layers	Convolution + pooling, compression factor 0.5
Global Pooling	Average pooling
Output Layer	Fully connected (1 unit) + Sigmoid
Total Trainable Parameters	Approximately 7.0 million

2.4 Cost-Sensitive Learning and Optimization Strategy

Despite dataset curation and augmentation, class imbalance remained substantial, with benign lesions still outnumbering malignant lesions by approximately 45.3:1 [12, 18]. To address this imbalance during optimization, cost-sensitive learning was incorporated through weighted binary cross-entropy loss using BCEWithLogitsLoss in PyTorch.

Instead of applying the full inverse class ratio directly, the positive class weight was empirically capped at 15.0 to prevent excessive gradient updates and over-penalization of benign samples, which could otherwise lead to unstable convergence and an inflated false-positive rate [18, 21].

This calibrated weighting strategy still amplifies the malignant class contribution substantially compared to an unweighted baseline, effectively encouraging the model to prioritize clinically important sensitivity (recall) during optimization.

To improve generalization and reduce overfitting, strong batch-level regularization techniques were incorporated through MixUp and CutMix augmentation [31, 32]. MixUp with α=0.4was applied to 25% of training batches, while CutMix with α=1.0was applied to another 25% of batches. The remaining 50% of batches were trained without batch-level mixing augmentation. Training was conducted for a maximum of 10 epochs using a batch size of 32, with validation AUC serving as the early stopping criterion. The complete hyperparameter configuration is summarized in Table 3.

Table 3. Training hyperparameters for the DenseNet-121 model

Hyperparameter	Value
Optimizer	Adam
Initial learning rate	1×10^{-4}
Learning rate scheduler	ReduceLROnPlateau (patience = 2, factor = 0.1)
Loss function	BCEWithLogitsLoss
Positive class weight	15.0
MixUp probability (α=0.4)	0.25
CutMix probability (α=1.0)	0.25
Batch size	32
Number of epochs	10 (early stopping on validation AUC)
Threshold selection criterion	$\{\theta \mid recall(\theta) \geq 0.95\}$

2.5 Dynamic Threshold Optimization

The default classification threshold of 0.5 is rarely optimal for medical screening tasks because false negatives for malignant lesions carry substantially higher clinical risk than false positives [13, 18, 29]. Therefore, threshold selection was optimized specifically for high-sensitivity melanoma screening.

The classification threshold was tuned exclusively on the validation set using a recall-constrained optimization strategy that guarantees at least 95% recall for malignant lesions [33]. Formally, the selected threshold satisfies: $\theta = \max\{\theta \mid r(\theta) \geq 0.95\}$, where r(θ) denotes malignant-class recall

at threshold θ.

This criterion selects the highest threshold that still preserves the required sensitivity level, thereby minimizing unnecessary false positives while maintaining clinically acceptable recall [33]. The resulting threshold for the DenseNet-121 model was approximately $\theta \approx 0.1854$. This relatively low threshold intentionally prioritizes sensitivity over precision, making the framework suitable for preliminary melanoma screening and triage applications.

3. Main Results

3.1. Training Performance and Convergence

All experiments were conducted using the curated 12,000-image subset of the ISIC 2024 dataset, partitioned at the patient level into training (70%, 8,296 images), validation (15%, 1,904 images), and testing (15%, 1,800 images) sets as described in Section 2.1. The DenseNet-121 architecture was initialized using ImageNet pretrained weights and optimized using the Adam optimizer with an initial learning rate of 1×10^{-4}. A ReduceLROnPlateau scheduler (patience = 2, factor = 0.1) dynamically adjusted the learning rate during training. The model employed BCEWithLogitsLoss with a positive class weight of 15.0 to mitigate severe class imbalance. Curriculum-based augmentation was applied online during training, and all experiments were executed using PyTorch and Albumentations on an NVIDIA Tesla P100 GPU with a fixed random seed (2024) to ensure reproducibility. An additional integrity verification confirmed that no patient appeared simultaneously across different dataset splits, eliminating potential patient-level data leakage.

Figure 1 illustrates the training and validation loss curves. Both losses decreased consistently throughout training, indicating stable optimization without severe overfitting. The validation AUC achieved its highest value of 0.9175 at epoch 3, after which only minor fluctuations were observed. Consequently, the epoch-3 checkpoint was selected as the final model for subsequent evaluation.

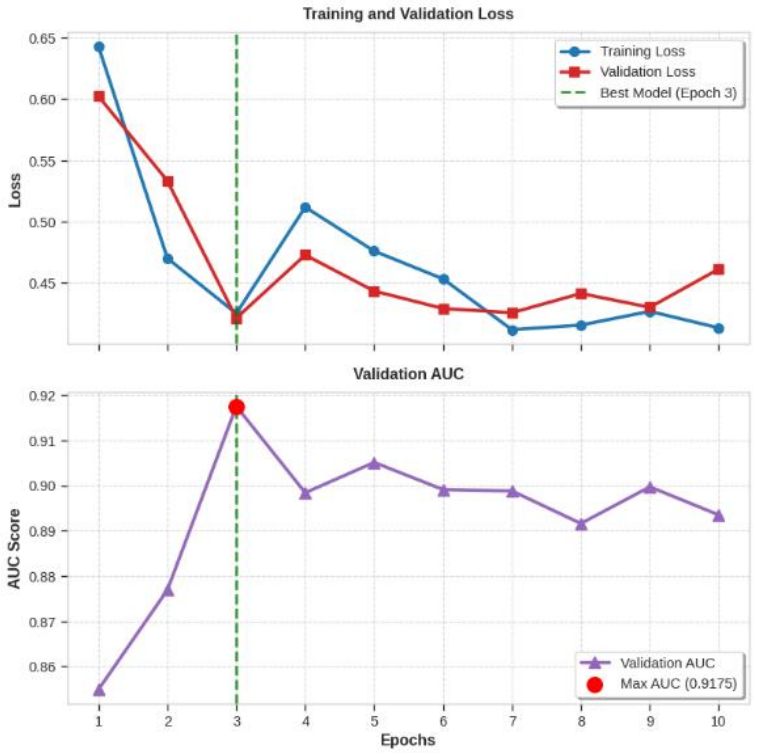

Figure 1. Training and validation loss curves.

3.2 Validation Performance and Threshold Optimization

Because false negatives in melanoma screening are clinically more critical than false positives, model evaluation was not limited to the default classification threshold of 0.5. Instead, threshold optimization was performed using the validation set to maximize malignant lesion recall while maintaining reasonable specificity.

A precision–recall curve was constructed from validation predictions, and the threshold was selected using the following recall-constrained criterion: $\theta = \max\{\theta \mid r(\theta) \geq 0.95\}$, where r(θ)represents malignant-class recall at threshold θ. The resulting clinically oriented threshold was determined to be 0.1854. Table 4 compares the validation performance obtained using the default threshold (0.5) and the optimized threshold.

Table 4. Validation set performance at different decision thresholds.

Threshold	AUC	Recall (Malignant)	Precision (Malignant)	Specificity (Benign)
0.5	0.9175	0.81	0.17	0.91
0.1854	0.9175	0.95	0.06	0.67

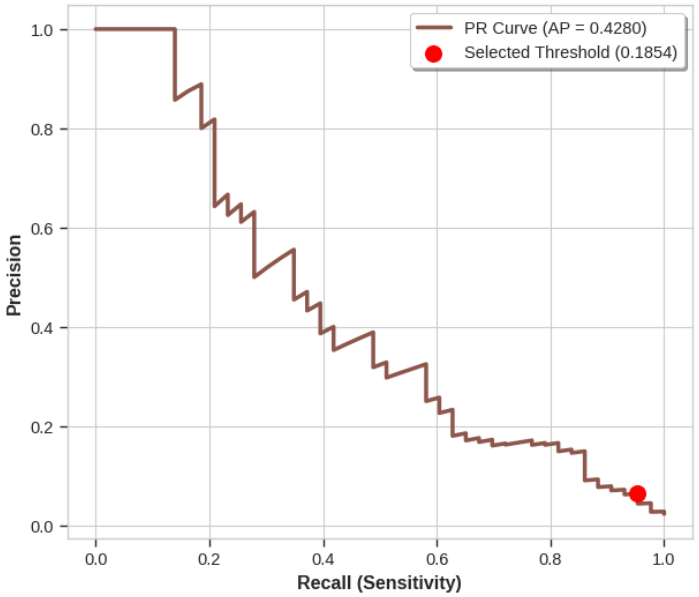

Figure 2. Precision–Recall curve on the validation set.

Using the optimized threshold of 0.1854, the model correctly identified 41 out of 43 malignant lesions on the validation set, corresponding to a recall of 95.35%. Only two malignant lesions were missed. Although the lower threshold substantially reduced precision to 6% and decreased specificity to 67%, this trade-off is clinically acceptable for screening applications where minimizing missed cancers is prioritized over reducing false alarms.

3.3 Test Set Evaluation and Malignant Detection Performance

The final evaluation was conducted on the held-out test set consisting of 1,800 images, including 41 malignant lesions. The optimized threshold of 0.1854 determined from the validation set was applied directly to the test data without further adjustment. Figure 3 presents the resulting confusion matrix, while the detailed evaluation metrics are summarized in Table 5.

Table 5. Test set performance of the DenseNet-121 model using threshold 0.1854.

Metric	Value
AUC	0.8959
Recall (Sensitivity, Malignant)	92.68%
Specificity (Benign)	64.07%
Precision (Malignant)	6.0%
F1 Score (Malignant)	0.11
False Negatives	3 / 41
False Positives	632 / 1759

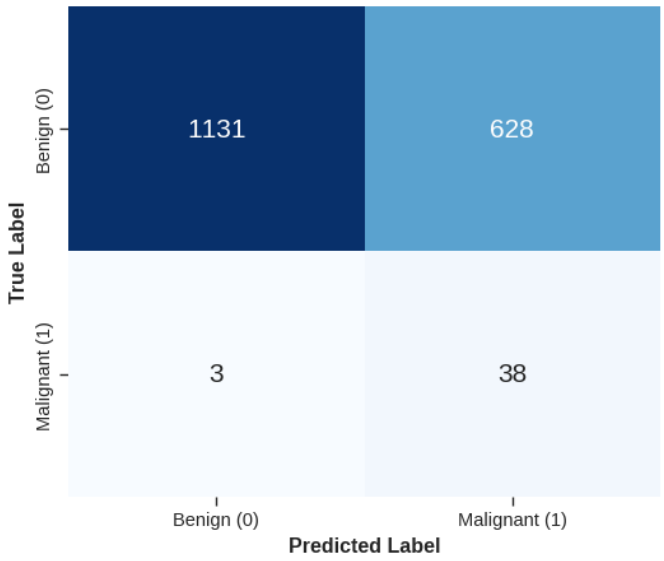

Figure 3. Confusion matrix on the test set.

The proposed framework achieved a test AUC of 0.8959, indicating strong discriminative capability despite the highly imbalanced dataset. Most importantly, the model successfully detected 38 of 41 malignant lesions, corresponding to a malignant recall of 92.68%, while missing only three melanoma cases. The false positive count remained relatively high, with 632 benign lesions incorrectly flagged as malignant. However, this behavior is expected in a clinically oriented triage system where maximizing sensitivity is substantially more important than minimizing false alarms. In practical screening scenarios, false positives can subsequently be filtered through further clinical examination, whereas false negatives may delay diagnosis and treatment.

Compared with the default threshold of 0.5, which achieved only 81% recall on the validation set, the proposed threshold optimization strategy significantly improved malignant detection performance and reduced the number of missed cancer cases.

3.4 Discussion

The experimental results demonstrate that the proposed framework effectively balances strong discriminative performance with clinically prioritized malignant sensitivity. The achieved test AUC of 0.8959 is competitive with recent skin lesion classification studies. For example, Alam et al. [1] reported an AUC of 0.92 and malignant recall of 85% on an ISIC 2020 subset using a conventional CNN-based framework, while Jayaseeli et al. [5] achieved an AUC of 0.91 using a

DenseNet-SE hybrid architecture. Although our framework attains a slightly lower AUC, it achieves substantially higher malignant recall through clinically motivated threshold optimization. Several factors contributed to the effectiveness of the proposed system. First, rigorous patient-level curation prevented data leakage and ensured realistic evaluation. Second, curriculum-based augmentation improved model generalization by progressively introducing increasingly complex transformations during training. Third, cost-sensitive learning explicitly penalized malignant misclassification through weighted optimization. Finally, dynamic threshold optimization transformed probabilistic outputs into clinically meaningful screening decisions by prioritizing sensitivity over raw accuracy.

Despite these promising results, several limitations remain. The experiments were conducted on a curated subset of ISIC 2024 and may not fully represent the diversity encountered in real-world dermatological practice. In addition, evaluation was limited to a single held-out test set, and external validation on independent datasets is still required to confirm generalizability. The intentionally high false positive rate may also reduce clinical efficiency and could potentially be improved through post-processing strategies or ensemble learning techniques.

Future work will investigate lightweight architectures such as MobileNet for deployment in resource-constrained environments, as well as explainability methods including Grad-CAM to provide interpretable visual evidence supporting model predictions. Ensemble-based strategies combining multiple backbones may also help improve specificity while preserving high malignant recall.

This study proposed a clinically oriented deep learning framework for imbalanced skin cancer classification by integrating rigorous patient-level dataset curation, curriculum-based progressive augmentation, cost-sensitive optimization, and dynamic threshold selection within a DenseNet-121 architecture. The proposed framework effectively addressed the severe class imbalance inherent in the ISIC 2024 dataset while preventing patient-level data leakage through strict dataset partitioning.

4. CONCLUSION

Experimental results demonstrated that the model achieved a test AUC of 0.8959 while attaining a malignant recall of 92.68%, correctly identifying 38 of 41 malignant lesions on the independent test set. By optimizing the decision threshold using a recall-constrained strategy, the framework substantially reduced false negatives compared with the conventional threshold, making it particularly suitable for melanoma screening where missing malignant cases carries substantially greater clinical risk than generating false-positive referrals.

Although the proposed approach intentionally sacrifices precision and specificity to maximize sensitivity, this trade-off is appropriate for preliminary clinical triage systems in which suspicious lesions can subsequently undergo specialist examination. Overall, the findings demonstrate that combining cost-sensitive learning, progressive augmentation, and clinically driven threshold optimization provides an effective strategy for improving malignant lesion detection in highly imbalanced dermoscopic datasets. Future research should validate the framework on external multi-center datasets, investigate lightweight architectures for real-world deployment, and incorporate explainable artificial intelligence techniques to improve model interpretability and clinical acceptance.

ACKNOWLEDGEMENT

The researcher sincerely acknowledges the financial support provided by the Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia, through the Faculty Research Grant No. 32.A/UN7.F8/PP/II/2026, which greatly contributed to the successful completion of this research.

Conceptualization: Puspita Kartikasari, methodology: Puspita Kartikasari and Rukun Santoso, software: Rukun Santoso and Novri Suhermi, validation: Puspita Kartikasari and Rizwan Arisandi, formal analysis: Puspita Kartikasari, Rizwan Arisandi, and Parlindungan Situmorang, investigation: Novri Suhermi, resources, writing—original draft preparation, supervision, writing—review and editing: Puspita Kartikasari, data curation: Parlindungan Situmorang, visualization: Rizwan Arisandi, Novri Suhermi and Parlindungan Situmorang; funding

acquisition, Universitas Diponegoro.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

REFERENCES

[1] M. Togacar, Z. Comert, B. Ergen, Intelligent Skin Cancer Detection Applying Autoencoder, MobileNetV2 and Spiking Neural Networks, Chaos Solitons Fractals 144 (2021), 110714. https://doi.org/10.1016/j.chaos.2021.110714.

[2] M.W. Purbandanu, R. Yanuarta, A. Kurniawan, Optimization of Skin Cancer Detection to Improve Accuracy with the Application of Efficient Convolutional Neural Network and EfficientNetB2 Models, J. Intell. Comput. Health Inform. 5 (2024), 43-50. https://doi.org/10.26714/jichi.v5i2.14338.

[3] M. Wang, X. Gao, L. Zhang, Recent Global Patterns in Skin Cancer Incidence, Mortality, and Prevalence, Chin. Med. J. 138 (2024), 185-192. https://doi.org/10.1097/cm9.0000000000003416.

[4] T.M. Alam, K. Shaukat, W.A. Khan, I.A. Hameed, L.A. Almuqren, et al., An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset, Diagnostics 12 (2022), 2115. https://doi.org/10.3390/diagnostics12092115.

[5] I. Suneetha, Hybrid Deep Learning Model for Skin Cancer Classification, E3S Web Conf. 591 (2024), 09010. https://doi.org/10.1051/e3sconf/202459109010.

[6] Y.S. Alsahafi, M.A. Kassem, K.M. Hosny, Skin-Net: a Novel Deep Residual Network for Skin Lesions Classification Using Multilevel Feature Extraction and Cross-Channel Correlation with Detection of Outlier, J. Big Data 10 (2023), 105. https://doi.org/10.1186/s40537-023-00769-6.

[7] J.D.D. Jayaseeli, J. Briskilal, C. Fancy, V. Vaitheeshwaran, R.S.M.L. Patibandla, et al., An Intelligent Framework for Skin Cancer Detection and Classification Using Fusion of Squeeze-Excitation-DenseNet with Metaheuristic-Driven Ensemble Deep Learning Models, Sci. Rep. 15 (2025), 7425. https://doi.org/10.1038/s41598-025-92293-1.

[8] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, I. Ellinge, Skin Lesion Classification Using Hybrid Deep Neural Networks, in: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 1229-1233. https://doi.org/10.1109/ICASSP.2019.8683352.

[9] V. Ravi, Attention Cost-Sensitive Deep Learning-Based Approach for Skin Cancer Detection and Classification, Cancers 14 (2022), 5872. https://doi.org/10.3390/cancers14235872.

[10] S. Shen, M. Xu, F. Zhang, P. Shao, H. Liu, et al., A Low-Cost High-Performance Data Augmentation for Deep Learning-Based Skin Lesion Classification, BME Front. 2022 (2022), 9765307. https://doi.org/10.34133/2022/9765307.

[11] L. Cino, C. Distante, A. Martella, P.L. Mazzeo, Skin Lesion Classification Through Test Time Augmentation and Explainable AI, J. Imaging 11 (2025), 15. https://doi.org/10.3390/jimaging11010015.

[12] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, M.H. Yap, Analysis of the ISIC Image Datasets: Usage, Benchmarks and Recommendations, Med. Image Anal. 75 (2022), 102305. https://doi.org/10.1016/j.media.2021.102305.

[13] M. Martin-Gonzalez, C. Azcarraga, A. Martin-Gil, C. Carpena-Torres, P. Jaen, Efficacy of a Deep Learning Convolutional Neural Network System for Melanoma Diagnosis in a Hospital Population, Int. J. Environ. Res. Public Health 19 (2022), 3892. https://doi.org/10.3390/ijerph19073892.

[14] P.N. Srinivasu, J.G. SivaSai, M.F. Ijaz, A.K. Bhoi, W. Kim, et al., Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM, Sensors 21 (2021), 2852. https://doi.org/10.3390/s21082852.

[15] M. Kreouzi, N. Theodorakis, G. Feretzakis, E. Paxinou, A. Sakagianni, et al., Deep Learning for Melanoma Detection: A Deep Learning Approach to Differentiating Malignant Melanoma from Benign Melanocytic Nevi, Cancers 17 (2024), 28. https://doi.org/10.3390/cancers17010028.

[16] ISIC Archive, ISIC Challenge Datasets, 2024. https://challenge.isic-archive.com/data/#2024.

[17] R. Kaur, H. GholamHosseini, M. Lindén, Advanced Deep Learning Models for Melanoma Diagnosis in Computer-Aided Skin Cancer Detection, Sensors 25 (2025), 594. https://doi.org/10.3390/s25030594.

[18] V. Ravi, Attention Cost-Sensitive Deep Learning-Based Approach for Skin Cancer Detection and Classification, Cancers 14 (2022), 5872. https://doi.org/10.3390/cancers14235872.

[19] A. Wibowo, J.D. Setiawan, H. Afrisal, A.A.S.M.M. Mertha, S.P. Santosa, et al., Optimization of Computational Resources for Real-Time Product Quality Assessment Using Deep Learning and Multiple High Frame Rate Camera Sensors, Appl. Syst. Innov. 6 (2023), 25. https://doi.org/10.3390/asi6010025.

[20] E. Goceri, Medical Image Data Augmentation: Techniques, Comparisons and Interpretations, Artif. Intell. Rev. 56 (2023), 12561-12605. https://doi.org/10.1007/s10462-023-10453-z.

[21] G.M. Silva, A.E. Lazzaretti, F.C. Monteiro, Deep Learning Techniques Applied to Skin Lesion Classification, in: 2023 2nd International Conference on Machine Learning, Control, and Robotics (MLCR), IEEE, 2023, pp. 18-23. https://doi.org/10.1109/mlcr61158.2023.00013.

[22] G.D. Finlayson, E. Trezzi, Shades of Gray and Colour Constancy, Color Imaging Conf. 12 (2004), 37-41. https://doi.org/10.2352/cic.2004.12.1.art00008.

[23] L. Cino, C. Distante, A. Martella, P.L. Mazzeo, Skin Lesion Classification Through Test Time Augmentation and Explainable AI, J. Imaging 11 (2025), 15. https://doi.org/10.3390/jimaging11010015.

[24] K. Cahyanto, K. Adi, C.E. Widodo, The L2-EffCANet: A Novel Overfitting-Resistant EfficientNetV2S with Attention Mechanism and L2 Regularization for Skin Disease Classification, Int. J. Online Biomed. Eng. 21 (2025), 113-129. https://doi.org/10.3991/ijoe.v21i13.57417.

[25] A. Buslaev, V.I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, et al., Albumentations: Fast and Flexible Image Augmentations, Information 11 (2020), 125. https://doi.org/10.3390/info11020125.

[26] O. Wisnu Ardhi, T. Retnaningsih Soeprobowati, K. Adi, E. Prakasa, A. Rachman, Enhanced You Only Look Once Approach for Automatic Phytoplankton Identification, IAES Int. J. Artif. Intell. 13 (2024), 3426. https://doi.org/10.11591/ijai.v13.i3.pp3426-3436.

[27] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 2261-2269. https://doi.org/10.1109/cvpr.2017.243.

[28] S.I. Hussain, E. Toscano, Enhancing Recognition and Categorization of Skin Lesions with Tailored Deep Convolutional Networks and Robust Data Augmentation Techniques, Mathematics 13 (2025), 1480. https://doi.org/10.3390/math13091480.

[29] J. Wu, W. Hu, Y. Wen, W. Tu, X. Liu, Skin Lesion Classification Using Densely Connected Convolutional Networks with Attention Residual Learning, Sensors 20 (2020), 7080. https://doi.org/10.3390/s20247080.

[30] M. Kreouzi, N. Theodorakis, G. Feretzakis, E. Paxinou, A. Sakagianni, et al., Deep Learning for Melanoma Detection: A Deep Learning Approach to Differentiating Malignant Melanoma from Benign Melanocytic Nevi, Cancers 17 (2024), 28. https://doi.org/10.3390/cancers17010028.

[31] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, mixup: Beyond Empirical Risk Minimization, arXiv:1710.09412, 2017. https://doi.org/10.48550/arXiv.1710.09412.

[32] S. Yun, D. Han, S. Chun, S.J. Oh, Y. Yoo, et al., CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2019, pp. 6022-6031. https://doi.org/10.1109/iccv.2019.00612.

[33] O. Rainio, J. Tamminen, M.S. Venäläinen, J. Liedes, J. Knuuti, et al., Comparison of Thresholds for a Convolutional Neural Network Classifying Medical Images, Int. J. Data Sci. Anal. 20 (2024), 2093-2099. https://doi.org/10.1007/s41060-024-00584-z.