Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 5, October 2025, pp. 4086~4099
ISSN: 2302-9285, DOI: 10.11591/eei.v14i5.10379

Enhancing skin cancer detection using transfer learning and AdaBoost: a deep learning approach

Hersatoto Listiyono[1], Retnowati[2], Purwatiningtyas[3], Eko Nur Wahyudi[3], Ali Maskur[4]
[1]Study Program of Graphic Multimedia Engineering Technology (D4), Vocational Faculty, Universitas Stikubank, Semarang, Indonesia
[2]Master's Program in Information Technology (S2), Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Indonesia
[3]Undergraduate Program in Informatics Engineering (S1), Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Indonesia
[4]Study Program of Finance and Banking (D3), Vocational Faculty, Universitas Stikubank, Semarang, Indonesia

Article Info

Article history:

Received Mar 29, 2025
Revised Aug 11, 2025
Accepted Sep 11, 2025

Keywords:

Adaptive boosting
Deep learning
Image classification
International Skin Imaging Collaboration dataset
Skin cancer detection

ABSTRACT

Skin cancer is one of the most prevalent types of cancer worldwide, with early detection playing a critical role in improving patient outcomes. In this study, we propose a deep learning model based on LeNet-7 combined with adaptive boosting (AdaBoost) to classify skin lesions as either benign or malignant using the International Skin Imaging Collaboration (ISIC) dataset. We evaluate the proposed model alongside other well-established deep learning architectures, such as residual network (ResNet), VGGNet, and the traditional LeNet model, through various performance metrics including precision, recall, F1-score, specificity, Matthew's correlation coefficient (MCC), area under the receiver operating characteristic curve (AUC-ROC), and testing accuracy. Our results demonstrate that the proposed model (LeNet-7+AdaBoost) significantly outperforms the other models, achieving a testing accuracy of 91.3%, precision of 0.92, recall of 0.91, and AUC-ROC of 0.93. The model successfully addresses issues of overfitting and generalization, providing a robust solution for skin cancer classification. However, some misclassifications of visually similar benign and malignant lesions highlight areas for future improvement. The proposed model shows promise in real-world medical applications and paves the way for further research into optimizing deep learning models for skin cancer detection.

Corresponding Author:

Hersatoto Listiyono
Study Program of Graphic Multimedia Engineering Technology (D4), Vocational Faculty
Universitas Stikubank
Tri Lomba Juang Road, Mugassari, South Semarang District, Semarang City, Central Java, Indonesia
Email: hersatotolistiyono@edu.unisbank.ac.id

1. INTRODUCTION

Skin cancer is one of the most common types of cancer worldwide, with the number of cases continuing to rise each year [1]. Early detection of skin cancer is crucial for improving patient survival rates, as earlier medical intervention can prevent the progression of cancer to more dangerous stages [2], [3]. However, conventional diagnosis of skin cancer still faces several challenges, including reliance on the expertise of dermatologists, subjectivity in assessments, and limited access to adequate healthcare services in some regions [1], [4]. Therefore, the development of artificial intelligence (AI)-based systems has emerged as a potential solution to enhance the efficiency and accuracy of skin cancer diagnosis [5], [6]. With advancements in deep learning [7]-[9], convolutional neural network (CNN) models have demonstrated excellent performance in medical image classification tasks, including skin cancer detection [10], [11]. One widely used dataset in related research is the International Skin Imaging Collaboration (ISIC), which contains thousands of skin lesion images labeled by dermatology experts [12]-[14]. However, a major challenge in using CNNs for skin cancer classification is the need for a large amount of data and the complexity of training models to achieve high accuracy.

In medical image classification, including skin cancer detection, various CNN architectures have been developed to improve model accuracy and efficiency. Popular architectures such as VGG16 and VGG19 [8], [15], feature deep network structures with stacked convolutional layers designed to capture complex features from medical images. Additionally, residual network (ResNet) [16], [17], including variants like ResNet50 and ResNet101, offers an innovative approach with skip connections that allow deeper networks to be trained without suffering from the vanishing gradient problem [18]. InceptionV3 and InceptionResNetV2 are also widely used due to their ability to capture features across different scales through Inception modules [19], [20]. Meanwhile, models such as [21], [22], which optimize the scale of the network automatically, have proven to be more efficient in medical image classification, achieving high accuracy with fewer parameters. By leveraging these architectures, CNN models can classify skin cancer images more accurately, contributing to better early detection and diagnosis. As described by [23], this article has several notable strengths and weaknesses. A major strength of this study is the use of transfer learning with the EfficientNet V2 architecture, which achieved a high accuracy of 84%, outperforming other models such as Inception V3 (82%) and generic CNNs (81%). Additionally, this study emphasizes the importance of model interpretability, which provides insights into key features influencing predictions, thus increasing trust in the model's recommendations. However, a limitation of this study is its reliance on a limited dataset, consisting of 3,297 images from the ISIC archive, which may not encompass all variations of skin lesions found in real-world scenarios, potentially affecting the model's generalizability. This aligns with the findings of [24], where the study highlighted several strengths and weaknesses. The primary strength of the study was the use of a modified GoogleNet model with transfer learning, which achieved high classification accuracy of 94.92% for eight classes of skin lesions, including melanoma and squamous cell carcinoma, demonstrating great potential for early skin cancer diagnosis. Additionally, the study addressed the issue of class imbalance in the dataset by applying image augmentation techniques, which improved the model’s sensitivity and precision. However, a drawback of this study was that, despite promising results, the initially low sensitivity and precision (53.3% and 62.5%, respectively) indicated that the model could still be improved, particularly in handling classes with very few images.

Based on the research objectives and references outlined, the primary challenge in this study is how to improve the accuracy of skin cancer classification using the ISIC dataset through an optimal deep learning approach [25]. Although CNNs have proven effective in medical image classification, challenges still exist, such as the limitation of adequate training data, the risk of overfitting, and the optimization of the model architecture to capture key features of skin lesions more accurately [26], [27]. Moreover, conventional methods are often not robust enough to handle the complex variations in color, texture, and shape of skin lesions [28]. Therefore, this research focuses on the application of transfer learning and boosting techniques as strategies to enhance the performance of CNN models in detecting skin cancer [21], [29]. By combining these two approaches, it is hoped that a more reliable, efficient model can be developed, capable of providing more accurate classification results to support early skin cancer detection and more precise medical decision-making [30], [31]. In recent years, transfer learning and ensemble learning techniques have emerged as promising approaches to improve the performance of CNN models [32], [33]. Transfer learning enables the use of pre-trained models on large datasets, speeding up the training process and improving the model’s generalization to new data. Meanwhile, boosting, as an ensemble learning technique, aims to combine several weak models into one strong model to improve classification accuracy. The combination of these two approaches has the potential to produce a more accurate and efficient model for detecting skin cancer from medical images.

This study aims to enhance the accuracy of skin cancer classification using the ISIC dataset through deep learning techniques, specifically transfer learning and boosting [28], [34]. By applying the appropriate optimization strategies, this research aims to develop a more reliable model to support early skin cancer diagnosis, ultimately assisting healthcare professionals in making more accurate and timely decisions.

2. METHOD

This study employs a quantitative experimental approach to develop and compare the performance of ISIC classification models using various deep learning architectures. The primary focus of the research is to compare the proposed LeNet-7+adaptive boosting (AdaBoost) model with individual models such as LeNet-7, ResNet, and VGG16. The experiments were conducted using the ISIC image dataset, which underwent preprocessing to ensure the quality of the input data [12], [30]. This chapter provides a detailed explanation of the methods used in the research, including the dataset employed, the proposed models, and the research design, which outlines the stages involved in the experimental process. The goal of this study is to improve the accuracy of skin cancer classification by implementing various CNN architectures, along with transfer learning and boosting techniques, using the ISIC dataset [11], [35].

2.1. Research dataset

The dataset used in this study is the ISIC dataset, which is a standard dataset in medical image research for skin cancer detection. ISIC has become a primary resource in various competitions and scientific studies related to the automated diagnosis of skin cancer using AI. Typically, this dataset is pre-labeled to indicate the type of lesion (benign or malignant), which facilitates the training and evaluation process of classification models.

The data source is from the kaggle.com website (https://www.kaggle.com/competitions/isic-2024-challenge/data). The dataset consists of diagnostic labeled images with additional metadata. The images are in JPEG format. The associated .csv file contains binary diagnostic labels (target), potential input variables (e.g., age_approx, sex, and anatom_site_general), and additional attributes (e.g., image source and exact diagnosis). The SLICE-3D dataset-skin lesion image slices extracted from 3D total body photography (TBP) for skin cancer detection. To mimic non-dermoscopic images, the competition uses standard cropped lesion images of lesions from 3D TBP. The data is divided into 2 parts, namely, 80% training data and 20% test data.

The dataset consists of thousands of skin lesion images that have been classified by dermatology experts into several main categories, including: Figure 1 presents the ISIC dataset used in this study, which focuses on skin cancer classification using deep learning and transfer learning techniques to enhance detection accuracy. The dataset consists of images of skin lesions categorized as benign or malignant, with variations in color, texture, and lighting conditions that reflect the challenges in medical diagnosis. The sample images in the dataset are divided into a training dataset for model training and a testing dataset for evaluating classification performance. Deep learning techniques using CNN architectures are employed to extract features from the images, while transfer learning utilizes pre-trained models such as ResNet, VGG, or Inception to improve classification performance. The primary challenges in this dataset include class imbalance, variations in image quality, and differences in lighting and image angles, which can impact the model's generalization. With the appropriate approach, this study aims to enhance the accuracy of skin cancer detection, contributing to more effective and accurate early diagnosis.

Figure 1 presents representative samples from the ISIC dataset employed in this study, which consists of dermoscopic images of skin lesions categorized into two main classes: benign and malignant. The dataset is divided into two subsets, with the training set used for model development and the testing set reserved for performance evaluation on previously unseen data. The ISIC dataset is widely recognized as a benchmark in skin cancer research due to its high variability in color, texture, size, and illumination, reflecting the real-world challenges faced in clinical diagnosis. This variability provides valuable opportunities for CNNs to learn more robust and discriminative feature representations. However, it also introduces the risk of overfitting, particularly when models are not carefully optimized. As illustrated in Figures 1(a) and (b), the dataset offers a diverse and representative collection of lesion images, making it well-suited for evaluating deep learning approaches in skin cancer classification. In this research, the ISIC dataset serves as the foundation for assessing the proposed LeNet-7 architecture combined with AdaBoost, aiming to improve classification accuracy and enhance the generalization of the model in distinguishing between benign and malignant lesions.

2.2. Proposed models

In this study, four models are proposed for skin cancer classification: LeNet-7, ResNet50, VGG16, and LeNet-7+AdaBoost. Each model has its own strengths and weaknesses in recognizing patterns from skin lesion images. In Table 1, the comparison presents four deep learning models—LeNet-7, ResNet50, VGG16, and LeNet-7 with boosting—based on their methods, main objectives, and potential uses. LeNet-7, which uses a CNN architecture, is designed for simple image classification with high efficiency on small datasets. ResNet50, applying residual learning, aims to address the vanishing gradient problem and extract complex features, making it optimal for large-scale medical datasets. VGG16, with its deep-layer CNN architecture, has a deeper feature extraction capability, resulting in high accuracy but requiring significant computational power. Meanwhile, LeNet-7 with boosting combines CNN with boosting techniques to improve accuracy through the combination of multiple models, offering a lighter alternative compared to more complex models. This comparison highlights that each model has its own strengths and limitations, meaning that the choice of model should be tailored to the dataset's needs and available computational resources. Figure 2, summarizing the standard LeNet-7 model and the LeNet-7 model enhanced with boosting.

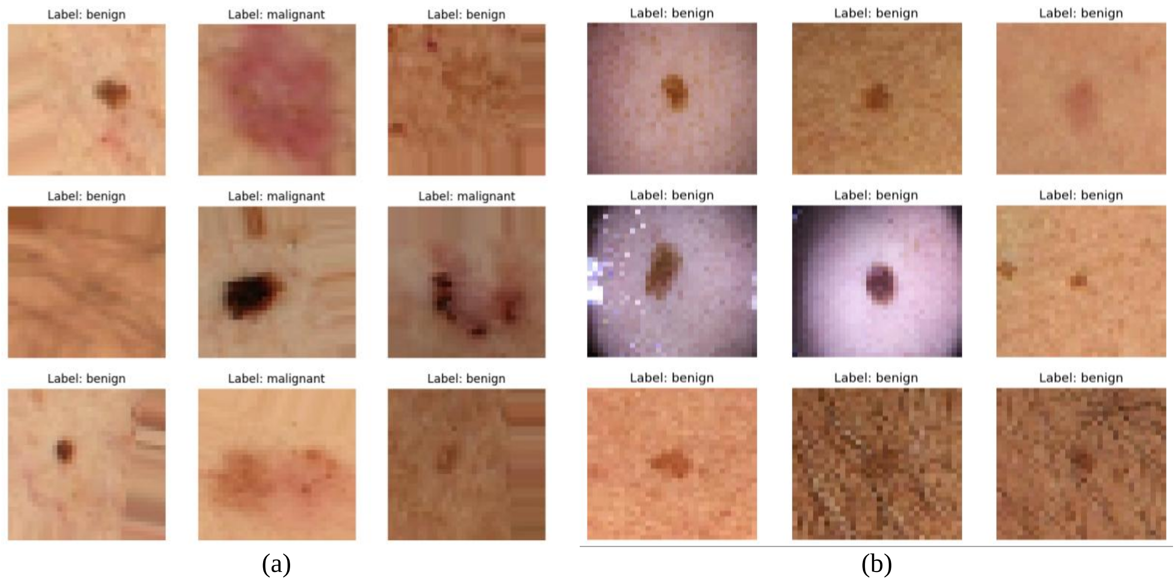


Figure 1. ISIC research dataset; (a) training dataset sample images and (b) testing dataset sample images

Table 1. Comparison of models

Model	Method	Main objective	Potential use
LeNet-7	CNN	Simple and fast image classification	Efficient for tasks with small datasets
ResNet50	Residual learning	Overcoming vanishing gradient and complex feature extraction	Optimal for large medical datasets
VGG16	Deep layer CNN	Detailed feature extraction with a deep architecture	High accuracy but requires significant computational power
LeNet-7+boosting	CNN+boosting	Improving accuracy through a combination of multiple models	A lighter alternative compared to more complex models

Figure 2 illustrates the architectures of LeNet-7 and LeNet-7+AdaBoost, which are two CNN-based models used for image classification. LeNet-7 serves as the baseline model, while LeNet-7+AdaBoost is an enhanced version incorporating boosting techniques. LeNet-7 Figure 2(a) consists of two convolutional layers, each followed by a max-pooling layer. The first convolutional layer has 456 parameters, and the second one has 2,416 parameters, both designed to extract spatial features from the images. After feature extraction, the data is flattened and passed through three dense layers with 129, 84, and 2 neurons, respectively, leading to a total of 10,164 parameters before the output layer.

Meanwhile, LeNet-7+AdaBoost Figure 2(b) maintains the basic structure of LeNet-7 but with several modifications to improve classification performance. The area circled with a red dashed line represents the optimization applied to the LeNet-7 architecture. This model has a higher number of parameters in each layer, with the first convolutional layer containing 896 parameters and the second convolutional layer containing 3,168 parameters, indicating more complex feature extraction. Additionally, the model includes a dropout layer to reduce overfitting and two extra dense layers with 128 neurons before reaching the output layer. A significant improvement in LeNet-7+AdaBoost is the incorporation of AdaBoost classifier, which serves as a decision enhancer by combining several weak models to increase classification accuracy. As a result, LeNet-7+AdaBoost performs better in handling dataset complexity, particularly in improving the model's generalization ability for more diverse images.

2.3. Research design

The research framework in this study is essential as it provides a systematic guide for each stage of the research, from data processing to result evaluation. In this study, the framework helps organize the comparison process of various CNN model architectures, such as LeNet, ResNet, VGGNet, and the proposed model (ResNet+Bagging). With a clear framework, researchers can perform comparisons transparently and accurately, allowing for a better understanding of each phase of the research. Additionally, this framework minimizes errors and ensures a comprehensive evaluation using various performance metrics, making the research results more valid and accountable. The research framework is shown in Figure 3.

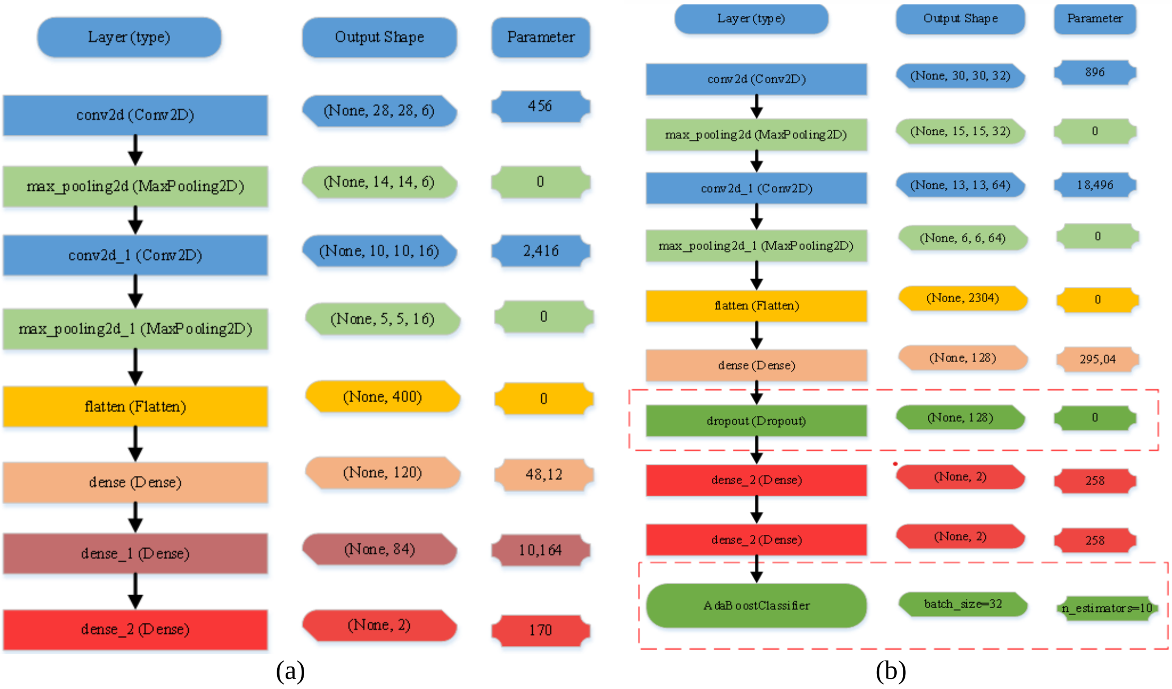


Figure 2. The illustrates architectures; (a) LeNet-7 and (b) LeNet-7+AdaBoost

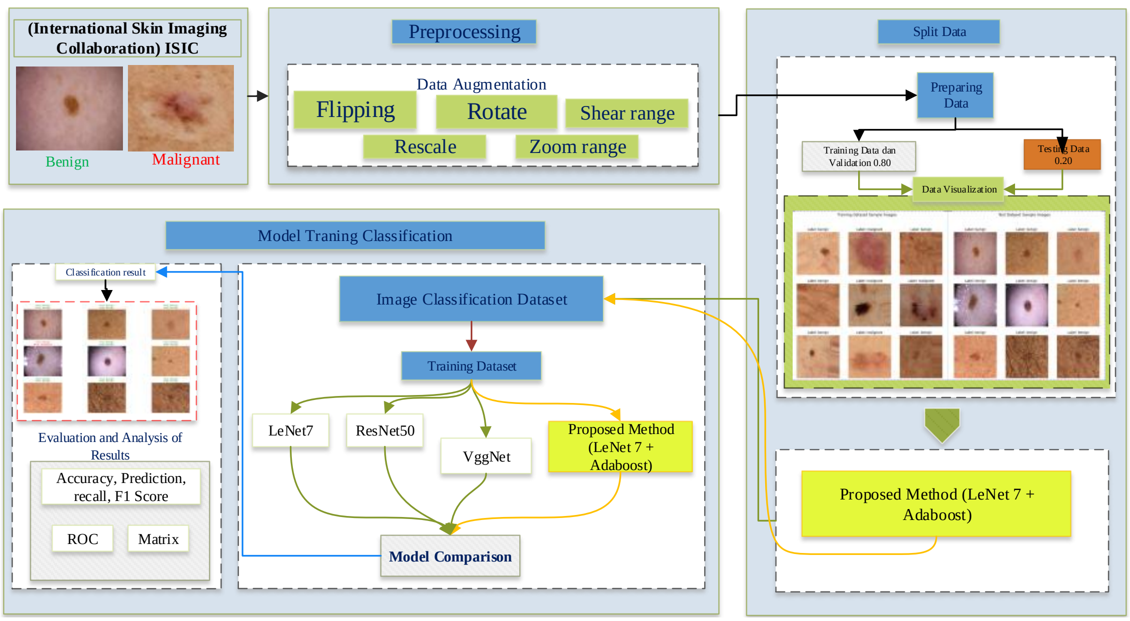


Figure 3. Research framework

The research framework, as depicted in Figure 3, outlines a workflow that begins with a skin CT scan dataset, which consists of two classes: benign and malignant. The stages of the research are as follows: First, data preprocessing is conducted, where the image data undergoes augmentation techniques to increase the variety within the training data. The augmentation methods applied include flipping, rotating, shearing, and rescaling. This step is aimed at enriching the dataset to improve the model's ability to recognize a wider range of patterns. Next, the dataset is divided into three parts: training data, which comprises 80% of the total data and is used for model training; testing data, which makes up the remaining 20% and is used to assess the final performance of the model after training; and data visualization, which allows for checking the distribution of data across each class. The third stage involves model training and classification, where the processed data is used to train various model architectures, such as LeNet-7, ResNet, VGGNet, and the modified proposed model (LeNet-7+Adaboost). All models were trained with the same dataset to develop a skin cancer classification model. Following the training, the fourth stage involves model comparison, where the performance of the LeNet-7, ResNet, and VGGNet models is compared to the proposed model. This comparison aims to evaluate how modifications in architecture and hyperparameter tuning have improved the accuracy of the proposed model. Finally, in the evaluation and analysis stage, the classification results from each model are evaluated and analyzed using performance metrics such as accuracy, precision, recall, F1-score, Matthew's correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC-ROC). The results are then visualized for easier interpretation and further analysis.

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

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

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

$$\mathrm{F1-score} = \frac{2\times\mathrm{precision}\times\mathrm{recall}}{\mathrm{precision+recall}} \quad (4)$$

$$\mathrm{MCC} = \frac{\mathrm{TPxTN-FPxFN}}{\sqrt{(TP+FP)(TP+FN)(TN+FP)TN+FN)}} \quad (5)$$

$$\mathrm{AUC-ROC\ (TPR)} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (6)$$

$$\mathrm{AUC-ROC\ (FRP)} = \frac{\mathrm{FP}}{\mathrm{FP+TN}} \quad (7)$$

3. RESULTS AND DISCUSSION

3.1. Data preprocessing and data augmentation

In the preprocessing stage, the images from the ISIC undergo several data enhancement processes aimed at improving the quality and diversity of the dataset. These processes include flipping, where the image is mirrored either horizontally or vertically to create variations in orientation. Additionally, rotation is applied by rotating the image at specific angles, which helps to increase the diversity within the dataset. Another technique, shearing, involves applying a shear transformation that creates slight distortions in the image, further enriching the dataset. Lastly, rescaling normalizes the pixel values to a range of [0, 1], which aids the model in better recognizing patterns within the data. The results of these augmentations are shown in Figure 4.

Figure 4. Sample ISIC scan dataset images after augmentation

Figure 4 displays the results of the data augmentation process on the ISIC scan images, which significantly enriches the dataset by introducing various variations. Augmentation techniques such as flipping, rotating, and rescaling effectively enhance the model's ability to generalize. The purpose of this augmentation is to ensure that the model does not become overly dependent on specific patterns and can adapt better to diverse test data. In this study, the experimental results are derived from the classification of brain scan images into two categories: hemorrhage and non-hemorrhage. The dataset undergoes preprocessing, including data augmentation using techniques like flipping, rotating, shearing, and rescaling. After preprocessing, the dataset is divided into training data (80%), validation data (10%), and testing data (10%). This process ensures that the model is evaluated on data it has never seen before, providing objective and realistic results.

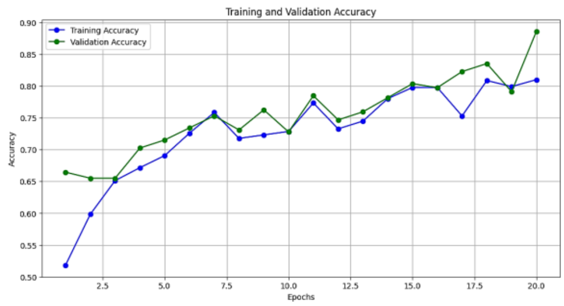
3.2. Model training classification

Several CNN models, such as LeNet, ResNet, VGGNet, and the proposed model, were tested and showed varying performance in ISIC classification. Figure 5 illustrates the training accuracy curves. Upon analyzing the training and validation accuracy curves of the models presented (Figure 5), we can observe significant differences in their performance. The LeNet model (Figure 5(a)) shows fluctuating training accuracy throughout the epochs, with no clear, consistent upward trend. The validation accuracy follows a similar pattern, often lagging behind the training accuracy. This suggests that LeNet may struggle to generalize well to unseen data, potentially due to overfitting, as the model's performance is inconsistent both during training and validation. The ResNet model (Figure 5(b)) demonstrates a gradual improvement in training accuracy, although it still exhibits fluctuations. Validation accuracy also shows some inconsistency, with occasional spikes and drops. While ResNet has some improvement over LeNet, its instability in validation accuracy points to difficulties in optimization and generalization to new data. The model shows moderate improvement but does not reach a stable or high level of performance. The VGGNet model (Figure 5(c)) shows a more stable increase in training accuracy compared to the previous two models. However, validation accuracy continues to lag behind, indicating some level of overfitting. Despite its smooth progress, the gap between the training and validation accuracies implies that VGGNet may be overfitting to the training data, as it struggles to generalize well to unseen data. In contrast, the proposed model (LeNet-7+Adaboost) (Figure 5(d)) outperforms all other models. It shows a steady and consistent increase in both training and validation accuracy, with minimal fluctuations across the epochs. The validation accuracy closely follows the training accuracy, indicating that the model is not only learning effectively but also generalizing well to new data. This model, combining LeNet-7 with Adaboost, exhibits the best performance in both training stability and generalization capability.

(a)

(b)

(c)



(d)

Figure 5. Training accuracy curves; (a) LeNet, (b) ResNet, (c) VGGNet, and (d) the proposed model

Upon analyzing the training and validation loss curves for each model (Figure 6), several key differences in performance can be observed. The LeNet model (Figure 6(a)) shows a gradual decrease in training loss, although it fluctuates somewhat throughout the epochs. The validation loss follows a similar trend but decreases at a slower rate and with more noticeable fluctuations. This indicates that while LeNet is improving during training, it struggles to generalize effectively to unseen data, as seen in the higher and more variable validation loss, suggesting potential overfitting. The ResNet model (Figure 6(b)) also exhibits a decreasing training loss, but the curve is more erratic, with several spikes indicating difficulties during the training process. The validation loss fluctuates significantly, failing to show a consistent downward trend. This variability suggests that ResNet, while showing some improvement, does not generalize well and struggles to handle the complexity of the dataset in a stable manner. The VGGNet model (Figure 6(c)) demonstrates a smoother decline in training loss, with fewer fluctuations compared to LeNet and ResNet. The validation loss also shows a steady decline, but the gap between training and validation loss remains visible, indicating that VGGNet may be prone to overfitting, as it is not able to generalize well to unseen data despite its strong performance during training. In contrast, the proposed model (LeNet-7+Adaboost) (Figure 6(d)) stands out as the best performer. Both the training and validation loss decrease smoothly and consistently throughout the epochs. The minimal fluctuation in both loss curves indicates that this model is learning effectively and generalizing well to unseen data. The validation loss follows the same smooth downward trend as the training loss, suggesting that the proposed model is not overfitting and can handle the dataset's complexity with superior generalization.

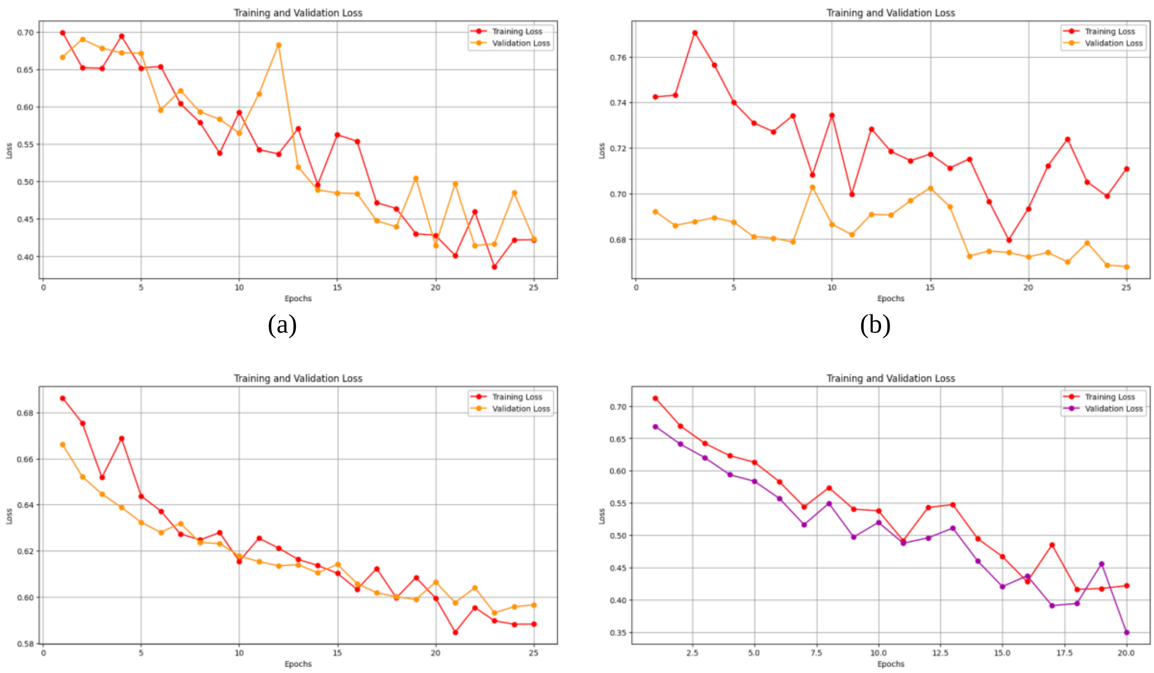
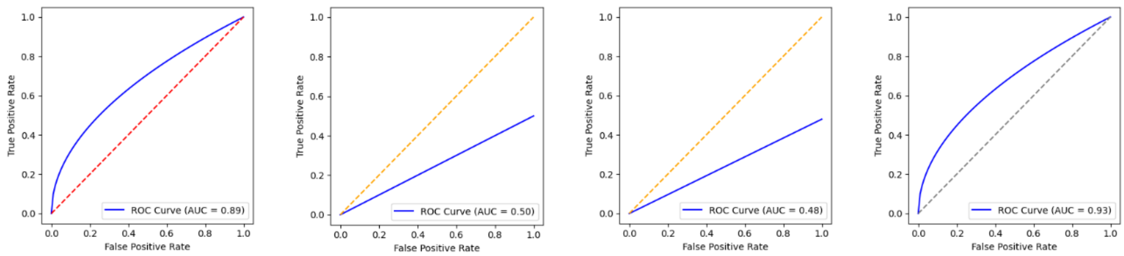


Figure 6. Training loss curves; (a) LeNet, (b) ResNet, (c) VGGNet, and (d) the proposed model

Upon analyzing the receiver operating characteristic (ROC) curves for the four models (Figure 7), we can observe notable differences in their performances, as indicated by the area under the curve (AUC) values and the shapes of their respective curves. The LeNet-7 model (Figure 7(a)) demonstrates strong performance, with an AUC of 0.89, reflecting its ability to correctly classify data points with high accuracy. The ROC curve for LeNet-7 shows a steady increase, indicating that the model effectively distinguishes between positive and negative classes. The curve is well above the diagonal line (representing a random classifier), and while it is not perfect, it indicates that the model performs significantly better than random guessing. The AUC of 0.89 places this model in the upper range of performance, but there is still some room for improvement, especially when compared to the proposed model. The ResNet model (Figure 7(b)), with an AUC of 0.50, presents a stark contrast to LeNet-7. The ROC curve is nearly a straight diagonal line, indicating that the model's performance is close to random guessing. An AUC of 0.50 is equivalent to a classifier that has no discriminative power between the two classes. This suggests that ResNet is either poorly trained or not suited for this particular classification task, as its performance does not exceed the level of chance. Similarly, the VGGNet model (Figure 7(c)) shows a poor performance, with an AUC of 0.48, which is slightly worse than ResNet. The ROC curve for VGGNet is almost identical to that of ResNet, showing that the model struggles to differentiate between positive and negative classes. This AUC value indicates a weak classifier, which misclassifies nearly as often as it classifies correctly. The model's slight deviation below the diagonal line suggests that it is slightly worse than random guessing, which indicates a need for either retraining or adjustment in the architecture for improved performance. In contrast to the other models, the proposed model (LeNet-7+Adaboost) (Figure 7(d)) achieves an exceptional performance, as reflected in its AUC of 0.93. The ROC curve for this model is steep and remains well above the diagonal, indicating that the model effectively distinguishes between the two classes with very few misclassifications. This near-perfect curve suggests that the model is highly effective, able to identify true positives while minimizing false positives and false negatives. With an AUC of 0.93, the proposed model demonstrates superior performance, making it the most reliable model in this analysis.



(a) (b) (c) (d)

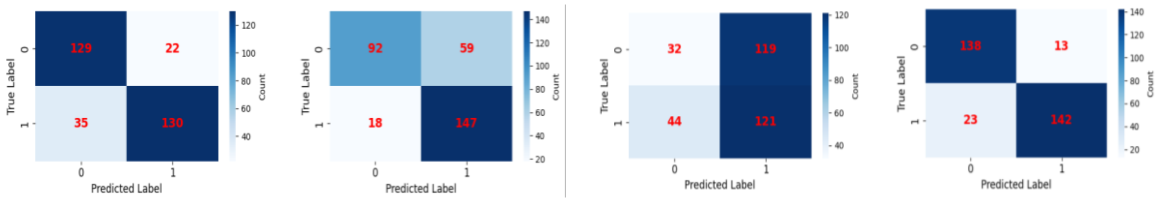
Figure 7. ROC curves; (a) LeNet, (b) ResNet, (c) VGGNet, and (d) the proposed model

3.3. Discussion

The results of training and testing revealed the confusion matrix for each model, demonstrating the classification capabilities across three categories: benign, malignant, and normal. The confusion matrix for each model can be seen in Figure 8. Upon analyzing the confusion matrices for each model, the LeNet model Figure 8(a) shows a moderate performance with 129 benign images and 130 malignant images correctly classified. However, it also exhibits considerable misclassification, with 22 benign images incorrectly predicted as malignant and 35 malignant images incorrectly classified as benign. This suggests that while LeNet is effective to some extent, it struggles with generalization, particularly in distinguishing between the two categories. The ResNet model (Figure 8(b)) performs better than LeNet, correctly predicting 92 benign images and 147 malignant images. However, it still misclassifies 59 benign images as malignant and 18 malignant images as benign. The high number of false positives suggests that the model may be overly sensitive in predicting malignant cases, leading to a higher rate of misclassifying benign images. The VGGNet model (Figure 8(c)) exhibits the poorest performance, with only 32 benign images correctly classified and 121 malignant images correctly predicted. It shows a high rate of misclassification, particularly in benign cases, where 119 benign images are misclassified as malignant. Additionally, 44 malignant images are incorrectly predicted as benign, indicating that VGGNet struggles significantly with distinguishing between the categories, making it prone to both false positives and false negatives. The last, the proposed model (LeNet-7+Adaboost) (Figure 8(d)) performs significantly better than the others in terms of both true positives and true negatives. This model correctly identifies 142 malignant images and 138 benign images, with only 13 benign images misclassified as malignant and 23 malignant images misclassified as benign. These low numbers of misclassifications demonstrate that the proposed model is highly effective at distinguishing between benign and malignant cases, minimizing both false positives and false negatives, and showcasing a strong ability to generalize to new, unseen data.

In Table 2, the LeNet-7 model demonstrates strong performance across various metrics, achieving a precision of 0.87, recall of 0.85, and an F1-score of 0.86. It also has a high specificity of 0.88 and an AUC-ROC of 0.89, which indicates its capability to classify both benign and malignant cases effectively. The testing accuracy of 86.5% further supports its reliable performance. On the other hand, ResNet performs significantly lower, with a precision of 0.52, recall of 0.50, and an F1-score of 0.51. The specificity is low (0.49), and its MCC (0.02) indicates a very weak model that is struggling with both types of errors. Additionally, the testing accuracy of 50.2% confirms that this model is far less effective in comparison.

Similarly, VGGNet shows poor performance with a precision of 0.48, recall of 0.46, and F1-score of 0.47. Its specificity (0.45) and AUC-ROC (0.48) are also quite low, which reflects its inability to accurately distinguish between benign and malignant categories. With a testing accuracy of only 48.7%, this model is not suitable for reliable skin cancer detection. The proposed model (LeNet-7+AdaBoost) outperforms all other models by a large margin, achieving a precision of 0.92, recall of 0.91, and an F1-score of 0.91. The specificity (0.94) and AUC-ROC (0.93) are the highest among the models, indicating its excellent ability to classify both benign and malignant cases with minimal misclassifications. With a testing accuracy of 91.3%, this model is the most accurate and reliable for skin cancer detection among the ones tested. The proposed model (LeNet-7+AdaBoost) is the best-performing model in terms of all metrics, including precision, recall, F1-score, specificity, and AUC-ROC. It significantly outperforms LeNet-7, ResNet, and VGGNet, making it the most suitable model for accurate and reliable skin cancer classification. Below are some images generated by the proposed model, showing the classification results, as seen in Figure 9.



(a) (b) (c) (d)

Figure 8. Confusion matrix; (a) LeNet, (b) ResNet, (c) VGGNet, and (d) the proposed model

Table 2. Testing results of the compared models

Model	Precision	Recall	F1-score	Specificity	MCC	AUC-ROC	Testing accuracy (%)
LeNet-7	0.87	0.85	0.86	0.88	0.78	0.89	86.5
ResNet	0.52	0.50	0.51	0.49	0.02	0.50	50.2
VGGNet	0.48	0.46	0.47	0.45	-0.01	0.48	48.7
LeNet-7+AdaBoost	0.92	0.91	0.91	0.94	0.85	0.93	91.3

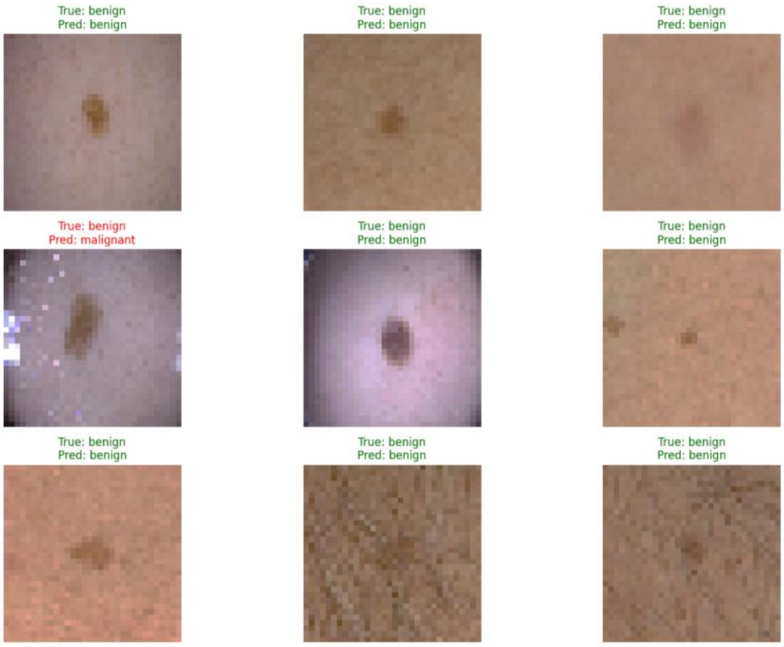


Figure 9. The classification results

Upon analyzing the classification results shown in Figure 9, it is evident that the proposed model (LeNet-7+AdaBoost) performs well in classifying skin lesions, but there are some areas that could be improved. In the top-left quadrant, the model correctly classifies benign lesions as benign, with labels indicating "True: benign, Pred: benign," which demonstrates its ability to accurately identify benign lesions. Similarly, in the top-right quadrant, the model continues to make correct predictions for benign lesions, showcasing its strength in handling variations in benign lesions. However, in the middle-left quadrant, a misclassification occurs where the model predicts "benign" for a malignant lesion, indicating that there are instances where the model struggles to distinguish between benign and malignant cases, especially when the lesions appear visually similar. This issue is further highlighted in the bottom-left quadrant, where another benign lesion is incorrectly classified as malignant. These errors suggest that while the model is highly effective, it still faces challenges in differentiating between certain benign and malignant lesions, particularly when the images share similar visual characteristics. On the other hand, the middle-right and bottom-right quadrants show successful classifications where benign lesions are accurately predicted as benign, reinforcing the model's overall ability to classify images with distinct features correctly.

4. CONCLUSION

This study presents an in-depth evaluation of several deep learning models for skin cancer classification using the ISIC dataset, focusing on the effectiveness of the proposed model (LeNet-7+AdaBoost). The results demonstrate that the proposed model significantly outperforms traditional models, including LeNet-7, ResNet, and VGGNet, in terms of classification accuracy, precision, recall and F1-score, specificity, MCC and AUC-ROC. The proposed model (LeNet-7+AdaBoost) achieved superior results, with a testing accuracy of 91.3%, precision of 0.92, recall of 0.91, and an AUC-ROC of 0.93. These results suggest that the model not only excels at distinguishing malignant from benign lesions but also demonstrates high generalization capabilities across different data distributions. The model's strong performance is attributed to the integration of the AdaBoost algorithm with LeNet-7, which effectively reduces overfitting and improves model robustness, particularly in dealing with the inherent challenges in skin cancer detection, such as dataset imbalance and feature variations. Despite the overall success, there are areas for improvement. Some misclassifications occurred, particularly when distinguishing between benign and malignant lesions that appear visually similar. These misclassifications highlight the need for further refinement in the model's decision-making process, such as enhancing feature extraction capabilities or incorporating more diverse data to address edge cases. In comparison, ResNet and VGGNet performed inadequately, with lower accuracy and more significant misclassification rates. These models struggled with overfitting and failed to generalize well, particularly when predicting benign lesions, leading to higher false positive rates. In conclusion, the proposed model (LeNet-7+AdaBoost) provides a highly reliable and efficient approach to skin cancer classification, showing excellent promise for real-world applications in medical diagnostics. Future work will focus on optimizing the model further, exploring hybrid architectures, and incorporating larger, more diverse datasets to enhance its generalization power, ultimately contributing to more accurate and effective skin cancer detection systems.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to colleagues at Universitas Stikubank for their valuable suggestions during discussions and for providing access to computing resources that supported the experiments in this research. The authors also acknowledge the technical assistance received from the Faculty of Information Technology and Industry in facilitating dataset management and laboratory support.

FUNDING INFORMATION

The authors declare that no specific funding was received for this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Hersatoto Listiyono	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	
Retnowati		✓			✓		✓	✓	✓	✓	✓	✓		
Purwatiningtyas	✓		✓	✓		✓			✓	✓		✓	✓	✓
Eko Nur Wahyudi		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Ali Maskur					✓		✓		✓			✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

INFORMED CONSENT

All images and metadata used in this research were obtained from the publicly available ISIC dataset. Since no direct interaction with human participants occurred, informed consent was not required.

ETHICAL APPROVAL

This study used secondary, publicly available datasets (ISIC archive). Therefore, ethical approval from an institutional review board was not required, as no new data collection involving human participants or animals was conducted.

DATA AVAILABILITY

The dataset supporting the findings of this study is publicly available from the ISIC Challenge repository at: https://www.kaggle.com/competitions/isic-2024-challenge/data.

REFERENCES

[1] A. K. Shakya, A. Ramola, and A. Vidyarthi, "Investigating Novel Coronavirus Through Statistical Methods for Biomedical Applications: An Engineering Student Perspective," in *Emergent Converging Technologies and Biomedical Systems: Select Proceedings of ETBS 2021*, 2022, pp. 341–347, doi: 10.1007/978-981-16-8774-7_28.

[2] M. H. Imam, N. Nahar, M. A. Rahman, and F. Rabbi, "Enhancing skin cancer classification using a fusion of Densenet and Mobilenet models: a deep learning ensemble approach," *Multidisciplinary Science Journal*, vol. 6, no. 7, p. 2024117, 2024, doi: 10.31893/multiscience.2024117.

[3] D. Deepa, V. Muthukumaran, V. Vinodhini, S. Selvaraj, M. S. Kumar, and J. Prabhu, "Uncertainty Quantification to Improve the Classification of Melanoma and Basal Skin Cancer Using ResNet Model," *Journal of Uncertain Systems*, vol. 16, no. 1, pp. 1–16, 2023, doi: 10.1142/S1752890922420107.

[4] N. J. Devabathini and P. Mathivanan, "Sign Language Recognition Through Video Frame Feature Extraction using Transfer Learning and Neural Networks," in *2023 International Conference on Next Generation Electronics (NEleX)*, 2023, pp. 1–6, doi: 10.1109/NEleX59773.2023.10421383.

[5] N. N. Prakash, V. Rajesh, D. L. Namakhwa, S. D. Pande, and S. H. Ahammad, "A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis," *Scientific African*, vol. 20, pp. 1–16, 2023, doi: 10.1016/j.sciaf.2023.e01629.

[6] K. L. Kohsasih and B. H. Hayadi, "Classification SARS-CoV-2 Disease based on CT-Scan Image Using Convolutional Neural Network," *Scientific Journal of Informatics*, vol. 9, no. 2, pp. 197–204, 2022, doi: 10.15294/sji.v9i2.36583.

[7] P. Alkhairi, W. Wanayumini, and B. H. Hayadi, "Analysis of the adaptive learning rate and momentum effects on prediction problems in increasing the training time of the backpropagation algorithm," *AIP Conference Proceedings*, vol. 3048, no. 1, p. 20049, 2024, doi: 10.1063/5.0203374.

[8] R. Pavez, J. Diaz, J. Arango-Lopez, D. Ahumada, C. Mendez-Sandoval, and F. Moreira, "Emo-mirror: a proposal to support emotion recognition in children with autism spectrum disorders," *Neural Computing and Applications*, vol. 35, no. 11, pp. 7913–7924, 2023, doi: 10.1007/s00521-021-06592-5.

[9] A. H. Shahid and M. P. Singh, "A Novel Approach for Coronary Artery Disease Diagnosis using Hybrid Particle Swarm Optimization based Emotional Neural Network," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1568–1585, 2020, doi: 10.1016/j.bbe.2020.09.005.

[10] K. Hauser *et al.*, "Explainable artificial intelligence in skin cancer recognition: A systematic review," *European Journal of Cancer*, vol. 167, pp. 54–69, 2022, doi: 10.1016/j.ejca.2022.02.025.

[11] N. H. Khan *et al.*, "Skin cancer biology and barriers to treatment: Recent applications of polymeric micro/nanostructures," *Journal of Advanced Research*, vol. 36, pp. 223–247, 2022, doi: 10.1016/j.jare.2021.06.014.

[12] R. Mohakud and R. Dash, "Skin cancer image segmentation utilizing a novel EN-GWO based hyper-parameter optimized FCEDN," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 9889–9904, 2022, doi: 10.1016/j.jksuci.2021.12.018.

[13] S. E. Cerminara *et al.*, "Diagnostic performance of augmented intelligence with 2D and 3D total body photography and convolutional neural networks in a high-risk population for melanoma under real-world conditions: A new era of skin cancer screening?," *European Journal of Cancer*, vol. 190, pp. 1–13, 2023, doi: 10.1016/j.ejca.2023.112954.

[14] A. Abdelhafeez and H. K. Mohamed, "Skin cancer detection using neutrosophic c-means and fuzzy c-means clustering algorithms," *Journal of Intelligent Systems and Internet of Things*, vol. 8, no. 1, pp. 33–42, 2023, doi: 10.54216/JISIoT.080103.

[15] S. Shivadekar, B. Kataria, S. Hundekari, K. Wanjale, V. P. Balpande, and R. Suryawanshi, "Deep Learning Based Image Classification of Lungs Radiography for Detecting COVID-19 using a Deep CNN and ResNet 50," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 1s, pp. 241–250, 2023, doi: 10.2139/ssrn.4951255.

[16] S. Sowmya and D. Jose, "Contemplate on ECG signals and classification of arrhythmia signals using CNN-LSTM deep learning model," *Measurement: Sensors*, vol. 24, pp. 1–7, 2022, doi: 10.1016/j.measen.2022.100558.
[17] G. Rajeshkumar *et al.*, "Smart office automation via faster R-CNN based face recognition and internet of things," *Measurement: Sensors*, vol. 27, pp. 1–9, 2023, doi: 10.1016/j.measen.2023.100719.
[18] Z. A. Sejuti and M. S. Islam, "A hybrid CNN–KNN approach for identification of COVID-19 with 5-fold cross validation," *Sensors International*, vol. 4, pp. 1–11, 2023, doi: 10.1016/j.sintl.2023.100229.
[19] S. Sarraf, D. D. Desouza, J. A. E. Anderson, and C. Saverino, "MCADNNet: Recognizing stages of cognitive impairment through efficient convolutional fMRI and MRI neural network topology models," *IEEE Access*, vol. 7, pp. 155584–155600, 2019, doi: 10.1109/ACCESS.2019.2949577.
[20] J. M. Ede and R. Beanland, "Adaptive learning rate clipping stabilizes learning," *Machine Learning: Science and Technology*, vol. 1, no. 1, pp. 1–11, 2020, doi: 10.1088/2632-2153/ab81e2.
[21] L. I. Kesuma, Ermatita, and Erwin, "ELREI: Ensemble Learning of ResNet, EfficientNet, and Inception-v3 for Lung Disease Classification based on Chest X-Ray Image," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 5, pp. 149–161, 2023, doi: 10.22266/ijies2023.1031.14.
[22] D. Kwak, J. Choi, and S. Lee, "Rethinking Breast Cancer Diagnosis through Deep Learning Based Image Recognition," *Sensors*, vol. 23, no. 4, pp. 1–17, 2023, doi: 10.3390/s23042307.
[23] I. Z. Renu *et al.*, "A Comprehensive Analysis on Skin Cancer Classification Using Transfer Learning," in *2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, 2024, pp. 1–6, doi: 10.1109/ICAEEE62219.2024.10561770.
[24] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020, doi: 10.1109/ACCESS.2020.3003890.
[25] S. Defit, A. P. Windarto, and P. Alkhairi, "Comparative Analysis of Classification Methods in Sentiment Analysis: The Impact of Feature Selection and Ensemble Techniques Optimization," *Telematika*, vol. 17, no. 1, pp. 52–67, 2024, doi: 10.35671/telematika.v17i1.2824.
[26] M. Ponnambalam, M. Ponnambalam, and S. S. Jamal, "A robust color image encryption scheme with complex whirl wind spiral chaotic system and quadrant-wise pixel permutation," *Physica Scripta*, vol. 99, no. 10, p. 105239, Sep. 2024, doi: 10.1088/1402-4896/ad7075.
[27] M. Ponnambalam and P. Mathivanan, "A modified Astable multi-vibrator-driven 3D chaotic circuit with Dual LC band stop filters," *Physica Scripta*, vol. 99, no. 6, p. 65265, May. 2024, doi: 10.1088/1402-4896/ad4c28.
[28] R. Li, R. Gao, and P. N. Suganthan, "A decomposition-based hybrid ensemble CNN framework for driver fatigue recognition," *Information Sciences*, vol. 624, pp. 833–848, 2023, doi: 10.1016/j.ins.2022.12.088.
[29] E. H. Houssein, M. M. Emam, and A. A. Ali, "An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators algorithm," *Neural Computing and Applications*, vol. 34, no. 20, pp. 18015–18033, 2022, doi: 10.1007/s00521-022-07445-5.
[30] A. Ramola, A. K. Shakya, and D. Van Pham, "Study of statistical methods for texture analysis and their modern evolutions," *Engineering Reports*, vol. 2, no. 4, pp. 1–24, 2020, doi: 10.1002/eng2.12149.
[31] A. K. Shakya, A. Ramola, S. Singh, and A. Vidyarthi, "Optimized Design of Plasmonic Biosensor for Cancer Detection: Core Configuration and Nobel Material Coating Innovation," *Plasmonics*, vol. 20, no. 4, pp. 1789–1810, 2025, doi: 10.1007/s11468-024-02400-7.
[32] B. Sterniczuk and M. Charytanowicz, "An Ensemble Transfer Learning Model for Brain Tumors Classification using Convolutional Neural Networks," *Advances in Science and Technology Research Journal*, vol. 18, no. 8, pp. 204–216, 2024, doi: 10.12913/22998624/193627.
[33] I. W. Harsono, S. Liawatimena, and T. W. Cenggoro, "Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 3, pp. 567–577, 2022, doi: 10.1016/j.jksuci.2020.03.013.
[34] N. Razali, A. Mustapha, N. Arbaiy, and P. C. Lin, "Deep Learning for Football Outcomes Prediction based on Football Rating System," *AIP Conference Proceedings*, vol. 2644, no. 1, p. 040007, 2022, doi: 10.1063/5.0104587.
[35] J. Ferguson, V. Eleftheriadou, and J. Nesnas, "Risk of Melanoma and Nonmelanoma Skin Cancer in People with Vitiligo: United Kingdom Population−Based Cohort Study," *Journal of Investigative Dermatology*, vol. 143, no. 11, pp. 2204–2210, 2023, doi: 10.1016/j.jid.2023.04.013.

BIOGRAPHIES OF AUTHORS

Hersatoto Listiyono was born in Pati - Central Java, June 30, 1972. He completed Bachelor in Informatics Management degrees at Stikubank University 1998 and Master in Computer Science degrees at Gadjah Mada University 'UGM' Yogyakarta 2010. He has been an active lecturer at Stikubank University since 1998, teaching in artificial intelligence, information systems, database, and decision support system. His research focuses on artificial intelligence, information systems, database, and decision support system. He have won research grant proposals from DIKTI in 2012, 2013, 2017. He can be contacted at email: hersatotolistiyono@edu.unisbank.ac.id.

Retnowati was born in Salatiga on December 22, 1971. Education: Doctoral degrees at Satya Wacana Christian University (SWCU) Salatiga in 2019, Master of Science in Information Systems and Operations Management in- 2000 at University of Manchester Institute of Science and Technology (UMIST), Bachelor of Sains in 1994 at Immanuel Christian University.Yogyakarta. She has been an active lecturer at Stikubank University since 2017, teaching in the information systems management project, software engineering, and research methodology. Her research focuses on technology acceptance model, information systems development, and gamification. She has won research grant proposals from DIKTI in 2023-2024. She can be contacted at email: retnowati@edu.unisbank.ac.id.

Purwatiningtyas was born in Kediri-East Java, August 17, 1966. Been lecture Stikubank since 1992. Completed his Masters Degree from Gadjah Mada University in 2002 in the computer science study program, Faculty of Mathematics and Natural Sciences. She has been active lecture in the information system, project management, and decision support system. Her research focuses on information system, E-Gov, and DSS. She has received several applied research grants from the Ministry of Research, Technology and Higher Education in 2014, 2015, 2019, 2021, 2022, and 2023. She can be contacted at email: purwati@edu.unisbank.ac.id.

Eko Nur Wahyudi was born in Semarang on September 12, 1971. He completed Bachelor in Informatics Management degrees at Stikubank University 1999 and Master in Computer Science degrees at Gadjah Mada University 'UGM' Yogyakarta 2010. He has been an active lecturer at Stikubank University since 1999, teaching in the information systems, data visualization and information, and data mining. His research focuses on artificial intelligence (decision support systems, data mining, and genetic algorithms). He can be contacted at email: eko@edu.unisbank.ac.id.

Ali Maskur was born in Tulungagung, East Java, July 18, 1964. He completed Bachelor in Economic Science degrees at Jendral Sudirman University 1991 and Master in Computer Science degrees at Gadjah Mada University 'UGM' Yogyakarta 2004. He has been an active lecturer at Stikubank University since 1992, teaching in digital marketing, finance and banking management, and business management. His research focuses on digital marketing, finance, and banking. He can be contacted at email: maskur@edu.unisbank.ac.id.