Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 13 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i13.57417

PAPER

The L2-EffCANet: A Novel Overfitting-Resistant EfficientNetV2S with Attention Mechanism and L2 Regularization for Skin Disease Classification

Kurnia Adi Cahyanto (✉), **Kusworo Adi, Catur Edi Widodo**

Universitas Diponegoro, Semarang, Central Java, Indonesia

kurnia@polindra.ac.id

ABSTRACT

Skin diseases are a common health problem that is often underestimated. However, some types of skin diseases can become cancerous and fatal if not treated properly, such as melanoma. Melanoma is caused by excessive exposure to ultraviolet rays and has a 99% cure rate if diagnosed early, but this figure drops to 20% in advanced stages. In developing countries, uneven distribution of medical personnel and geographical challenges lead to many skin diseases going undiagnosed. This study develops a multi-class model for classifying skin diseases using transfer learning, leveraging pre-trained models like EfficientNetV2S to address overfitting and improve accuracy. The model is trained on a dataset from DermNet consisting of 23 classes and a total of 19,559 skin disease images. Data augmentation is used to reduce class imbalance. The EfficientNetV2S model with the addition of CA and L2 regularization at the end of the model architecture achieves a test accuracy of 71.78% and demonstrates stable superiority, surpassing previous research. The study shows that deep learning can help with the early detection of skin diseases, thus improving healthcare services.

KEYWORDS

skin disease, diagnosis, L2 regularization, EfficientNetV2S, accuracy, health

1 INTRODUCTION

Skin diseases are common in the community due to health factors that are considered relatively harmless. However, some types of skin diseases can be cancerous and can become deadly if not treated properly [1]. Some skin disease problems are caused by environmental influences, while others may be genetically inherited. Some skin conditions are mild, while others can be fatal and cancerous [2]. Melanoma, which is often caused by overexposure to ultraviolet (UV) light, has a 5-year survival rate of up to 99% if diagnosed early, but this rate drops to around

Cahyanto, K. A., Adi, K., Widodo, C. E. (2025). The L2-EffCANet: A Novel Overfitting-Resistant EfficientNetV2S with Attention Mechanism and L2 Regularization for Skin Disease Classification. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(13), pp. 113–129. https://doi.org/10.3991/ijoe.v21i13.57417

Article submitted 2025-07-03. Revision uploaded 2025-09-08. Final acceptance 2025-09-08.

20% in advanced stages. In developing countries, access to dermatology care is often limited due to uneven distribution of medical personnel and geographical barriers to care [3]. As a result, many severe skin conditions go undiagnosed or are detected late, reducing the chances of a cure [4].

Given the above facts, improving the survival rate requires early diagnosis of the disease [5]. Medical capabilities sometimes limit the early detection of skin diseases, so an alternative is needed. Therefore, machine learning using computers has become a necessity that crosses all fields [6], including those related to skin health. Among such technologies is deep learning, which uses convolutional neural networks (CNNs) to classify different forms of skin diseases based on visual images of affected areas [7], with the aim of identifying the type of skin disease in a patient as early as possible [8]. Transfer learning utilizes pre-trained models to handle image data. However, when handling multi-class skin diseases that are often unbalanced [9], the training accuracy results can differ significantly from the accuracy observed when applying the model to test data. This leads to overfitting, where the model only remembers the training data with limited generalization ability [10]. Therefore, this study aims to overcome data imbalance and improve the accuracy of previous studies in classifying multi-class skin diseases.

The main contribution of this research is the creation and assessment of a novel deep learning model named L2-EffCANet. This model is a changed version of the EfficientNetV2S design, and it cleverly added Coordinate Attention (CA) and applies L2 regularization techniques to enhance the model's ability to accurately classify different types of human skin diseases. L2-EffCANet is designed to solve these issues by adding a penalty to the model's weights during training, unlike older methods that often struggle with overfitting due to uneven data and the complexity of visual features in different skin disease categories.

The rest of this paper is organized as follows:

Section 2 reviews related works on multi-class skin disease classification using deep learning. Section 3 describes the methodology, including dataset preparation, augmentation strategies, the proposed L2-EffCANet architecture, and training configuration. Section 4 presents the experimental setup, results, and comparative analysis with previous studies. Section 5 discusses the findings, limitations, and potential improvements. Finally, Section 6 concludes the paper and outlines directions for future research.

2 RELATED WORKS

Although there have been many studies examining the use of machines in skin disease classification, there are still limited studies that specifically discuss classification with a multi-class approach. Therefore, this study only reviews related studies that involve a relatively large number of classes.

In 2021, in their review paper, Li et al. [5] present a thorough examination of deep learning methodologies for the diagnosis of skin diseases, focusing specifically on multi-class categorization encompassing as many as 23 categories of skin conditions. They emphasize the efficacy of CNNs and transfer learning methodologies in managing extensive and intricate datasets like DermNet. Despite these improvements, problems still exist, especially with uneven representation of different skin conditions and the similar appearances of many skin disorders, which negatively impact how accurately they can be classified. The review emphasizes the essential need for data preparation and augmentation techniques to improve the efficacy and resilience of models for classifying multiple skin diseases.

Aboulmira et al. [11] conducted a detailed study comparing how well different CNN models worked in identifying 23 types of skin conditions [12]. The research looked at how well well-known CNN models such as VGG16, ResNet50, InceptionV3, and DenseNet121 worked by using a carefully selected set of skin images. Their tests showed that using transfer learning was crucial for achieving high accuracy in classification, especially when pre-trained models were improved with skin images. Of the models evaluated, DenseNet121 had superior performance, attaining greater accuracy in multi-class classification relative to other architectures. The classification accuracy achieved in this study is 68.97%, while the precision is 67.3% and the recall is 67.2%.

In the meantime, Cahyanto et al. [13] investigated the effectiveness of L2 regularization in enhancing the performance of EfficientNetV2S for the classification of multi-class cutaneous diseases. This study investigates the overfitting issue that is frequently encountered in deep learning models that are constructed from dermatology datasets that are imbalanced. Superior generalization and enhanced accuracy across 23 skin disease categories were achieved by incorporating L2 regularization into the EfficientNetV2S training regimen. This study underscores the significance of regularization techniques in the retention of learning and the enhancement of deep neural networks for complex medical image categorization tasks. After the integration of L2 regularization into the EfficientNet architecture, the accuracy value for testing increased from 58% to 61.4%. Nevertheless, L2 regularization is exclusively implemented on the kernel side of the regularizer in this context.

Balansundaram et al. [14] performed classification with a genetic algorithm (GA) approach to find the optimal model combination between ResNet50, DenseNet121, and a simple CNN. The dataset used was the DermNet dataset, which had 23 classes. The accuracy obtained by this model is 45.83%, and it still struggles to overcome the increasing level of overfitting.

Table 1. Summary of literature related to the classification of multi-class skin disease

Ref	Authors	Year	Goal Points	Model	Dataset	Result
[11]	Aboulmira et al.	2022	Classification of 23 skin disease	DenseNet201	DermNet	Test accuracy: 68.97%
[13]	Cahyanto et al.	2023	Classification of 23 skin disease	EfficientNetV2S + L2 on kernel side	DermNet	Test accuracy: 61.4%
[14]	Balasundaram et al.	2024	Classification of 23 skin disease	GA for or ensemble method	DermNet	Test accuracy: 45.83%

Table 1 encapsulates the details of each research study pertaining to the detection of skin illnesses and other diseases, together with the datasets employed in the investigations. References are sorted by the oldest year.

3 RESEARCH METHOD

3.1 System workflow framework

Broadly speaking, the steps to conduct this research include data collection, data augmentation, data splitting, modeling, and finally model evaluation, which is then implemented to classify skin disease images. Figure 1 illustrates the flow of this research framework.

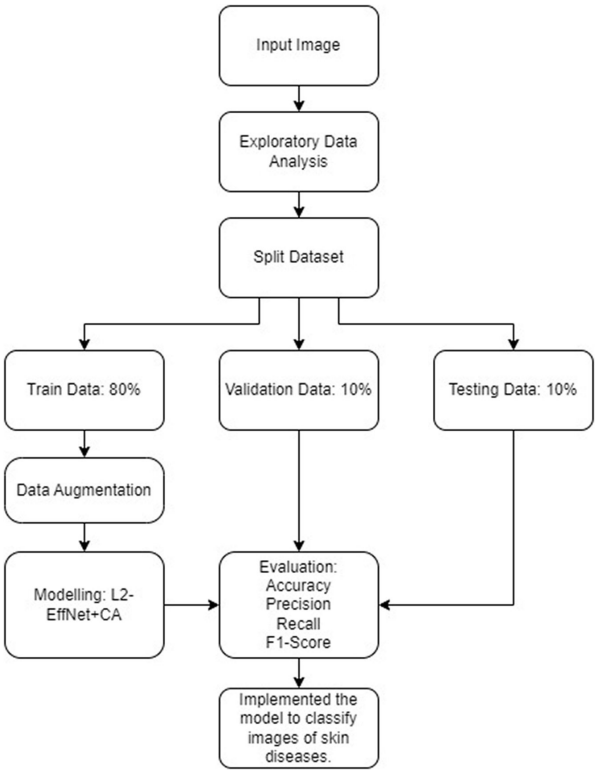

Fig. 1. Framework of the proposed skin disease classification model

3.2 Data collection and data analysis

The dataset used in this study is a public dataset. The DermNet dataset comprises photographs of 23 distinct skin illnesses sourced from https://DermNetnz.org. The overall count of photos is approximately 19,500, with around 15,500 allocated to the training set and the remainder designated for the testing set. The photos are in Joint Photographic Experts Group (JPEG) format and comprise three channels, namely red, green, and blue (RGB). The resolution varies among images and categories. However, these photographs are predominantly of low resolution. Figure 2 illustrates the distribution of images across the 23 classes for skin diseases.

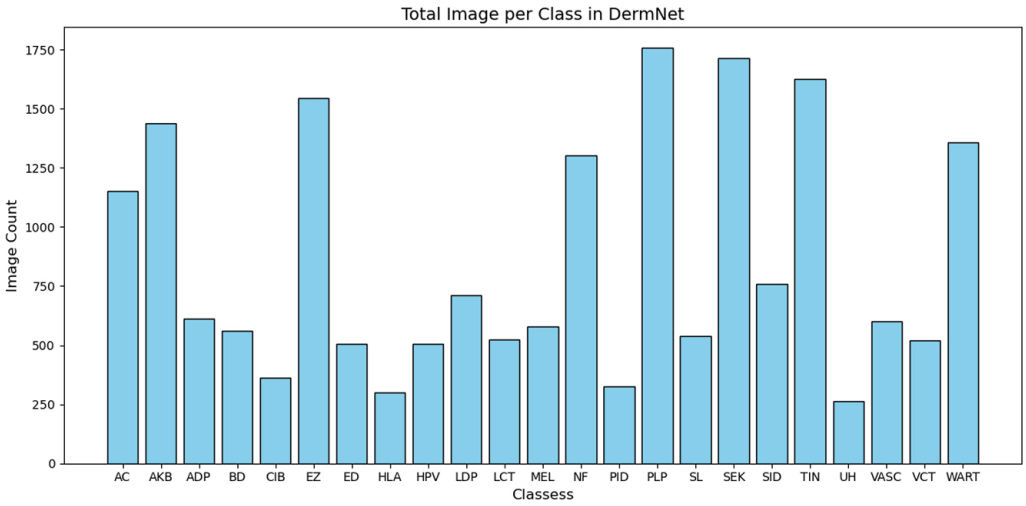

Fig. 2. DermNet training dataset distribution

Table 2 describes the types of skin disease diagnoses and their initials, along with the number of images in the training and testing datasets from the DermNet dataset.

Table 2. DermNet dataset

Label	Diagnostic Category	Count	
		Train	Test
AC	Acne and Rosacea	840	312
AKB	Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions	1149	288
ADP	Atopic Dermatitis	489	123
BD	Bullous Disease	448	113
CIB	Cellulitis Impetigo and other Bacterial Infections	288	73
EZ	Eczema	1235	309
ED	Exanthems and Drug Eruptions	404	101
HLA	Hair Loss Photos Alopecia and other Hair Diseases	239	60
HPV	Herpes HPV and other STDs	405	102
LDP	Light Diseases and Disorders of Pigmentation	568	143
LCT	Lupus and other Connective Tissue Diseases	420	105
MEL	Melanoma Skin Cancer Nevi and Moles	463	116
NF	Nail Fungus and other Nail Disease	1040	261
PID	Poison Ivy and other Contact Dermatitis	260	65
PLP	Psoriasis Pictures Lichen Planus and Related Diseases	1405	352
SL	Scabies Lyme Disease and other Infestations and Bites	431	108
SEK	Seborrheic Keratoses and other Benign Tumors	1371	343
SID	Systemic Disease	606	152
TIN	Tinea Ringworm Candidiasis and other Fungal Infections	1300	325
UH	Urticaria Hives	212	53
VASC	Vascular Tumors	482	121
VCT	Vasculitis	416	105
WART	Warts Molluscum and other Viral Infections	1086	272

Dermatoscopy facilitates the diagnosis of a variety of dermatological disorders and the evaluation of suspicious skin lesions, thereby improving the physician's examination of the skin. This method frequently improves the identification of benign and malignant pigmented skin lesions in comparison to visual examination [15]. To help understand the skin disease image collection, Figure 3 display some example images of each type of skin lesion from the DermNet dataset.

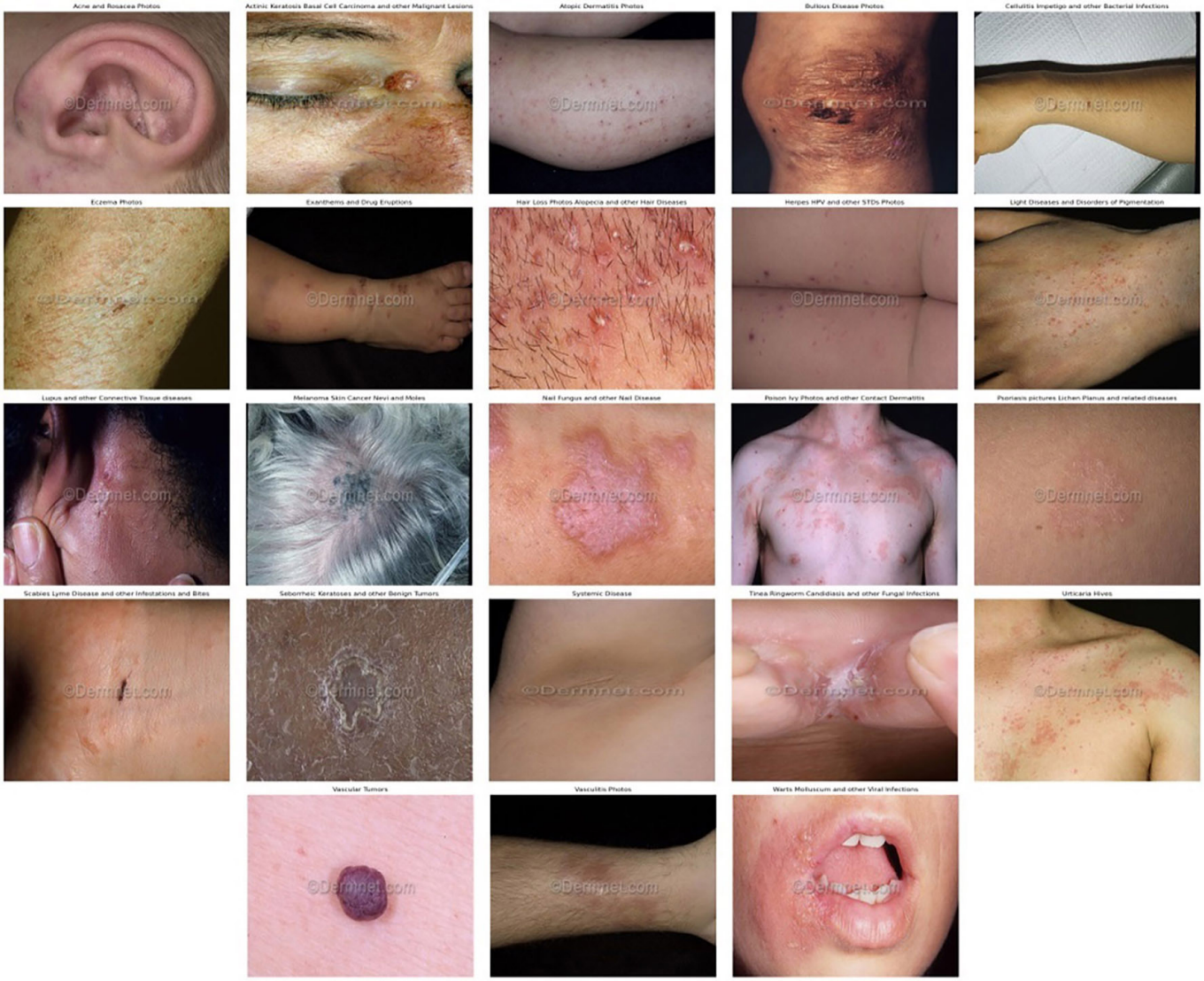

Fig. 3. 23 classes of skin disease images from the DermNet dataset

3.3 Data splitting

In this study, the dataset is split with a composition of about 80% for training data, 10% for validation data, and 10% for test data, because it is relatively frequently used and is considered effective for training and evaluating models that classify skin diseases [16].

In the original source, the DermNet dataset was initially divided into training data and testing data, with an initial composition of 79.54% for training data and 20.46% for testing data. Therefore, in this study, the training data and test data were first combined and then divided again with a composition of 80% for training data, 10% for validation data, and 10% for testing data.

Figures 4 illustrate the distribution of picture quantities across each class within the training, validation, and test datasets. This graphic reveals a class imbalance, characterized by a markedly greater quantity of training data compared to the validation and test datasets. The disparity in figures is significant, as it may influence

the efficacy of the classification model, particularly regarding its applicability to minority classes.

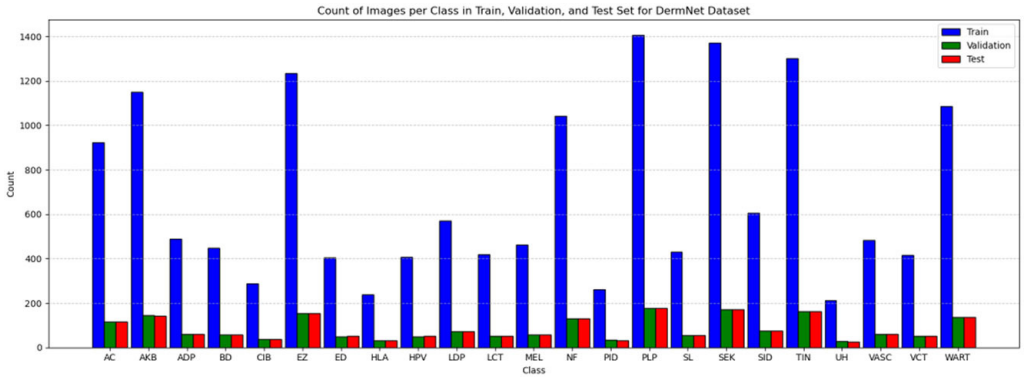

Fig. 4. Count of images per class in train, validation, and test set for DermNet

The DermNet dataset included in this study is characterized by an inherent class imbalance, as the total number of pictures available for each disease category varies in the original source. To guarantee representativeness and preserve the natural distribution of the data, researchers employed a stratified splitting technique within each class, allocating 80% of the images for training, 10% for validation, and the final 10% for testing. This methodology maintains a uniform proportional distribution among classes [17]. Nonetheless, the total number of images in each subset necessarily varies based on the initial class distribution. This technique reflects real-world data situations and facilitates the evaluation of the model's robustness in handling imbalanced datasets.

3.4 Data augmentation

Augmentation techniques are frequently implemented to enhance the generalization and performance capabilities of machine learning models, particularly when the quantity of available data is restricted or unbalanced. Machine learning models are able to learn more robust features and generalize to new data more effectively by applying such transformations to actual skin disease images, resulting in a larger and more diverse dataset [18]. Rotation, shear range, zoom range, shift range, rotation range, horizontal and vertical flip are the geometric augmentation techniques that are implemented. While introducing variability in orientation and spatial scale, these transformations preserve the semantic content of the image, which is essential for robust feature learning in visual classification tasks.

Several geometric augmentation techniques were applied during the training process to increase the diversity of the training dataset and reduce class imbalance. Initially, all images were adjusted to a scale of 1/255. Next, random shift and zoom transformations were applied in the range of 0.2, as well as width and height adjustments of up to 20%. Horizontal and vertical rotations were applied to accommodate orientation variations, along with random rotations of up to 20 degrees. Local structure was preserved without introducing artifacts by dynamically expanding the images using the 'nearest neighbor' filling mode. Overall, these operations were used to enhance the model's robustness and its ability to generalize to various visual patterns associated with skin diseases.

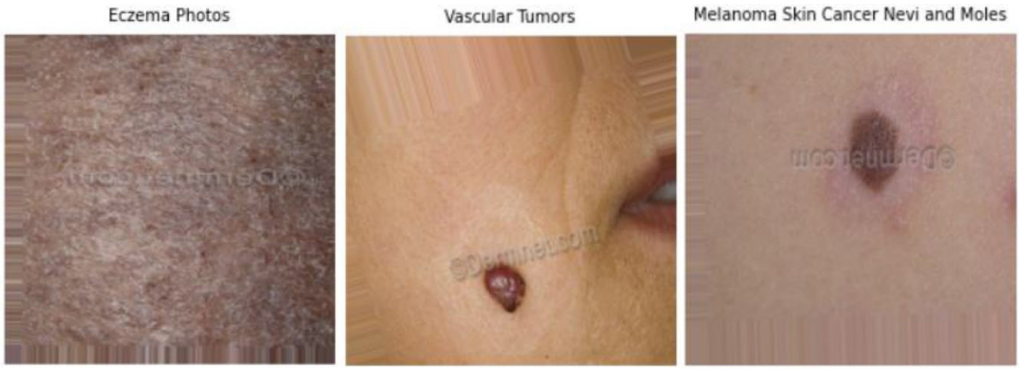

Fig. 5. Some examples from augmentation data result

Considering the relatively large number of classes in the DermNet dataset, batch size 16 and on-the-fly augmentation techniques are used, which dynamically apply transformations to the images during the training process without permanently storing the augmented results in memory, thus saving memory usage [19]. In addition, maintaining the expanded dataset in memory can provide a challenge, especially if the dataset size increases. Applying augmentation to large datasets exacerbates this problem. Some of the augmented images are illustrated in Figure 5.

3.5 EfficientNetV2S

In general, EfficientNet is superior to other pretrained architectures, both in terms of a more compact number of parameters and accuracy, so it can relatively accelerate the computation process with good accuracy [20]. This step completes the data modeling process to develop the classification model. This model is developed by using the EfficientNetV2 architecture, a novel neural network scaling technique. With 11× faster training speed and a 6.8× smaller model size, the EfficientNetV2 model set of convolutional neural networks (CNN) offers better image categorization capabilities. EfficientNetV2 uses a smaller 3×3 filter with several layers, combines fused-MBConv and fused-MBConv at the first layer, and includes a mobile inverted bottleneck convolution (MBConv) block that has a smaller expansion ratio [21].

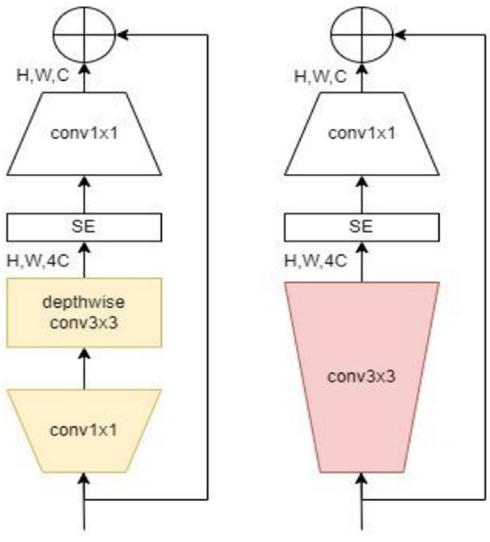

Fig. 6. Description of the structure of MBConv and Fused-MBConv

The Figure 6 depicts the architecture of MBConv and Fused-MBConv. In EfficientNetV2, there are slight differences in the layers used by the Fused-MBConv and Mobile Inverted Bottleneck Convolution (MBConv) blocks. Fused-MBConv uses Conv3×3, while MBConv uses Conv3×3 and Conv1×1 in depth.

3.6 L2-EffNet architecture

EfficientNetV2S serves as the primary base for this multi-class model that classifies skin diseases. Subsequently, the output from EfficientNetV2S is passed to the attention mechanism model, specifically coordinate attention, which can handle both channel and spatial features, enabling the model to better understand and recognize the position of skin disease objects. CA is a relatively lightweight attention mechanism, providing improved focus without significantly increasing the number of model parameters used [22]. The general formulation of CA is as follows:

$$y_c(i, j) = x(i, j) \cdot \sigma(W_h f(i)) \cdot \sigma(W_w f(j)) \quad (1)$$

Where:
$x(i, j)$ is input feature
f is shared transform results from coordinate pooling
W_h, W_w are separation of height and width paths
σ sigmoid function for normalization

L2 regularization is employed to impose a penalty on weights that are excessively large. Subsequently, the model is able to decrease the size of the weights in order to reduce overfitting [23]. This phase is essential for the model to become more generalizable to data that has never been observed before. This parameter is incorporated into the existing loss (L) function to form an updated version, with λ controlling the intensity of regularization. As λ increases, the magnitude of the resulting weights decreases.

A comprehensive L2 regularization approach is applied using a fully connected dense layer with 256 units to improve the model's generalization ability and reduce the risk of overfitting. In this study, we exclusively apply L2 regularization to kernel weights, unlike standard methods that may ignore regularization or limit it to certain layers. This method reduces overfitting and promotes a more balanced representation during training by applying a penalty to significant kernel weights. The implementation and evaluation of L2 regularization on EfficientNetV2S for multi-class skin disease classification, although commonly used in other architectures, has not been thoroughly investigated in previous studies, making it a significant contribution of our study.

Feature extraction:

$$Z_1 = f_{EffNet}(x) \quad (2)$$

Global Average Pooling (GAP):

$$Z_2 = GAP(Z_1) \quad (3)$$

Batch Normalization (BN):

$$Z_3 = BN(Z_2) \quad (4)$$

Dense Layer with ReLU regularization and activation:

$$Z_4 = ReLU(W_1 Z_3 + b_1) \tag{5}$$

with L2 regularization penalty:

$$Loss_{new} = Loss + \lambda_1 \left\| W_1 \right\|_2^2 \tag{6}$$

Dropout

$$Z_5 = Dropout(Z_4, rate = 0.25) \tag{7}$$

Output

$$\hat{y} = Softmax(W_2 Z_5 + b_2) \tag{8}$$

In the proposed EfficientNetV2S construction, after fully connected (FC), batch normalization momentum 0.99, dense layer with L2 0.01, and DropOut 0.25 to prevent overfitting. The learning rate (LR) is set temporarily and will continue to decrease as the accuracy increases. When no improvement in validation accuracy occurs after 10 epochs, the computation will be stopped. The L2-EffNet architecture is illustrated in Figure 7.

Fig. 7. The L2-EffNet architecture

The LR setting starts at 0.001, but it is configured so that if there is no improvement in validation accuracy for 3 epochs, the learning rate value will be multiplied by 20% with a maximum epoch of 100. The optimizer employed is "Adamax" due to the significant variability and imbalance in the data [24]. This characteristic renders it especially appropriate for datasets exhibiting significant heterogeneity and class imbalance, such as DermNet, since it mitigates the danger of unstable parameter updates and improves convergence resilience [25]. The loss function used is categorical cross entropy because the dataset is multi-class [26].

3.7 Evaluation metrics

This experiment will be tested with a confusion matrix commonly used for classification performance testing, which consists of true negative (TN), false positive (FP), false negative (FN), and true positive (TP) [27]. These evaluation metrics are calculated using certain formulas [28].

Accuracy denotes the ratio of correct predictions made by the model, determined by dividing the count of accurate predictions by the total predictions made. The accuracy of a model is determined by dividing the sum of TPs and TNs by the total of true positives, FPs, TNs, and false negatives.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{9}$$

Precision denotes the ratio of correct positive predictions generated by the model in relation to the overall number of positive predictions made.

$$Precision = \frac{(TP)}{(TP + FP)} \tag{10}$$

Recall is the proportion of TP cases to the total number of positive cases.

$$Recall = \frac{(TP)}{(TP + FN)} \tag{11}$$

The F1-score is derived by computing the harmonic mean of the recall and precision scores. The F1-score offers an exact evaluation of recall and precision. The F1-score is computed by multiplying precision and recall, followed by multiplying the result by 2. Then divide this value by the sum of precision and recall.

$$F1 - Score = 2 * \frac{(precision * recall)}{(precision + recall)} \tag{12}$$

All evaluations were performed on the test dataset to ensure unbiased assessment of the model's generalization capabilities.

4 EXPERIMENTAL RESULT AND DISCUSSION

The network training procedure uses equipment with hardware specifications of an 11th Generation Intel(R) Core(TM) i7-1165G7 processor, 40 GB RAM, and GeForce RTX 4060 Ti VGA. The software used is the Windows 10 Education 64-bit operating system, Python 3.9.13, and TensorFlow 2.10.0.

4.1 Result and analysis

This part will present the experimental results together with the interpretation and analysis of the acquired data. The experiments aim to assess the performance of the model across different parameter combinations throughout the training phase. The primary emphasis is on the dynamics of altering training parameters, including epoch and learning rate, and their influence on the ultimate model outcomes. The results from these studies offer a comprehensive analysis of the method's efficiency and effectiveness.

This experiment was conducted with a maximal epoch configuration of 100. The computation was terminated at the 47th epoch after a duration of approximately 7 hours, 47 minutes, and 19.35 seconds. The learning rate decreased throughout the computation period until the 20th epoch, at which point it reached 10^{-6}. Table 3 provides a summary of the accuracy value's change as the learning rate decreases at each epoch session.

Table 3. Accuracy value in each decrease in learning rate

Epoch Session	LR	Train Accuracy (%)	Validation Accuracy (%)
1–7	10^{-3}	66.19	30.62
8–19	$2*10^{-4}$	90.67	69.48
20–22	$4*10^{-5}$	92.77	70.65
23–25	$8*10^{-6}$	93.12	70.76
26–29	$1.6*10^{-6}$	92.99	70.96
30–36	$1*10^{-6}$	93.44	70.65
37	$1*10^{-6}$	**93.42**	**71.17**
38–47	$1*10^{-6}$	93.23	70.60

The end of the training process resulted in a training process accuracy of 93.23%, while the validation process accuracy was 70.60%. The system automatically reduces the learning rate by five times its initial value. The best validation accuracy value achieved during the training stage will be taken as the model to classify the test data. Thus, the best result is a training accuracy of 93.42% and a validation accuracy of 71.17%. Figure 8 illustrates the training and validation accuracy curves of the proposed model throughout 47 epochs, where the best validation model falls at epoch 37.

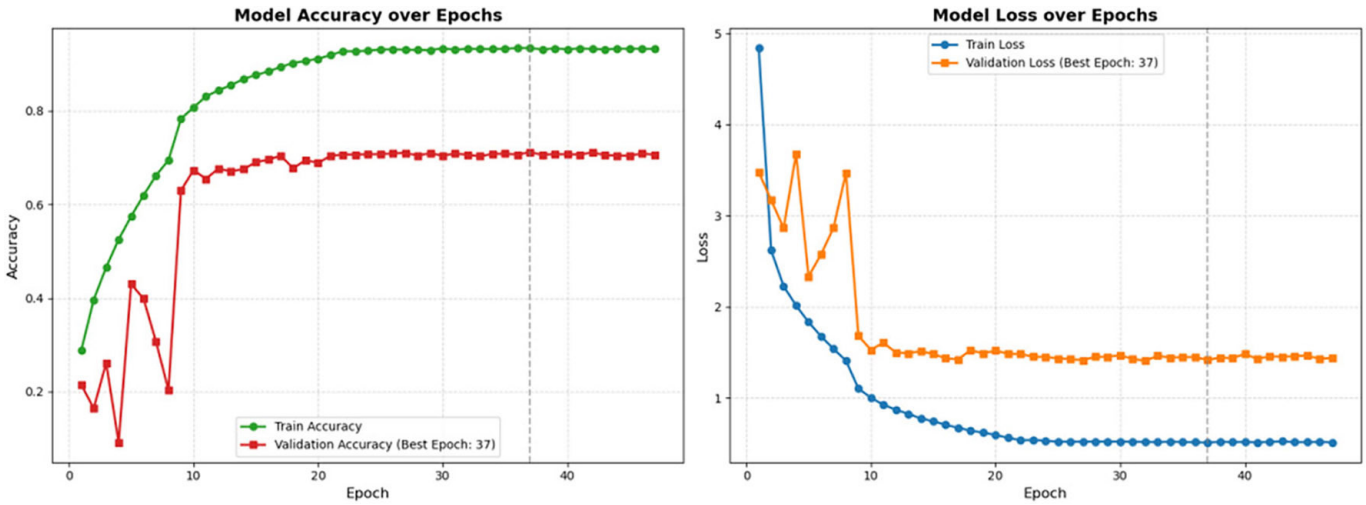

Fig. 8. Proposed deep learning model learning curves: (left) accuracy and (right) loss on training and validation

Figure 9 illustrates the confusion matrix for 23 disease classes, highlighting misclassification patterns and class-level performance, and also showing the accuracy rate for each class [29].

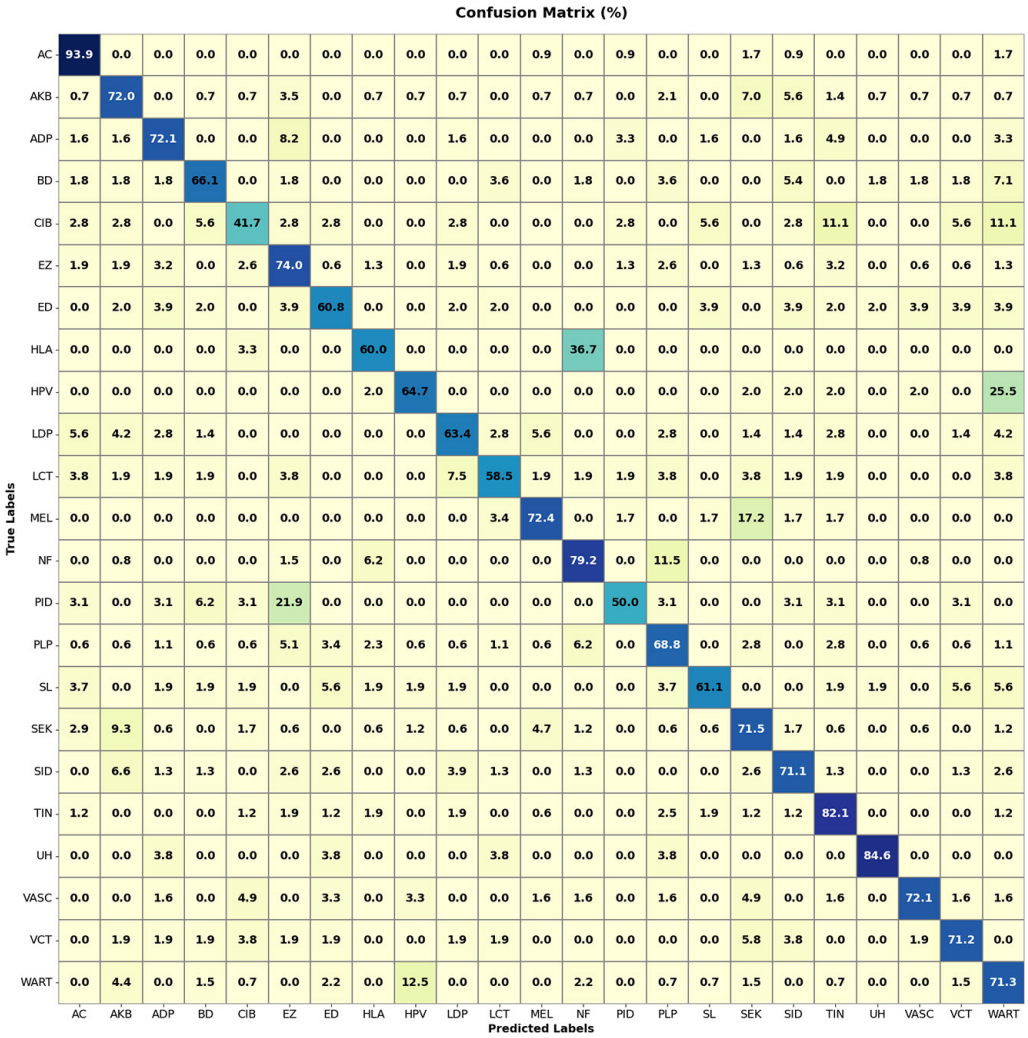

Fig. 9. Confusion matrix for multi-class skin disease classification

4.2 Discussion

After the model was tested using the test dataset, the final result obtained for testing accuracy was 71.78%. Then for other metrics, precision was 71.93%, recall 71.78%, and the F1-score 71.71%. The model showed a consistent upward trend in training accuracy, reaching about 93.44% at the end. This result shows that the model learns and adapts to the training data effectively, with the overfitting rate tending to stabilize. Overfitting is still somewhat evident due to class imbalance, where the majority class dominates and reduces sensitivity to minority categories, and high image diversity and inter-class similarity, as variations in lighting, resolution, skin color, and lesion appearance create feature overlap.

This study contributes significantly to both the medical business and the research community. Previously, several studies contributed to the field of skin disease diagnosis, but the models provided in these studies were often limited, particularly in terms of accuracy. Table 4 presents a comparative analysis of various past research studies alongside the findings of this investigation.

Table 4. Comparative analysis with previous studies

Ref	Title	Year	Dataset	Model	Result
[11]	Aboulmira et al.	2022	DermNet	DenseNet201	Test accuracy: 68.97%
[13]	Cahyanto et al.	2023	DermNet	EfficientNetV2S	Test accuracy: 61.4%
[14]	Balasundaram et al.	2024	DermNet	GA for or ensemble method	Test accuracy: 45.83%
Baseline EfficientNetV2S			DermNet	EfficientNetV2S	Test accuracy: 69.81%
L2-EffNet			DermNet	L2 regularization is included for EfficientNetV2S	Test accuracy: 70.36%
L2-EffCANet			DermNet	CA and L2 regularization is included for EfficientNetV2S	Test accuracy: 71.78%

4.3 Implemented the model to classify images of skin disease

This research must be conducted to prove that this method is effective, as indicated by the evaluation results. In the context of dermatological image interpretation, visual representations not only demonstrate the predictive performance of the model but also facilitate a qualitative evaluation of its diagnostic accuracy. Figure 10 displays the results of the image classification in this case. Both images contain explanations of the skin disease names and whether the predictions are correct or incorrect. The image on the left shows the correct prediction results, while the image on the right shows the incorrect prediction results.

Fig. 10. Skin disease image classification results

5 CONCLUSION AND FUTURE WORKS

This study introduces L2-EffCANet, an improved deep learning model that enhances EfficientNetV2S with an attention mechanism and utilizes L2 regularization to address the challenges of classifying various types of skin diseases. This new model demonstrates that it can better understand and correctly classify skin diseases, even when some categories have fewer examples and when previous models struggled with overfitting, using a robust dataset from DermNet that includes 23 different types of skin conditions. The model's ability to learn important features is enhanced while remaining computationally efficient, thanks to the use of L2 regularization, which helps control large weight values. Additionally, the addition of CA

features can help the base model better recognize the position of skin disease objects without significantly increasing the number of parameters. Extensive experiments indicate that L2-EffCANet outperforms standard architectures and state-of-the-art approaches in terms of test accuracy under the same conditions. This performance indicates that the technique could be a useful tool for quickly and accurately identifying skin diseases in clinics, especially when there are not enough specialists available.

Despite the training accuracy achieved already being relatively high, this study still prioritizes reducing overfitting by developing models and formulations related to this issue. Therefore, future research will focus on the main subject of skin disease images by applying attention mechanisms, loss control, and explainable artificial intelligence (XAI) transparency, as well as model confidence in real-world healthcare settings, such as mobile health applications and point-of-care systems.

6 ACKNOWLEDGEMENT

Appreciation and thanks to the Indonesian Education Scholarship, the Center for Higher Education Funding and Assessment, the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, the Endowment Fund for Education Agency and the Ministry of Finance of the Republic of Indonesia.

7 REFERENCES

[1] M. A. Khan, K. Muhammad, M. Sharif, T. Akram, and V. H. C. D. Albuquerque, "Multi-class skin lesion detection and classification via teledermatology," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 12, pp. 4267–4275, 2021. https://doi.org/10.1109/JBHI.2021.3067789

[2] A. Rafay and W. Hussain, "EfficientSkinDis: An EfficientNet-based classification model for a large manually curated dataset of 31 skin diseases," *Biomed. Signal Process. Control*, vol. 85, p. 104869, 2023. https://doi.org/10.1016/j.bspc.2023.104869

[3] I. K. E. Purnama *et al.*, "Disease classification based on dermoscopic skin images using convolutional neural network in teledermatology system," in *2019 Int. Conf. Comput. Eng. Network, Intell. Multimedia, CENIM 2019 – Proceeding*, Surabaya, Indonesia, 2019, pp. 1–5. https://doi.org/10.1109/CENIM48368.2019.8973303

[4] N. B. Setiawan, F. Natalia, F. V. Ferdinand, S. Sudirman, and C. S. Ko, "Classification of skin diseases and disorders using convolutional neural network on a mobile application," *ICIC Express Lett. Part B Appl.*, vol. 12, no. 8, pp. 715–721, 2021. https://doi.org/10.24507/icicelb.12.08.715

[5] H. Li, Y. Pan, J. Zhao, and L. Zhang, "Skin disease diagnosis with deep learning: A review," *Neurocomputing*, vol. 464, pp. 364–393, 2021. https://doi.org/10.1016/j.neucom.2021.08.096

[6] T. Shanthi, R. S. Sabeenian, and R. Anand, "Automatic diagnosis of skin diseases using convolution neural network," *Microprocess. Microsyst.*, vol. 76, p. 103074, 2020. https://doi.org/10.1016/j.micpro.2020.103074

[7] R. Karthik, T. S. Vaichole, S. K. Kulkarni, O. Yadav, and F. Khan, "Eff2Net: An efficient channel attention-based convolutional neural network for skin disease classification," *Biomed. Signal Process. Control*, vol. 73, p. 103406, 2022. https://doi.org/10.1016/j.bspc.2021.103406

[8] M. A. Al-masni, D. H. Kim, and T. S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Comput. Methods Programs Biomed.*, vol. 190, p. 105351, 2020. https://doi.org/10.1016/j.cmpb.2020.105351

[9] C. Calderón, K. Sanchez, S. Castillo, and H. Arguello, "BILSK: A bilinear convolutional neural network approach for skin lesion classification," *Comput. Methods Programs Biomed. Updat.*, vol. 1, pp. 1–7, 2021. https://doi.org/10.1016/j.cmpbup.2021.100036

[10] M. M. Bejani and M. Ghatee, "A systematic review on overfitting control in shallow and deep neural networks," *Artificial Intelligence Review,* vol. 54, no. 8, pp. 6391–6438, 2021. https://doi.org/10.1007/s10462-021-09975-1

[11] A. Aboulmira, H. Hrimech, and M. Lachgar, "Comparative study of multiple CNN models for classification of 23 skin diseases," *Int. J. Online Biomed. Eng.*, vol. 18, no. 11, pp. 127–142, 2022. https://doi.org/10.3991/ijoe.v18i11.32517

[12] S. Goel and B. Hall, "Dermnet," [Online]. Available: https://www.kaggle.com/datasets/shubhamgoel27/dermnet/data

[13] K. A. Cahyanto, M. Yani, Rendi, A. Kustiyo, and T. Haryanto, "The effects of L2 regularization in EfficientNet for human skin disease multi-class classification," in *Proc. – 2023 6th Int. Conf. Comput. Informatics Eng. AI Trust. Risk Secur. Manag. (AI Trism), IC2IE 2023*, 2023, pp. 262–265. https://doi.org/10.1109/IC2IE60547.2023.10331108

[14] A. Balasundaram, A. Shaik, B. R. Alroy, A. Singh, and S. J. Shivaprakash, "Genetic algorithm optimized stacking approach to skin disease detection," *IEEE Access*, vol. 12, pp. 88950–88962, 2024. https://doi.org/10.1109/ACCESS.2024.3412791

[15] S. Chatterjee, J. M. Gil, and Y. C. Byun, "Early detection of multiclass skin lesions using transfer learning-based IncepX-Ensemble model," *IEEE Access*, vol. 12, pp. 113677–113693, 2024. https://doi.org/10.1109/ACCESS.2024.3432904

[16] J. R. N. Fernandes *et al.*, "Artificial intelligence on diagnostic aid of leprosy: A systematic literature review," *J. Clin. Med.*, vol. 13, no. 1, p. 180, 2024. https://doi.org/10.3390/jcm13010180

[17] J. Sadaiyandi, P. Arumugam, A. K. Sangaiah, and C. Zhang, "Stratified sampling-based deep learning approach to increase prediction accuracy of unbalanced dataset," *Electronics*, vol. 12, no. 21, p. 4423, 2023. https://doi.org/10.3390/electronics12214423

[18] A. Patil, A. Mehto, and S. Nalband, "Enhancing skin lesion diagnosis with data augmentation techniques: A review of the state-of-the-art," *Multimed. Tools Appl.*, vol. 84, pp. 25325–25364, 2024. https://doi.org/10.1007/s11042-024-20145-7

[19] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, 2019. https://doi.org/10.1186/s40537-019-0197-0

[20] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *36th Int. Conf. Mach. Learn. ICML 2019*, 2019, pp. 10691–10700.

[21] M. Tan and Q. V. Le, "EfficientNetV2: Smaller models and faster training," *Proc. Mach. Learn. Res.*, vol. 139, pp. 10096–10106, 2021.

[22] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350

[23] H. Soumare, A. Benkahla, and N. Gmati, "Deep learning regularization techniques to genomics data," *Array*, vol. 11, p. 100068, 2021. https://doi.org/10.1016/j.array.2021.100068

[24] G. Habib and S. Qureshi, "Optimization and acceleration of convolutional neural networks: A survey," *J. King Saud Univ. – Comput. Inf. Sci.*, vol. 34, no. 7, pp. 4244–4268, 2022. https://doi.org/10.1016/j.jksuci.2020.10.004

[25] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd Int. Conf. Learn. Represent. ICLR 2015 – Conf. Track Proc.*, 2015, pp. 1–15.

[26] A. Panthakkan, S. M. Anzar, S. Jamal, and W. Mansoor, "Concatenated Xception-ResNet50 — A novel hybrid approach for accurate skin cancer prediction," *Comput. Biol. Med.*, vol. 150, p. 106170, 2022. https://doi.org/10.1016/j.compbiomed.2022.106170

[27] V. A. Kumar, C. Chandana, G. Supraja, D. Haripriya, and E. Ravalika, "SCCNet: Skin cancer detection and multi-class classification using deep CNN model with estimated disease probabilities," *SN Comput. Sci.*, vol. 5, no. 6, 2024. https://doi.org/10.1007/s42979-024-03072-0

[28] S. Asif, Qurrat-ul-Ain, S. U. R. Khan, K. Amjad, and M. Awais, "SKINC-NET: An efficient lightweight deep learning model for multiclass skin lesion classification in dermoscopic images," *Multimed. Tools Appl.*, vol. 84, pp. 12531–12557, 2025. https://doi.org/10.1007/s11042-024-19489-x

[29] S. Hamida, D. Lamrani, M. A. Bouqentar, O. El Gannour, and B. Cherradi, "An integrated multimodal deep learning framework for accurate skin disease classification," *Int. J. Online Biomed. Eng.*, vol. 20, no. 2, pp. 78–94, 2024. https://doi.org/10.3991/ijoe.v20i02.43795

8 AUTHORS

Kurnia Adi Cahyanto received his master's degree in Informatics Engineering from Dian Nuswantoro University in 2014. Currently, he is pursuing his doctoral program at Diponegoro University Graduate School in the Information Systems Doctoral Study Program. His research interests include data mining and image processing (E-mail: kurnia@polindra.ac.id).

Kusworo Adi is a professor at the Department of Physics, Faculty of Science and Mathematics at Diponegoro University. He earned his doctorate degree from Bandung Institute of Technology. He currently holds a position as a lecturer at Diponegoro University's Graduate School of Information Systems Doctoral Study Program. His research interests include image processing and computer vision.

Catur Edi Widodo is a professor at the Department of Physics, Faculty of Science and Mathematics at Diponegoro University. He earned his doctorate degree from Gajah Mada University. Diponegoro University currently lists him as one of the lecturers at the Graduate School of Information Systems Doctoral Study Program. His research interests include computational physics.