narra j

Original Article

Evaluation of atopic dermatitis severity using artificial intelligence

Aga Maulana[1], Teuku R. Noviandy[1], Rivansyah Suhendra[2], Nanda Earlia[3,4], Mikyal Bulqiah[3], Ghazi M. Idroes[5], Nurdjannah J. Niode[6], Hizir Sofyan[7], Muhammad Subianto[7] and Rinaldi Idroes[8*]

[1]Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh, Indonesia; [2]Department of Information Technology, Faculty of Engineering, Universitas Teuku Umar, Meulaboh, Indonesia; [3]Dermatology Division, Dr. Zainoel Abidin Hospital, Banda Aceh, Indonesia; [4]Department of Dermatology and Venereology, Faculty of Medicine, Universitas Syiah Kuala, Banda Aceh, Indonesia; [5]Department of Occupational Health and Safety, Faculty of Health Sciences, Universitas Abulyatama, Aceh Besar, Indonesia; [6]Department of Dermatology and Venereology, Faculty of Medicine, Sam Ratulangi University, Manado, Indonesia; [7]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh, Indonesia; [8]Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh, Indonesia

*Corresponding author: rinaldi.idroes@usk.ac.id

Abstract

Atopic dermatitis is a prevalent and persistent chronic inflammatory skin disorder that poses significant challenges when it comes to accurately assessing its severity. The aim of this study was to evaluate deep learning models for automated atopic dermatitis severity scoring using a dataset of Aceh ethnicity individuals in Indonesia. The dataset of clinical images was collected from 250 patients at Dr. Zainoel Abidin Hospital, Banda Aceh, Indonesia and labeled by dermatologists as mild, moderate, severe, or none. Five pre-trained convolutional neural networks (CNN) architectures were evaluated: ResNet50, VGGNet19, MobileNetV3, MnasNet, and EfficientNetB0. The evaluation metrics, including accuracy, precision, sensitivity, specificity, and F1-score, were employed to assess the models. Among the models, ResNet50 emerged as the most proficient, demonstrating an accuracy of 89.8%, precision of 90.00%, sensitivity of 89.80%, specificity of 96.60%, and an F1-score of 89.85%. These results highlight the potential of incorporating advanced, data-driven models into the field of dermatology. These models can serve as invaluable tools to assist dermatologists in making early and precise assessments of atopic dermatitis severity and therefore improve patient care and outcomes.

Keywords: Atopic dermatitis, severity scoring, deep learning, convolutional neural network, image classification

Introduction

Atopic dermatitis (AD) is a prevalent chronic inflammatory skin disorder marked by persistent eczema and itching [1]. Patients often face a diminished quality of life, experiencing increased rates of sleep disturbances, anxiety, depression, and even suicidal thoughts [2]. It impacts all age groups, with approximately 15 million affected in the United States alone, of which 60% are children under 12 years-old [3]. Data from the Indonesian Pediatric Dermatology Research Group indicates that AD topped the list of pediatric dermatological diseases in 2022, with over 703,270 reported cases in Indonesia [4].

In recent years, artificial intelligence (AI) has revolutionized various sectors [5-7]. Notably, the healthcare sector has seen a surge in AI-driven devices, leading to enhanced accuracy,

https://narraj.org/

Received: November 6, 2023 - Accepted: December 18, 2023 - Published online: December 19, 2023

efficiency, and productivity [3,8-10]. Specifically in dermatology, the use of deep learning and convolutional neural networks (CNN) has enhanced image classification, object detection, and other analytical processes [11,12].

Traditional methods for diagnosing AD have predominantly relied on visual inspections by dermatologists. However, such approaches are susceptible to human error, making the diagnosis inherently subjective and potentially less optimal [13-15]. Recognizing this challenge, recent dermatological studies have harnessed machine learning and deep learning for enhanced skin disease diagnosis and severity assessment [15-18]. For instance, Medela *et al.* leveraged 604 dermatological images and utilized the Deep Expectation technique to achieve an 84.6% diagnosis accuracy [16]. Wu *et al.* applying LASSO Regression and Random Forest on datasets from NCBI obtained an 84.4% accuracy in assessing AD severity [17]. Another study processed 1,000 AD images via a CNN, achieving 94.44% accuracy [18]. Al-Masni *et al.* applied full-resolution CNN on melanoma images and noted 77.11% testing accuracy [19]. However, a significant limitation of these studies is their predominant reliance on datasets primarily composed of Caucasian skin types. This presents potential biases, as the morphological presentation of skin conditions can vary across different ethnicities. Without representation of diverse skin tones, the models might not perform as effectively or accurately when diagnosing or assessing skin conditions in non-Caucasian populations [20].

Addressing this gap, this study sought to develop a deep learning model for scoring the severity of AD, with a focus on the Acehnese ethnicity. Dataset from Dr. Zainoel Abidin Hospital in Banda Aceh, Indonesia was collected. By using and modifying pre-trained models like ResNet50, VGGNet19, MobileNetV3, MnasNet, and EfficientNetB0, our aim was to achieve robust AD severity classification. Our approach promotes more inclusive skin disease diagnosis by focusing on an underrepresented population. Additionally, automating severity scoring could lead to faster and more consistent assessments, reducing the workload for dermatologists.

Methods

Dataset

We utilized a dataset sourced from the dermatology division of Dr. Zainoel Abidin Hospital, Banda Aceh, Indonesia. This dataset contained images of Acehnese patients, captured with a 12-megapixel smartphone camera under the supervision of a dermatologist. To create the dataset, skin lesion images were manually cropped from the original patient photos. In total, 3037 images were collected from 250 patients. The patients included in the dataset were diagnosed with atopic dermatitis and received treatment between 2021 and 2023. The age range of the patients was between 18 and 65 years. The severity of the atopic dermatitis was assessed using the objective SCORAD (SCORing Atopic Dermatitis) score, which considered criteria such as erythema, oedema, crusting, excoriation, lichenification, and skin dryness. These images were then categorized into four severity levels: none, mild, moderate, and severe. Visual examples of images representing each severity level are presented in **Figure 1**. The training dataset was comprised of 2126 labeled images across the four severity classes, while the testing dataset contained 911 additional labeled images. Specifically, within the training data, there were 703 images classified as none, 401 classified as mild, 948 classified as moderate, and 985 classified as severe.

Model architectures

We utilize several CNN architectures as core components of the proposed methodology, including ResNet50 [21], VGGNet19 [22], MobileNetV3 [23], MnasNet [24], and EfficientNetB0 [25]. The rationale behind this selection was twofold. Firstly, deeper models were opted like ResNet50 and VGGNet19 to ensure that complex patterns and intricacies of the skin conditions could be captured. These models, with their comprehensive architectures, could discern nuanced features indicative of varying severity levels of AD. Secondly, efficient models such as MobileNetV3, MnasNet, and EfficientNetB0 were incorporated to ensure fast and resource-efficient predictions, making the proposed system practical for real-world clinical applications. Furthermore, these models have proven successful across a wide range of computer vision applications, as substantiated in previous works [26-32].

All of our models were fine-tuned from ImageNet weights [33]. By initializing the models with pre-trained weights from the ImageNet dataset, we were not only leveraging their ability to capture general visual features foundational for our approach but also optimizing them for the specifics of our dataset and objective. The strengths inherent to each architecture, whether in terms of efficiency, scalability, or suitability for resource-constrained environments, make them particularly well-suited for our methodology.

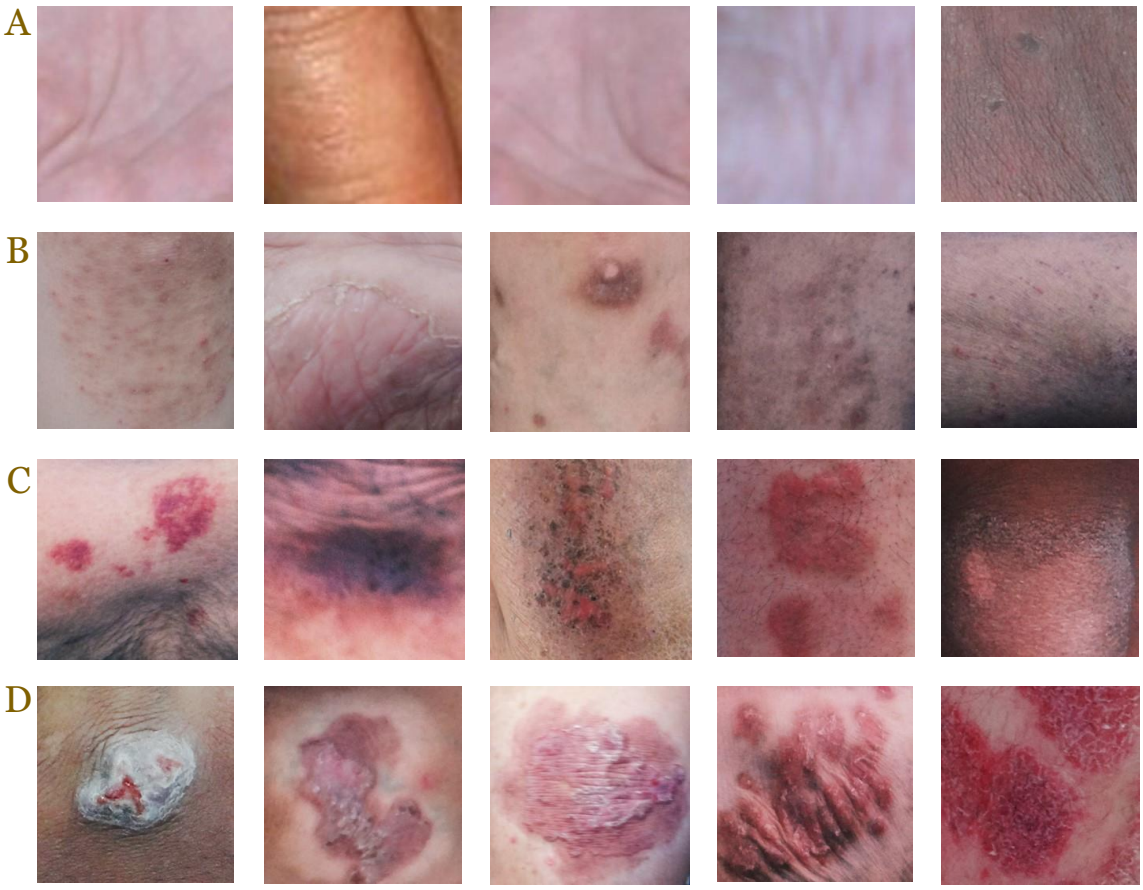

Figure 1. Representative of dataset used in this study based on severity: (A) none; (B) mild; (C) moderate; and (D) severe.

Data preparation

To prepare the data for modeling, several necessary preprocessing steps [34,35] were conducted. The first step was to manually crop the region of interest (ROI) with a 1:1 aspect ratio on the AD wound area. Next, data were labelled using SCORAD under the supervision of a dermatologist. Then, the labeled data were resized to 250×250 pixels and converted to the portable network graphics (PNG) format, with the aim of speeding up the training process. Subsequently, data underwent an augmentation process, which included flip, rotate, distortion, skew, and zoom. This enhanced the training process, making it more efficient and reliable [36]. The parameters and values of data augmentation are presented in **Table 1**.

Table 1. Parameters and values of data augmentation

Data augmentation method	Parameter value	Action
Flip	0.5	Flip the image horizontal and vertical with a probability of 0.5
Rotate	10	Rotate the image left and right rotation with a probability of 0.1
Distortion	8	Distortion the image with magnitude and grid dimension with a probability 0.08
Skew	True	Skew the image
Zoom	1	Scaling the image

Proposed approach

Each model was modified by adding a global average pooling 2D layer, a dense layer, and a dropout layer. To ensure that the pre-trained layers remained intact during training, they were set to non-trainable. The global average pooling 2D layer was used to condense the spatial information from the output of the modified model, reducing its spatial dimensions while retaining essential features [37]. Following this, a dense layer comprising 256 units was incorporated, tailored to the demands of AD classification. This layer captured higher-level features, enhancing the model's representational abilities. To mitigate the risk of overfitting, a dropout layer was introduced subsequent to the dense layer. By randomly deactivating certain input units during training, it ensured that the network did not overly depend on specific units, thereby fostering a more robust and generalizable data representation [38]. Finally, the severity scoring was made using a dense layer that contained a single unit with softmax activation.

To begin training our modified models, the initial step was the setting up of the hyperparameters. Hyperparameters are predefined configurations external to the model, which significantly influence its learning process and effectiveness. The employed hyperparameter are presented in **Table 2.** The training was conducted over ten epochs with a batch size of 32, facilitating iterative refinement of the models. The Adam optimizer with a learning rate of 0.001 was used for optimization, owing to its proven effectiveness in deep learning tasks [39]. The categorical cross-entropy loss function was chosen, being particularly suited for multi-class classification problems.

Table 2. Hyperparameter used to train of all modified models

Hyperparameter	Value
Optimizer	Adam
Epoch	10
Learning rate	0.001
Loss function	Categorical cross-entropy
Batch size	32

Model evaluation

In this study, we utilized various evaluation metrics to gauge the models' performance. These metrics encompass accuracy, precision, sensitivity, specificity, and F1-score. Because the severity scoring of AD is a multiclass problem, the weighted average of these metrics was used to ensure that each class's contribution was proportional to its representation in the dataset. The formulas for calculating accuracy, precision, sensitivity, F1-score, and the weighted average of these metrics are presented in equations 1–9 [40].

$$\text{Accuracy} = \frac{\text{TP} + \text{FN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \quad (1)$$

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

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

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (4)$$

$$\text{F1} - \text{score} = \frac{2\ \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (5)$$

$$\text{Weighted average precision} = \frac{\Sigma(\text{precision}_i * \text{weight}_i)}{\Sigma\ \text{weight}_i} \quad (6)$$

$$\text{Weighted average sensitivity} = \frac{\Sigma(\text{sensitivity}_i * \text{weight}_i)}{\Sigma\ \text{weight}_i} \quad (7)$$

$$\text{Weighted average specificity} = \frac{\Sigma(\text{specificity}_i * \text{weight}_i)}{\Sigma\ \text{weight}_i} \quad (8)$$

$$\text{Weighted average F1} - \text{score} = \frac{\Sigma(\text{F1} - \text{score}_i * \text{weight}_i)}{\Sigma\ \text{weight}_i} \quad (9)$$

True positive (TP) is the count of positive cases that were correctly identified; false negative (FN) represents the number of cases erroneously classified as negative; false positive (FP) denotes the number of cases incorrectly labeled as positive; and true negative (TN) indicates the count of correctly identified negative cases.

Results

We successfully developed the models specifically trained to assess AD severity using exclusive dataset, which had not been employed previously. The training and validation accuracy, along with the training and validation loss throughout the model's training, are presented in **Figure 2**. Our data indicated that ResNet50, VGGNet19, MobileNetV3, and EfficientNetB0 learned effectively, with validation accuracy generally increased as epochs progress, despite occasional spikes. This indicates that these models are benefiting from extended training, optimizing their weights to better generalize and predict unseen data. Such a trend suggests that the models are converging and the algorithms behind them are appropriately adjusting based on the training data, which can lead to better performance on AD severity scoring.

However, an interesting pattern emerged for the MNasNet model. The training accuracy was lower than the validation accuracy, and the validation loss diminished significantly after two epochs. This behavior suggested that the model might be underfitting the training data. Underfitting occurs when a model is too simplistic and fails to capture the underlying patterns of the data. The fact that the training accuracy is lower than the validation accuracy further supports this hypothesis, as it indicates the model struggled even with data it has seen before. The rapid decrease in validation loss might indicate that the model has found a general solution that works reasonably well for the validation data but may not be capturing all the intricacies of the training dataset.

The evaluation of our trained models revealed varying levels of performance in assessing AD severity (**Table 3**). The deep architecture of ResNet50 produced the best results, with an accuracy of 89.80%. This likely stems from its ability to learn intricate patterns, yielding a high precision of 90.00%, sensitivity of 89.80%, specificity of 96.70% and F1-score of 89.95% (**Table 3**). EfficientNetB0 followed closely behind, with an accuracy of 85.20% and nearly equivalent precision and sensitivity. VGGNet19 and MobileNetV3 had the middle ground, with moderate outcomes. VGGNet19 had an accuracy of 83.88%, while MobileNetV3 had 81.09%. Their precision, sensitivity, and other metrics aligned with these moderate numbers. MnasNet, however, demonstrated the least favorable performance with an accuracy of just 63.16%. It can be attributed to the architecture design that trades some accuracy for reduced model size and faster inference, possibly restricting its ability to learn deeper dataset patterns. Despite a higher specificity of around 87.72%, its lowest F1-score indicated weaker balanced classification capabilities compared to the other models.

Table 3. Weighted average performance of modified ResNet50, VGGNet, MobileNetV3, MnasNet, and EfficientNet models to determine the atopic dermatitis severity

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
ResNet50	89.80	90.00	89.80	96.60	89.85
VGGNet	83.88	84.19	84.50	94.68	83.88
MobileNetV3	81.09	80.99	81.09	93.70	80.76
MnasNet	63.16	64.97	63.16	87.72	60.81
EfficientNet	85.20	85.30	85.20	95.07	85.23

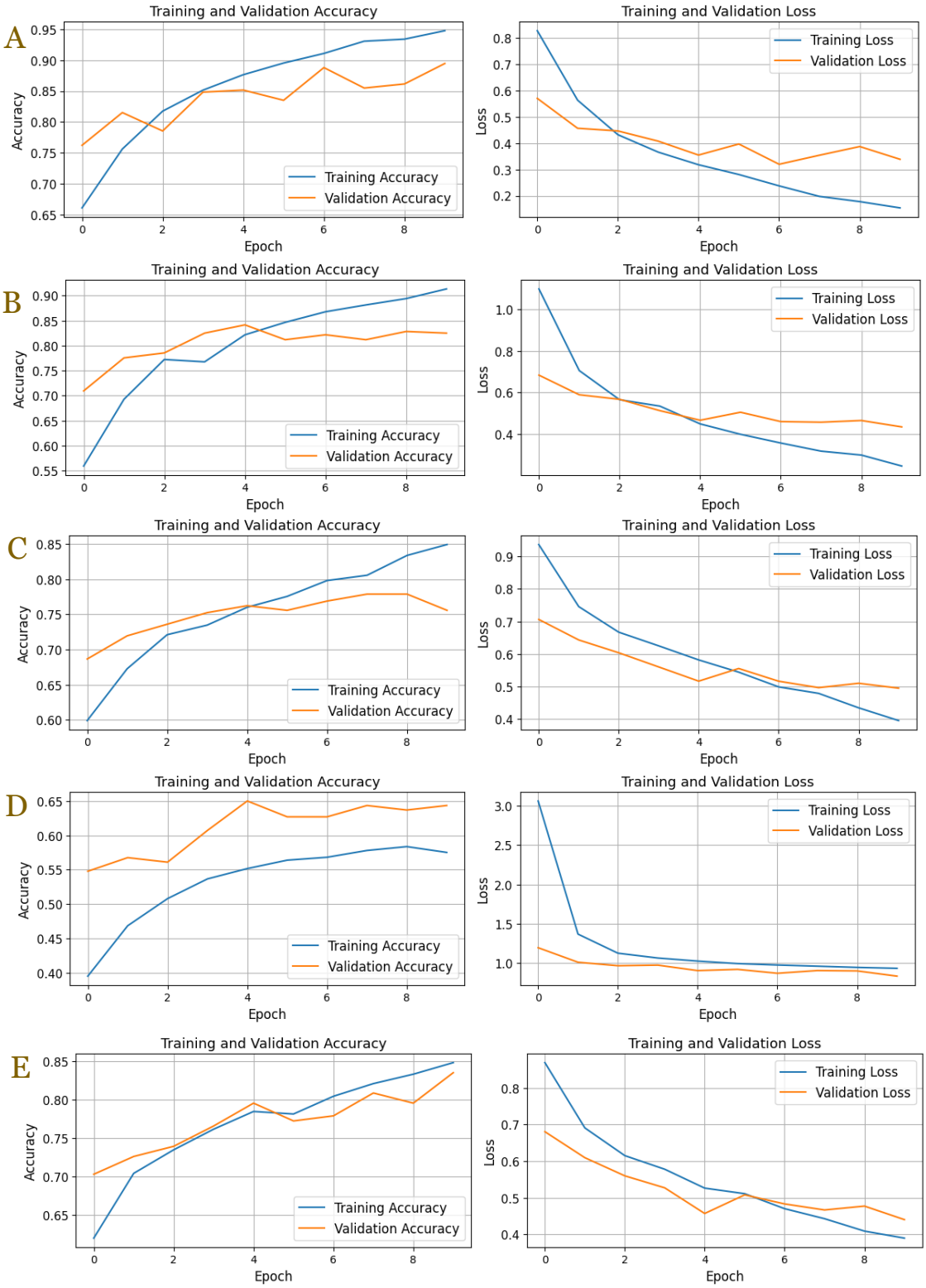

Figure 2. Training and validation results of modified (A) ResNet50; (B) VGGNet; (C) MobileNetV3; (D) MnasNet; and (E) EfficientNet.

To obtain a deeper insight into the models' severity scoring, the confusion matrix depicted in **Figure 3** offers a comprehensive overview of its classifications. It compared the actual values with the values predicted by the model. The matrix helps in understanding not just the errors of a model, but more importantly, the types of errors that are being made.

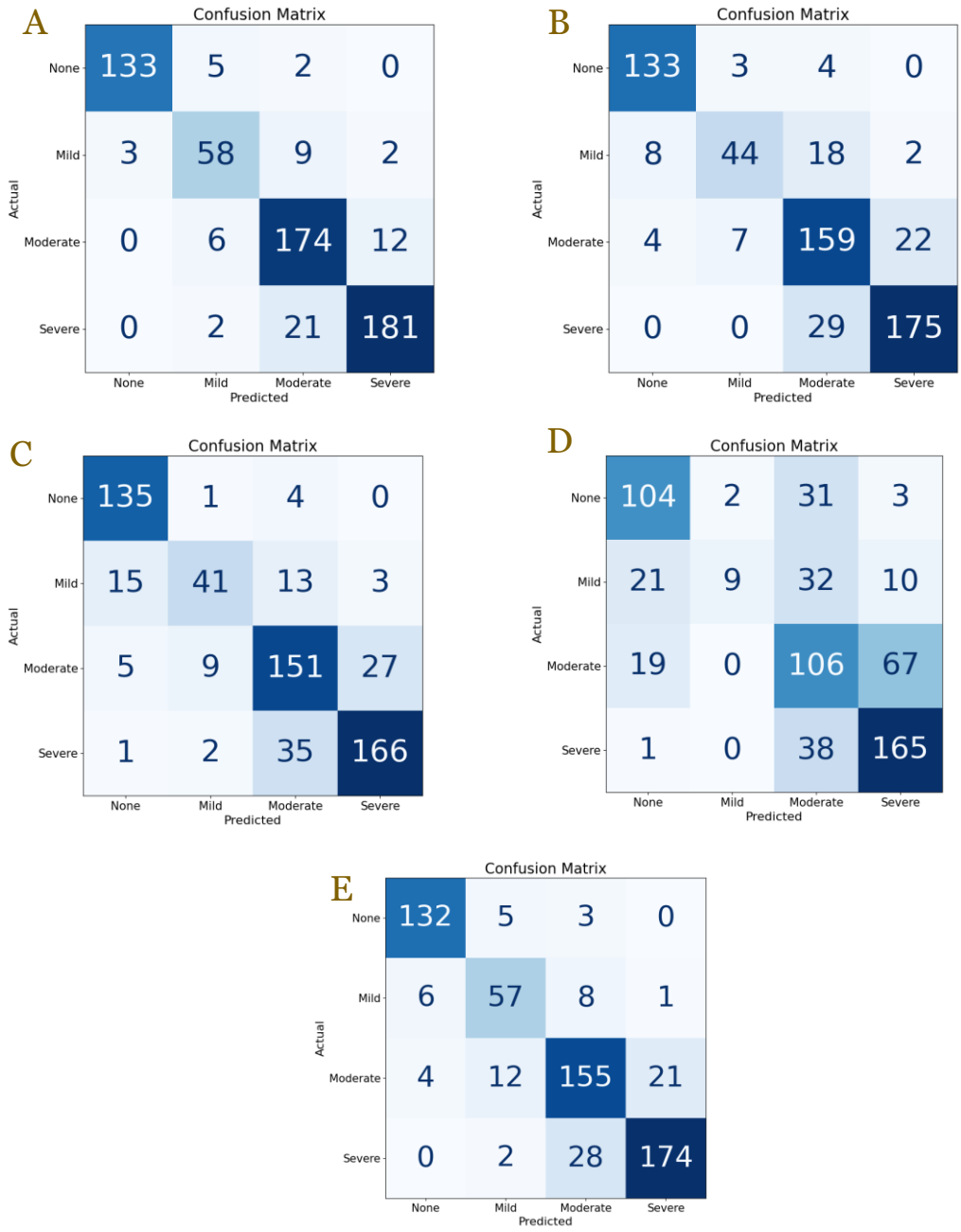

Figure 3. Confusion matrix on testing set of modified models: (A) ResNet50; (B) VGGNet; (C) MobileNetV3; (D) MnasNet; and (E) EfficientNet. Each cell represents the count of predictions, where one class was predicted as another. The diagonal cells (top-left to bottom-right) represent correct predictions, where the predicted class matches the actual class. The off-diagonal cells show the misclassifications. Cells with higher numbers are colored more intensely, and cells with lower numbers are lighter. The optimal number in the diagonal cells should be as high as possible, indicating correct predictions, while the off-diagonal cells should ideally contain low numbers, indicating fewer errors.

The confusion matrix indicated that ResNet50 model demonstrated commendable accuracy, particularly for the none and moderate severities. Although there were instances of misclassification, they were relatively minimal, suggesting that the model has a strong capability in its predictions across the categories. The VGGNet19 model, while having strengths in identifying the none and severe severities, revealed areas of potential improvement, particularly between the mild and moderate categories (**Figure 3**). The misclassifications in these mid-range categories indicate room for refinement in the models' performance. The MobileNetV3 model showcased impressive accuracy, especially for the none and severe classifications. However,

similar to VGGNet19, it had challenges in differentiating between the mild and moderate severities, suggesting areas of potential enhancement (**Figure 3**). MnasNet performance was more varied. While it demonstrated strength in the severe category, there were vulnerabilities in differentiating between mild and moderate severities. Notably, the none to moderate predictions showed significant false positives, highlighting areas for improvement (**Figure 3**). Finally, the EfficientNetB0 model displayed a robust performance, particularly in the extreme categories none and severe (**Figure 3**). Misclassifications were present but relatively fewer, indicating a balanced prediction capability across all severity levels.

Discussion

The integration of AI in dermatology, specifically for assessing AD severity, represents a significant advancement in the field. Studies have demonstrated that AI models, particularly those with high accuracy, can be effectively integrated into diagnostic workflows [41,42]. This integration not only assists dermatologists in making more precise and consistent assessments but also minimizes the likelihood of human error and subjective judgments in severity evaluations [43]. Such advancements are important in tailoring treatment strategies and leading to improved patient outcomes.

Beyond the immediate clinical setting, the adoption of AI in dermatology represents a broader shift in healthcare towards precision medicine, where treatments and interventions are customized for individual patients. With the ability to rapidly analyze and interpret vast amounts of data, AI could offer insights into patient-specific factors that may influence disease progression and treatment responses. Moreover, as AI models evolve to consider diverse datasets, they could help ensure equity in healthcare by ensuring all populations benefit from the latest advancements in medical technology.

Overall, our models show significant promise in aiding dermatologists and general practitioners to more accurately classify AD severity levels, leading to more accurate diagnoses and improved patient care. However, challenges remain, such as potential bias due to the regional specificity of the dataset and the subjectivity in data labeling affecting the models' generalizability. For instance, studies have shown variations in AI model performance when applied to data sets predominantly composed of Caucasian patients compared to more diverse populations [43-46].

Future studies should focus on cross-regional validation, model refinement, clinical integration, and dataset diversification to further develop the potential of deep learning in dermatology. Expanding the dataset to better represent global population diversity can improve generalizability and access to quality dermatological care worldwide. Additional collaboration between computer scientists and clinical experts could fine-tune models for enhanced performance and applicability in real-world settings. Research into model interpretations and uncertainty quantification is also important to promote understanding, trust, and adoption among practitioners. While this study provides a promising first step, continued work is needed to address limitations and fully harness the potential of deep learning in dermatology in a responsible and equitable manner.

Conclusion

We have successfully explored the application of deep learning models in the assessment of AD severity using a dataset of Aceh ethnicity individuals in Indonesia. We evaluated a diverse range of CNN architectures, from deeper models like ResNet50 and VGGNet19 to more efficient models such as MobileNetV3, MnasNet, and EfficientNetB0. By benchmarking these models, we gained an important understanding of their comparative strengths for dermatological analysis. The results represent a significant advancement in the utilization of deep learning models as valuable tools in dermatological practice. The models offer the potential for more accurate and efficient assessments of AD severity, making them a crucial asset in advancing dermatological care, particularly in regions with limited specialty expertise.

Ethics approval

The study has been ethically approved by the Institutional Ethics Committee of the Dr. Zainoel Abidin Hospital (protocol code no.23–04–091, 16 May 2023). Informed consent was obtained from all subjects involved in the study.

Acknowledgments

We would like to thank Dr. Zainoel Abidin Hospital and Universitas Syiah Kuala for supporting this study.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study was funded by Direktorat Riset, Teknologi, dan Pengabdian Kepada Masyarakat - Direktorat Jendral Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi and Universitas Syiah Kuala through "Penelitian Fundamental Reguler" scheme with grant number: 168/E5/PG.02.00.PL/2023.

Underlying data

Derived data supporting the findings of this study are available from the corresponding author on request.

How to cite

Maulana A, Noviandy TR, Suhendra R, *et al.* Evaluation of atopic dermatitis severity using artificial intelligence. Narra J 2023; 3 (3): e511 - http://doi.org/10.52225/narra.v3i3.511.

References

1. Renert-Yuval Y, Guttman-Yassky E. What's new in atopic dermatitis. Dermatol Clin 2019;37(2):205-213.
2. Yew YW, Thyssen JP, Silverberg JI. A systematic review and meta-analysis of the regional and age-related differences in atopic dermatitis clinical characteristics. J Am Acad Dermatol 2019;80(2):390-401.
3. Litjens G, Kooi T, Bejnordi BE, *et al.* A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.
4. Dimawan RSA, Prakoeswa FRS, Pramuningtyas R. Pediatric viral and bacterial skin infection profile. Berk Ilmu Kesehat Kulit Dan Kelamin 2022;34(3):184-188.
5. Noviandy TR, Maulana A, Idroes GM, *et al.* Integrating genetic algorithm and lightgbm for QSAR modeling of acetylcholinesterase inhibitors in alzheimer's disease drug discovery. Malacca Pharm 2023;1(2):48-54.
6. Rufo DD, Debelee TG, Ibenthal A, *et al.* Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM). Diagnostics 2021;11(9):1714.
7. Hidayat T, Hadinata E, Damanik IS, *et al.* Implementation of hybrid CNN-XGBoost method for leukemia detection problem. Infolitika J Data Sci 2023;1(1):15-21.
8. Idroes R, Noviandy TR, Maulana A, *et al.* Retention index prediction of flavor and fragrance by multiple linear regression and the genetic algorithm. Int Rev Model Simul IREMOS 2019;12(6):373.
9. Noviandy TR, Maulana A, Sasmita NR, *et al.* The prediction of Kovats retention indices of essential oils at gas chromatography using genetic algorithm. J Eng Sci Technol 2020;17(1):0306-0326.
10. Maulana A, Faisal FR, Noviandy TR, *et al.* Machine Learning Approach for Diabetes Detection Using Fine-Tuned XGBoost Algorithm. Infolitika J Data Sci 2023;1(1):1-7.
11. Li Z, Koban KC, Schenck TL, *et al.* Artificial intelligence in dermatology image analysis: Current developments and future trends. J Clin Med 2022;11(22):6826.
12. Idroes R, Noviandy TR, Maulana A, *et al.* Application of genetic algorithm-multiple linear regression and artificial neural network determinations for prediction of Kovats retention index. Int Rev Model Simul IREMOS 2021;14(2):137.
13. Shrivastava VK, Londhe ND, Sonawane RS, *et al.* First review on psoriasis severity risk stratification: An engineering perspective. Comput Biol Med 2015;63:52-63.

14. Suhendra R, Suryadi S, Husdayanti N, *et al.* Evaluation of gradient boosted classifier in atopic dermatitis severity score classification. Heca J Appl Sci 2023;1(2):54-61.
15. Maulana A, Noviandy TR, Suhendra R, *et al.* Performance analysis and feature extraction for classifying the severity of atopic dermatitis diseases. 2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE). 2023.
16. Medela A, Mac Carthy T, Aguilar Robles SA, *et al.* Automatic SCOring of atopic dermatitis using deep learning: A pilot study. JID Innov 2022;2(3):100107.
17. Wu S, Lei L, Hu Y, *et al.* Machine learning-based prediction models for atopic dermatitis diagnosis and evaluation. Fundam Res 2023;15:14.
18. Guimarães P, Batista A, Zieger M, *et al.* Artificial Intelligence in multiphoton tomography: Atopic dermatitis diagnosis. Sci Rep 2020;10(1):7968.
19. Al-masni MA, Al-antari MA, Choi MT, *et al.* Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Comput Methods Programs Biomed 2018;162:221-231.
20. Groh M, Harris C, Soenksen L, *et al.* Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
21. He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
22. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv Prepr ArXiv14091556 2014.
23. Howard A, Sandler M, Chu G, *et al.* Searching for mobilenetv3. Proceedings of the IEEE/CVF international conference on computer vision. 2019.
24. Zoph B, Vasudevan V, Shlens J, *et al.* Learning transferable architectures for scalable image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
25. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning. PMLR; 2019.
26. Florez R, Palomino-Quispe F, Coaquira-Castillo RJ, *et al.* A CNN-based approach for the driver drowsiness detection by real-time eye state identification. Appl Sci 2023;13:7849.
27. Talukder MdA, Islam MdM, Uddin MdA, *et al.* An efficient deep learning model to categorize brain tumor using reconstruction and fine-tuning. Expert Syst Appl 2023;230:120534.
28. Prusty S, Patnaik S, Dash SK. ResNet50V2: A transfer learning model to predict pneumonia with chest x-ray images. 2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS). IEEE; 2022.
29. Idroes GM, Maulana A, Suhendra R, *et al.* TeutongNet: A fine-tuned deep learning model for improved forest fire detection. Leuser J Environ Stud 2023;1(1):1-8.
30. Elfatimi E, Eryigit R, Elfatimi L. Beans leaf diseases classification using MobileNet models. IEEE Access 2022;10:9471-9482.
31. Xin C, Liu Z, Zhao K, *et al.* An improved transformer network for skin cancer classification. Comput Biol Med 2022;149:105939.
32. Albelwi S, Mahmood A. A framework for designing the architectures of deep convolutional neural networks. Entropy 2017;19(6):242.
33. Kornblith S, Shlens J, Le Q V. Do better imagenet models transfer better? Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
34. Kottath A V, Bharathi SVS. Image preprocessing techniques in skin diseases prediction using deep learning: A review. 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE; 2022.
35. Noviandy TR, Maulana A, Idroes GM, *et al.* Ensemble machine learning approach for quantitative structure activity relationship based drug discovery: A Review. Infolitika J Data Sci 2023;1(1):32-41.
36. Huang L, Qin J, Zhou Y, *et al.* Normalization techniques in training DNNs: Methodology, analysis and application. IEEE Trans Pattern Anal Mach Intell 2023:1-20.
37. Islam MdM, Tasnim N, Baek JH. Human gender classification using transfer learning via pareto frontier CNN networks. Inventions 2020;5(2):16.
38. Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929-1958.

39. Nguyen LC, Nguyen-Xuan H. Deep learning for computational structural optimization. ISA Trans 2020;103:177-191.
40. Noviandy TR, Maulana A, Emran TB, *et al.* QSAR classification of beta-secretase 1 inhibitor activity in alzheimer's disease using ensemble machine learning algorithms. Heca J Appl Sci 2023;1(1):1-7.
41. De A, Sarda A, Gupta S, *et al.* Use of artificial intelligence in dermatology. Indian J Dermatol 2020;65(5):352.
42. Du-Harpur X, Watt FM, Luscombe NM, *et al.* What is AI? Applications of artificial intelligence to dermatology. Br J Dermatol 2020;183(3):423-430.
43. Jeong HK, Park C, Henao R, *et al.* Deep learning in dermatology: A systematic review of current approaches, outcomes, and limitations. JID Innov 2023;3(1):100150.
44. Fletcher RR, Nakeshimana A, Olubeko O. Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. Front Artif Intell 2021;3:561802.
45. Daneshjou R, Vodrahalli K, Liang W, *et al.* Disparities in dermatology AI: Assessments using diverse clinical images. ArXiv Prepr ArXiv211108006 2021.
46. Aggarwal P, Papay FA. Artificial intelligence image recognition of melanoma and basal cell carcinoma in racially diverse populations. J Dermatol Treat 2022;33(4):2257-2262.