Indonesian Journal of Data and Science

Volume 5 Issue 3
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v5i3.200

Research Article

Performance Comparison of CNN and ResNet50 for Skin Cancer Classification Using U-Net Segmented Images

Aris Wahyu Murdiyanto [1,*], Dian Hafidh Zulfikar [2], Bagus Satrio Waluyo Poetro [3], Alda Cendekia Siregar [4]

[1] Universitas Jenderal Achmad Yani Yogyakarta, Yogyakarta, Indonesia, ariswahyumurdiyanto@gmail.com
[2] UIN Raden Intan Lampung, Lampung, dianhafidhzulfikar_uin@radenintan.ac.id
[3] Universitas Islam Sultan Agung, Indonesia, bagusswp@unissula.ac.id
[4] Universitas Muhammadiyah Pontianak, Indonesia, alda.siregar@unmuhpnk.ac.id

Correspondence should be addressed to Aris Wahyu Murdiyanto; ariswahyumurdiyanto@gmail.com

Received 11 September 2024; Accepted 22 December 2024; Published 31 December 2024

Abstract:

Skin cancer is a significant global health issue, with melanoma, basal cell carcinoma, and actinic keratosis being the most common types. Early and accurate detection is critical to improve survival rates and treatment outcomes. This study evaluates the performance of Convolutional Neural Networks (CNN) and ResNet50 in classifying segmented images of skin lesions. The dataset, sourced from Kaggle, was pre-processed using U-Net for lesion segmentation to enhance the quality of input data. Both models were trained and evaluated using accuracy, precision, recall, and F1-score metrics. The CNN model demonstrated a balanced performance across classes, with a weighted F1-score of 47%, but suffered from overfitting, as indicated by the divergence between training and validation losses. ResNet50 achieved better recall for basal cell carcinoma (100%) but failed to classify actinic keratosis and melanoma, resulting in a macro F1-score of 23%. The findings reveal that U-Net segmentation improved classification focus but was insufficient to address dataset imbalance and model-specific limitations. This study highlights the challenges of skin cancer classification using deep learning and underscores the importance of addressing data imbalance and overfitting. Future research should explore advanced techniques, such as ensemble methods, data augmentation, and transfer learning, to improve the generalization and clinical applicability of these models. The proposed framework serves as a foundation for further investigation into automated skin cancer detection systems.

Keywords: Skin Cancer Detection, Convolutional Neural Network (CNN), ResNet50, U-Net Segmentation, Deep Learning.
Dataset link: https://www.kaggle.com/datasets/nodoubttome/skin-cancer9-classesisic

1. Introduction

Skin cancer is a significant global health challenge, with its incidence continually rising due to factors such as increased ultraviolet radiation exposure and changes in lifestyle. It is primarily categorized into melanoma, basal cell carcinoma, and squamous cell carcinoma, with melanoma being the most aggressive and responsible for the majority of skin cancer-related deaths. Early and accurate detection is essential, as delayed diagnosis often results in metastasis, leading to a poor prognosis. While traditional diagnostic methods like biopsies remain reliable, they are invasive, time-consuming, and expensive. Consequently, there is a growing demand for automated, non-invasive diagnostic methods to assist dermatologists in identifying skin cancer efficiently and accurately.

Despite advancements in deep learning for medical imaging, achieving high accuracy in skin cancer classification remains challenging. Variations in lesion appearance, limited dataset sizes, and the complexity of feature extraction are critical obstacles. Previous studies have demonstrated the potential of convolutional neural networks (CNNs) and transfer learning models such as ResNet50 for skin cancer classification [1]–[3]. However, there is a lack of comparative analysis of these architectures when applied to images processed using segmentation techniques. This research aims to address this gap by employing U-Net for precise lesion segmentation and comparing the classification performance of CNN and ResNet50, offering insights into their strengths and limitations for skin cancer detection [4]–[6].

The objectives of this study are threefold: first, to evaluate the role of U-Net in enhancing input data quality for skin cancer classification; second, to compare the performance of CNN and ResNet50 in classifying actinic keratosis, basal cell carcinoma, and melanoma; and third, to identify the strengths and limitations of these architectures to inform future developments in automated skin cancer diagnostics. This research builds upon existing literature, such as the work of Reddy et al., which highlights the importance of pre-processing in skin disease classification using CNN-powered segmentation. Other studies, including those by [7]. And [8], demonstrate the efficacy of ensemble classifiers and multi-class CNNs, respectively, in addressing challenges like misclassification and limited data. Meanwhile, [9]. combined Xception and ResNet50 architectures to achieve high accuracy through feature concatenation, and [10]. highlighted the impact of optimizer selection on ResNet50 and MobileNetV2 performance. Additionally, [11] demonstrated the potential of feature fusion techniques with ResNet50 for melanoma detection.

These studies underscore the importance of pre-processing and advanced architectures in improving skin cancer classification. However, none have specifically examined the combined impact of U-Net segmentation and comparative performance analysis of CNN and ResNet50. This study addresses this gap, providing a comprehensive evaluation of these methods and contributing a novel framework that integrates segmentation and deep learning for improved skin cancer detection.

2. Method:

This section describes the methodology employed in this study, including the dataset, pre-processing techniques, model architecture, training process, and evaluation metrics. The research stages are shown in **Figure 1**.

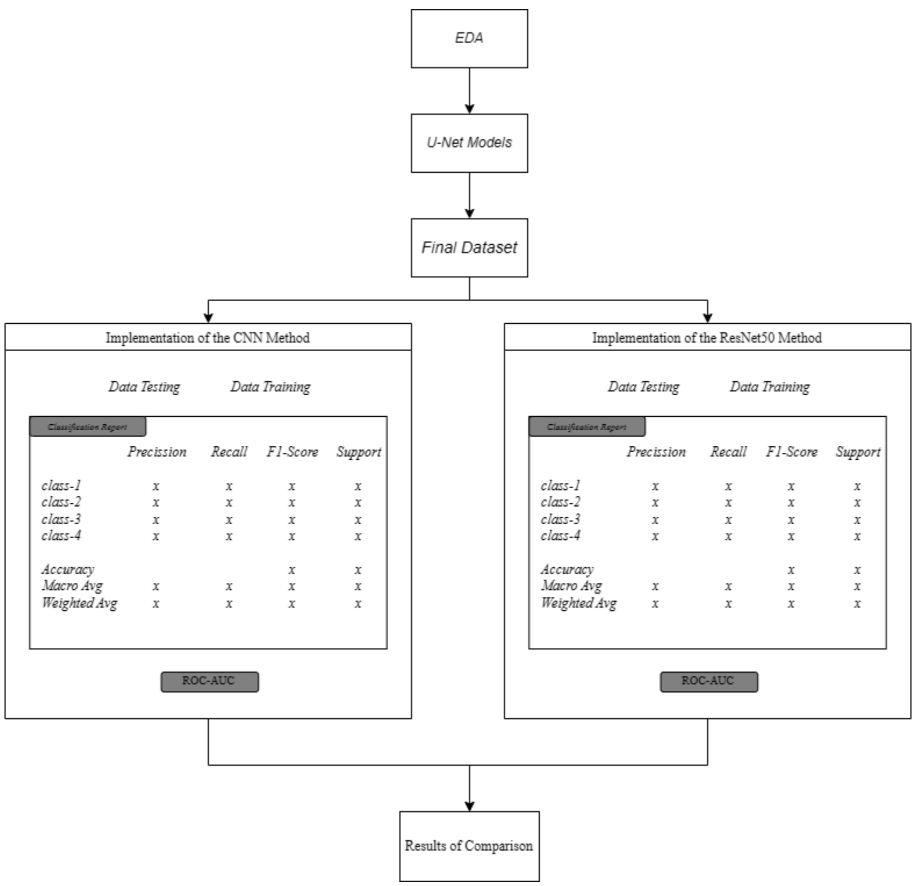

Figure 1. General Research Design Stages

Data Collection Process

The dataset used in this research was sourced from Kaggle and comprises images categorized into three classes: actinic keratosis, basal cell carcinoma, and melanoma, sample of dataset show in **Figure 2**. This dataset was chosen due to its diversity and clinical relevance, providing a robust foundation for training and evaluating deep learning models. The images were pre-processed and divided into training, validation, and test sets to ensure a comprehensive evaluation of the models

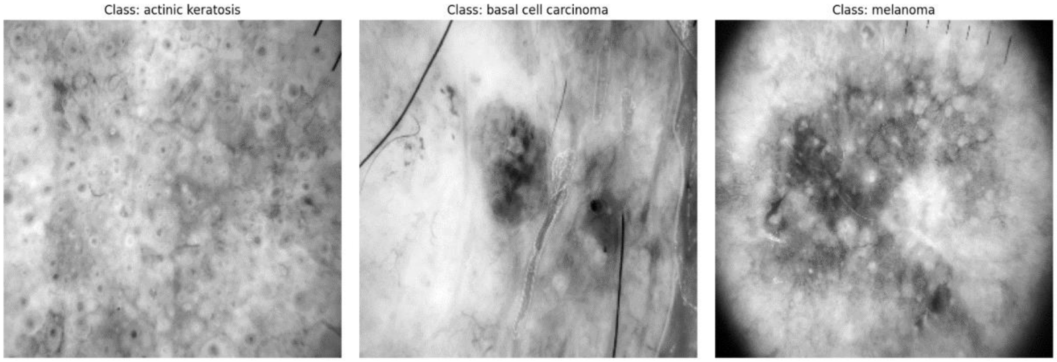

Figure 2. Sample of Dataset

Data Pre-processing

To enhance the classification accuracy, U-Net was utilized for segmentation, isolating the lesion regions and removing irrelevant background information [12], [13]. U-Net operates as an encoder-decoder network where:

1. **Encoder** extracts features from the input image using convolutional layers.
2. **Decoder** reconstructs the segmented image using transposed convolutions.

Mathematically, the U-Net loss function for segmentation can be defined as the combination of cross-entropy loss and Dice coefficient to balance pixel-wise accuracy and overlap similarity, segmentation result sample show in **Figure 3**

$$\mathcal{L}_{U-Net} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i) + \frac{2\sum_{i=1}^{N} y_i\hat{y}_i}{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} \hat{y}_i^2}, \tag{1}$$

Where:

y_i is the ground truth mask

$\hat{y}_i$ is the predicted mask

N is the total number of pixels

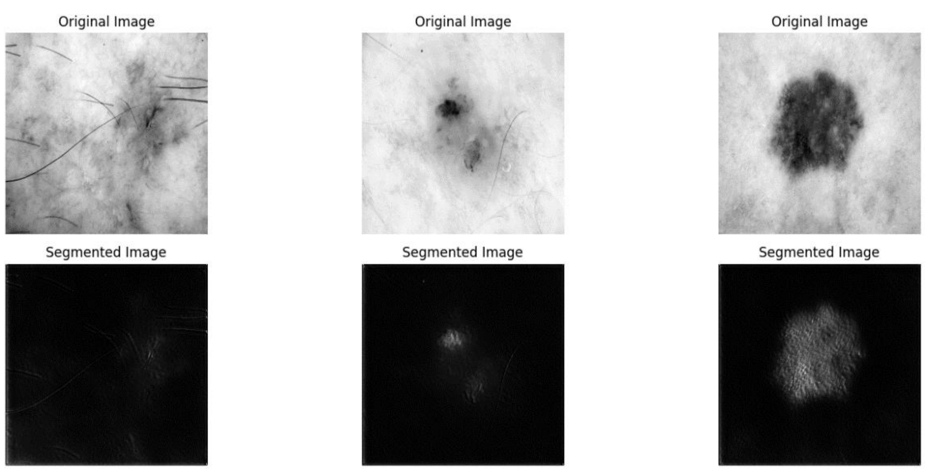

Figure 3. Sample of Dataset after Pre-processing

Model Architectures

Two deep learning architectures were used for classification:

1. **Convolutional Neural Network (CNN):** A custom CNN architecture was implemented with:
   a. Convolutional layers for feature extraction.
   b. Max-pooling layers for dimensionality reduction.
   c. Fully connected layers for classification.

   The convolution operation is defined as [4], [14], [15]:

$$z_{i,j=}\sum_{m=-k}^{k}\sum_{n=-k}^{k}w_{m,n}x_{i}+m,j+n+b \quad (1)$$

   Where x is the input, w represents the filter weights, b is the bias, and k defines the kernel size.

2. **ResNet50:** ResNet50, a 50-layer deep residual network, was employed to leverage the power of transfer learning [1], [16]–[18]. It introduces shortcut connections to prevent the vanishing gradient problem, where the output of layer l is defined as:

$$y\iota = F(x_\iota, \{W_\iota\}) + x_\iota, \quad (1)$$

   Where F represents the residual mapping, x_ι is the input to layer ι, and $\{W_\iota\}$ denotes the weights of the layer.

Training Process

Both models were trained on segmented data with:

a. Loss Function: Categorical Cross-Entropy

$$\mathcal{L}_{CCE} = -\sum_{C=1}^{C} y_c \log(\widehat{y_c}), \quad (1)$$

   Where C is the number classes, y_c is the ground truth probability for class c, and $\widehat{y_c}$ is the predicted probability.
b. Optimizer: Adam optimizer with learning rate $\eta = 0.001$.
c. Batch Size: 32 images per batch
d. Epoch: 50, with early stopping to prevent overfitting.

Evaluation Metrics

The performance of the models was evaluated using the following metrics [19]–[21]:

1. **Accuracy**:

$$Recall = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

2. **Recall**:

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

3. **F1-Score:**

$$F - measure = \frac{2(presisi \times recall)}{(presisi + recall)} \quad (4)$$

4. **Confusion Matrix**: A tabular representation of model predictions to identify class-wise performance.
   a. **Precision:** The ratio of true positive predictions to the total positive prediction:

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

3. Results and Discussion

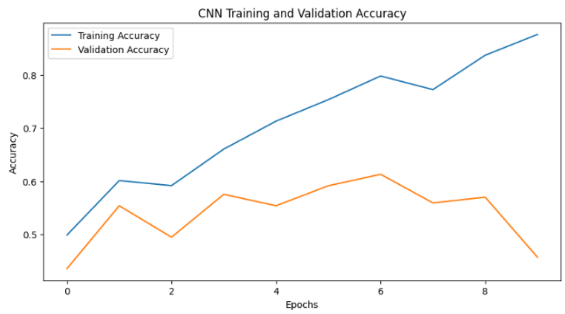

Figure 4: CNN Training and Validation Accuracy

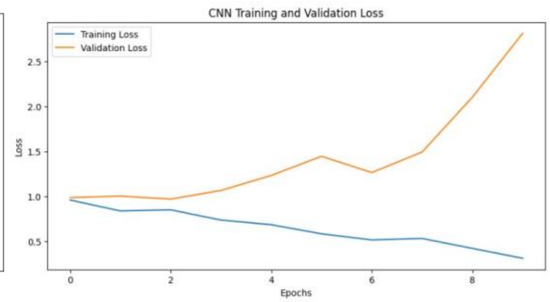

Figure 5: CNN Training and Validation Loss

The performance of CNN and ResNet50 was evaluated through training and validation accuracy, confusion matrices, and classification reports, as shown in the accompanying figures. For the CNN model, the training accuracy increased steadily across epochs, reaching approximately 88% by the 10th epoch, while validation accuracy stagnated around 46% (**Figure 4**). This indicates potential overfitting, as confirmed by the sharp divergence between training and validation loss (**Figure 5**). The confusion matrix for CNN (**Figure 7**) shows that the model performed relatively well in detecting melanoma, achieving a precision of 73% and recall of 57% (**Table 1**). However, it struggled to classify actinic keratosis and basal cell carcinoma correctly, resulting in a weighted F1-score of 47%. The rise in validation loss after the 6th epoch further emphasizes the model's poor generalization to unseen data.

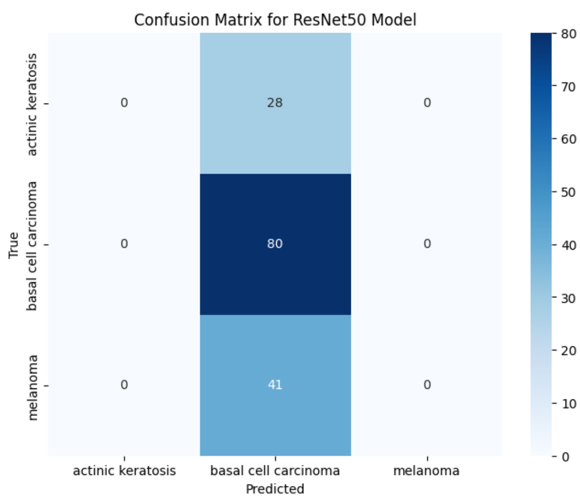

Figure 6: Confusion Matrix for ResNet50

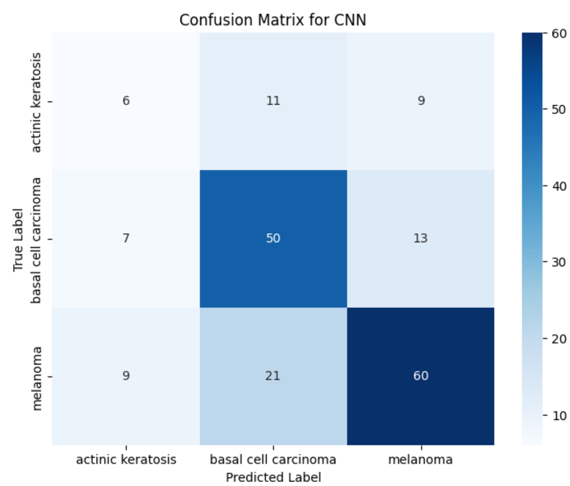

Figure 7: Confusion Matrix for CNN

Figure 8: ResNet50 Training and Validation Accuracy

In contrast, ResNet50 achieved slightly better training accuracy but failed to improve validation accuracy beyond 54%, as illustrated in **Figure 8**. The confusion matrix for ResNet50 (**Figure 6**) reveals its inability to classify actinic keratosis and melanoma, while it achieved perfect recall for basal cell carcinoma. This skewed performance across classes resulted in a macro F1-score of 23% (Table 2). The significant disparity in precision and recall between classes suggests that ResNet50 was heavily affected by the dataset's imbalance, focusing predominantly on the dominant class. Additionally, the consistent gap between training and validation loss (**Figure 8**) highlights ResNet50's overfitting to the training data.

The segmentation provided by U-Net contributed to isolating lesion regions, which improved CNN's ability to detect melanoma. However, it was insufficient to resolve the issues stemming from class imbalance, particularly for ResNet50. Both models displayed overfitting tendencies, which could be mitigated in future studies through data augmentation, dropout layers, or regularization techniques. Compared to previous studies, where ResNet50 achieved better results with ensemble methods and optimizer tuning, the current implementation could benefit from such enhancements to improve generalization. Furthermore, balancing the dataset with techniques like SMOTE or using class-weighted loss functions may improve the classification of underrepresented classes. These limitations must be addressed to make these models more suitable for clinical applications, where balanced performance across all classes is critical.

Table 1. CNN Classification Report

Classification Report:

	precision	recall	f1-score	support
0	0.22	0.65	0.33	26
1	0.42	0.24	0.31	70
2	0.73	0.57	0.64	90
accuracy			0.46	186
macro avg	0.46	0.49	0.43	186
weighted avg	0.54	0.46	0.47	186

Table 2. ResNet50 Classification Report

Classification Report:

	precision	recall	f1-score	support
actinic keratosis	0.00	0.00	0.00	28
basal cell carcinoma	0.54	1.00	0.70	80
melanoma	0.00	0.00	0.00	41
accuracy			0.54	149
macro avg	0.18	0.33	0.23	149
weighted avg	0.29	0.54	0.38	149

This analysis underscores the challenges of deep learning in medical image classification, particularly with imbalanced datasets. While CNN displayed more balanced performance, its overfitting issues limit its applicability. On the other hand, ResNet50's strong focus on one class at the expense of others highlights the need for further refinements, including the use of ensemble approaches or transfer learning with advanced tuning methods.

4. Conclusion

This study investigated the effectiveness of CNN and ResNet50 models in classifying skin cancer into three categories—actinic keratosis, basal cell carcinoma, and melanoma—using segmented images processed with U-Net. The results highlight the challenges of achieving balanced classification performance in the presence of imbalanced datasets and model-specific limitations. The CNN model demonstrated a more balanced performance across classes compared to ResNet50, achieving higher precision and recall for melanoma. However, its validation accuracy stagnated at 46%, and the rising validation loss indicated significant overfitting. On the other hand, ResNet50 achieved slightly better overall accuracy (54%) but struggled with imbalanced data, performing well only on the dominant class (basal cell carcinoma) while failing to classify actinic keratosis and melanoma. This imbalance resulted in a low macro F1-score of 23%, underscoring its inability to generalize effectively.

The segmentation process using U-Net improved the models' focus on lesion regions, providing clearer input for classification. However, the segmentation alone was insufficient to address the broader challenges posed by the dataset's imbalance and the models' overfitting tendencies. These findings emphasize the need for advanced techniques such as data augmentation, class-balancing strategies (e.g., SMOTE or class-weighted loss functions), and ensemble methods to improve model performance and generalization. In conclusion, while CNN demonstrated potential for more balanced classification, its overfitting issues and limited accuracy require further optimization. ResNet50, despite its strong performance on one class, requires enhancements to handle imbalanced data effectively. Future research should focus on addressing these limitations by integrating advanced deep learning techniques, optimizing hyperparameters, and using larger, more balanced datasets. By doing so, these models can become viable tools for clinical applications, offering reliable and automated solutions for skin cancer detection and diagnosis.

References:

[1] S. Prasher, "Resnet 50 Based Classification Model for Skin Cancer Detection Using Dermatoscopic Images," *2023 3rd Asian Conference on Innovation in Technology, ASIANCON 2023*. 2023, doi: 10.1109/ASIANCON58793.2023.10270158.

[2] Y. Filali, "Analysis and Classification of Skin Cancer Based on Deep Learning Approach," *2022 International Conference on Intelligent Systems and Computer Vision, ISCV 2022*. 2022, doi: 10.1109/ISCV54655.2022.9806087.

[3] A. Vatsa, "Comparing the Performance of Classification Algorithms for Melanoma Skin Cancer," *13th IEEE Integrated STEM Education Conference, ISEC 2023*. pp. 375–380, 2023, doi: 10.1109/ISEC57711.2023.10402205.

[4] T. S. Kumar, "Classification of Skin Cancer Images using Lightweight Convolutional Neural Network," *2023 4th International Conference for Emerging Technology, INCET 2023*. 2023, doi: 10.1109/INCET57972.2023.10170637.

[5] S. G. Ahmed, "Skin Cancer Classification Utilizing a Hybrid Model of Machine Learning Models Trained on Dermoscopic Images," *2023 3rd International Conference on Emerging Smart Technologies and Applications, eSmarTA 2023*. 2023, doi: 10.1109/eSmarTA59349.2023.10293619.

[6] N. V. Y. Lankadasu, "Skin Cancer Classification Using a Convolutional Neural Network: An Exploration into Deep Learning," *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems, ICETSIS 2024*. pp. 1047–1052, 2024, doi: 10.1109/ICETSIS61505.2024.10459368.

[7] S. A. Alshawi, "Skin Cancer Image Detection and Classification by CNN based Ensemble Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 5, pp. 710–717, 2023, doi: 10.14569/IJACSA.2023.0140575.

[8] I. Lihacova, "Multi-Class CNN for Classification of Multispectral and Autofluorescence Skin Lesion Clinical Images," *J. Clin. Med.*, vol. 11, no. 10, 2022, doi: 10.3390/jcm11102833.

[9] A. Panthakkan, "Concatenated Xception-ResNet50 — A novel hybrid approach for accurate skin cancer prediction," *Comput. Biol. Med.*, vol. 150, 2022, doi: 10.1016/j.compbiomed.2022.106170.

[10] E. Firasari, "Performance Evaluation of ResNet50 and MobileNetV2 in Skin Cancer Image Classification with Various Optimizers," *Proceeding - 2024 International Conference on Information Technology Research and Innovation, ICITRI 2024*. pp. 376–380, 2024, doi: 10.1109/ICITRI62858.2024.10698943.

[11] S. D. Thepade, "Enhancing Melanoma Skin Cancer Detection Through Feature Fusion of Pre-Trained Deep Convolutional Neural Network ResNet50 and Thepade Sorted Block Truncation Coding," *SN Comput. Sci.*, vol. 5, no. 4, 2024, doi: 10.1007/s42979-024-02821-5.

[12] M. N. Hasan, "Fetal Brain Planes Classification Using Deep Ensemble Transfer Learning from U-Net Segmented Fetal Neurosonography Images," *Int. J. Image, Graph. Signal Process.*, vol. 16, no. 4, pp. 74–86, 2024, doi: 10.5815/ijigsp.2024.04.06.

[13] T. T. Fousiya, "Diabetic Retinopathy Classification Based on Segmented Retinal Vasculature of Fundus Images Using Attention U-NET," *INDICON 2022 - 2022 IEEE 19th India Council International Conference*. 2022, doi: 10.1109/INDICON56171.2022.10039734.

[14] M. A. Arshed, “Multi-Class Skin Cancer Classification Using Vision Transformer Networks and Convolutional Neural Network-Based Pre-Trained Models,” *Inf.*, vol. 14, no. 7, 2023, doi: 10.3390/info14070415.

[15] S. Deshmukh, “Skin Cancer Classification Using CNN,” *International Conference on Applied Intelligence and Sustainable Computing, ICAISC 2023*. 2023, doi: 10.1109/ICAISC58445.2023.10199440.

[16] R. Bharti, “Comparative Analysis of Potato Leaf Disease Classification Using CNN and ResNet50,” *2024 International Conference on Data Science and Its Applications, ICoDSA 2024*. pp. 87–91, 2024, doi: 10.1109/ICoDSA62899.2024.10651649.

[17] G. B. Alghanimi, “CNN and ResNet50 Model Design for Improved Ultrasound Thyroid Nodules Detection,” *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems, ICETSIS 2024*. pp. 1000–1004, 2024, doi: 10.1109/ICETSIS61505.2024.10459588.

[18] M. Singla, “ResNet50 Utilization for Bag Classification: A CNN Model Visualization Approach in Deep Learning,” *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems, ICITEICS 2024*. 2024, doi: 10.1109/ICITEICS61368.2024.10624847.

[19] A. P. Wibowo, M. Taruk, T. E. Tarigan, and ..., “Improving Mental Health Diagnostics through Advanced Algorithmic Models: A Case Study of Bipolar and Depressive Disorders,” *Indones. J. …*, 2024, doi: 10.56705/ijodas.v5i1.122.

[20] R. Setiawan and H. Oumarou, “Classification of Rice Grain Varieties Using Ensemble Learning and Image Analysis Techniques,” *Indones. J. Data …*, 2024, doi: 10.56705/ijodas.v5i1.129.

[21] B. S. W. Poetro, E. Maria, H. Zein, and ..., “Advancements in Agricultural Automation: SVM Classifier with Hu Moments for Vegetable Identification,” *Indones. J. …*, 2024, doi: 10.56705/ijodas.v5i1.123.