Original Article

Check for updates

Available online at
www.heca-analitika.com/hjas

Heca Journal of Applied Sciences

Vol. 2, No. 2, 2024

Explainable Deep Learning Approach for Mpox Skin Lesion Detection with Grad-CAM

Ghazi Mauer Idroes [1,2], Teuku Rizky Noviandy [3,*], Talha Bin Emran [4] and Rinaldi Idroes [5]

[1] Graduate School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; idroesghazi_k3@abulyatama.ac.id (G.M.I);
[2] Department of Occupational Health and Safety, Faculty of Health Sciences, Universitas Abulyatama, Aceh Besar 23372, Indonesia
[3] Interdisciplinary Innovation Research Unit, Graha Primera Saintifika, Aceh Besar 23771, Indonesia; trizkynoviandy@gmail.com (T.R.N.)
[4] Department of Pharmacy, BGC Trust University Bangladesh, Chittagong 4381, Bangladesh; talhabmb@bgctub.ac.bd (T.B.E.)
[5] Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: trizkynoviandy@gmail.com

Article History

Received 14 July 2024
Revised 31 August 2024
Accepted 6 September 2024
Available Online 19 September 2024

Keywords:
Artificial intelligence
Medical imaging
Monkeypox
XAI

Abstract

Mpox is a viral zoonotic disease that presents with skin lesions similar to other conditions like chickenpox, measles, and hand-foot-mouth disease, making accurate diagnosis challenging. Early and precise detection of mpox is critical for effective treatment and outbreak control, particularly in resource-limited settings where traditional diagnostic methods are often unavailable. While deep learning models have been applied successfully in medical imaging, their use in mpox detection remains underexplored. To address this gap, we developed a deep learning-based approach using the ResNet50v2 model to classify mpox lesions alongside five other skin conditions. We also incorporated Grad-CAM (Gradient-weighted Class Activation Mapping) to enhance model interpretability. The results show that the ResNet50v2 model achieved an accuracy of 99.33%, precision of 99.34%, sensitivity of 99.33%, and an F1-score of 99.32% on a dataset of 1,594 images. Grad-CAM visualizations confirmed that the model focused on relevant lesion areas for its predictions. While the model performed exceptionally well overall, it struggled with misclassifications between visually similar diseases, such as chickenpox and mpox. These results demonstrate that AI-based diagnostic tools can provide reliable, interpretable support for clinicians, particularly in settings with limited access to specialized diagnostics. However, future work should focus on expanding datasets and improving the model's capacity to distinguish between similar conditions.

1. Introduction

Mpox, or monkeypox, is a viral zoonotic disease increasingly recognized as a significant public health concern due to its growing incidence worldwide [1, 2]. According to the World Health Organization (WHO), between January 1, 2022, and January 1, 2023, more than 85,000 confirmed cases of mpox were reported across 110 countries, marking a dramatic surge compared to previous years where the disease was predominantly confined to Central and West African regions [3]. This increase in cases, particularly in non-endemic countries, has underscored the urgent need for enhanced surveillance, diagnostics, and containment strategies [4].

Mpox is characterized by symptoms that include fever, swollen lymph nodes, and, most notably, distinct skin lesions [5]. These lesions are often similar in appearance

to those caused by other diseases, such as chickenpox, measles, and hand-foot-mouth disease, making accurate diagnosis challenging [6, 7]. With the potential for rapid transmission and severe complications, especially in vulnerable populations, the early and precise identification of mpox is critical for effective treatment and outbreak control.

Traditional diagnostic methods for mpox, such as polymerase chain reaction (PCR), are considered the gold standard for confirming the presence of the virus [8, 9]. However, these methods are often time-consuming and expensive and require specialized laboratory equipment and expertise that may not be readily available in resource-limited settings [10]. In underdeveloped regions, the delays and costs associated with traditional diagnostics can hinder timely diagnosis and treatment, exacerbating the spread of the disease. This not only affects individual health outcomes but also places a significant economic burden on healthcare systems, contributing to increased healthcare costs and resource strain [11]. Consequently, the impact on public health is magnified, highlighting the need for more accessible and cost-effective diagnostic solutions.

Artificial intelligence (AI) has shown great promise in addressing some of the challenges associated with traditional diagnostic methods [12–15]. In particular, AI-driven image analysis has emerged as a powerful tool for rapidly and accurately detecting various diseases, including those that present with skin lesions [16, 17]. AI systems can analyze medical images to identify patterns and features that may not be immediately apparent to the human eye, thereby improving diagnostic accuracy and speed [18]. This technology offers a valuable complement to existing diagnostic methods, particularly in settings where access to specialized laboratory testing is limited.

Various studies have demonstrated the effectiveness of AI in medical diagnostics, particularly in the field of dermatology. Convolutional neural networks (CNNs) have been widely used to classify skin conditions based on image data, such as dermatitis atopic [19, 20], psoriasis [21, 22], and fungal infections [23, 24], with high accuracy, often matching or surpassing the diagnostic abilities of dermatologists. Among these, ResNet (Residual Network) models have been particularly successful due to their ability to learn complex features from images and their robustness in handling large datasets [25]. These studies underscore AI's potential to transform how diseases like mpox are diagnosed, offering a scalable and accessible solution for healthcare providers.

Explainable AI (XAI) techniques, which improve the transparency of AI models by explaining their decision-making processes, have become increasingly important in recent years [26, 27]. In this study, we use deep learning along with Grad-CAM (Gradient-weighted Class Activation Mapping) to enhance the interpretability of the model's predictions [28]. XAI allows users to understand which parts of an image contributed most to a decision, increasing the trustworthiness of the AI system and providing visual explanations that align with expert knowledge [29–31]. Our goal is to develop a diagnostic tool that is both accurate and interpretable, ensuring that professionals can confidently trust and rely on the model's predictions.

This study aims to develop a deep learning-based approach using modified ResNet50v2, EfficientNetB4, and DenseNet169 models to accurately detect mpox from skin lesion images. By training these models on a comprehensive dataset that includes images of mpox alongside other similar dermatological conditions, we seek to create a reliable tool capable of distinguishing mpox from diseases with overlapping presentations. This research aims to provide a practical, efficient, and scalable diagnostic aid that can be utilized in various healthcare settings, especially in areas where traditional diagnostic resources are limited.

The potential impact of this research is substantial, particularly in resource-limited settings where access to advanced diagnostic tools is scarce. By providing a low-cost, automated solution for mpox detection, we can help healthcare providers make informed decisions and implement appropriate public health interventions. Furthermore, the proposed model could be integrated into mobile health applications, enabling remote diagnosis and monitoring, thus extending the reach of healthcare services to underserved populations.

2. Materials and Methods

2.1. Dataset

The dataset utilized in this study was obtained from a study by Ali et al. [32, 33], and it contains images of patients with mpox and five other non-mpox skin conditions: chickenpox, cowpox, hand-foot-mouth disease (HFMD), healthy, and measles. The comprehensive dataset consists of 1,594 images, allowing for effective training and testing of the deep learning models in distinguishing between mpox and other dermatological conditions that present with similar visual symptoms.

The images were divided into three subsets, with 70% allocated for training, 20% for validation, and 10% for

Table 1. Image distribution of six classes into training, validation, and testing sets.

Subset	Class						Total
	Chickenpox	Cowpox	HFMD	Healthy	Measles	Mpox	
Training	75	66	161	114	55	284	755
Validation	55	49	117	78	39	204	542
Testing	34	30	72	42	26	93	297
Total							**1594**

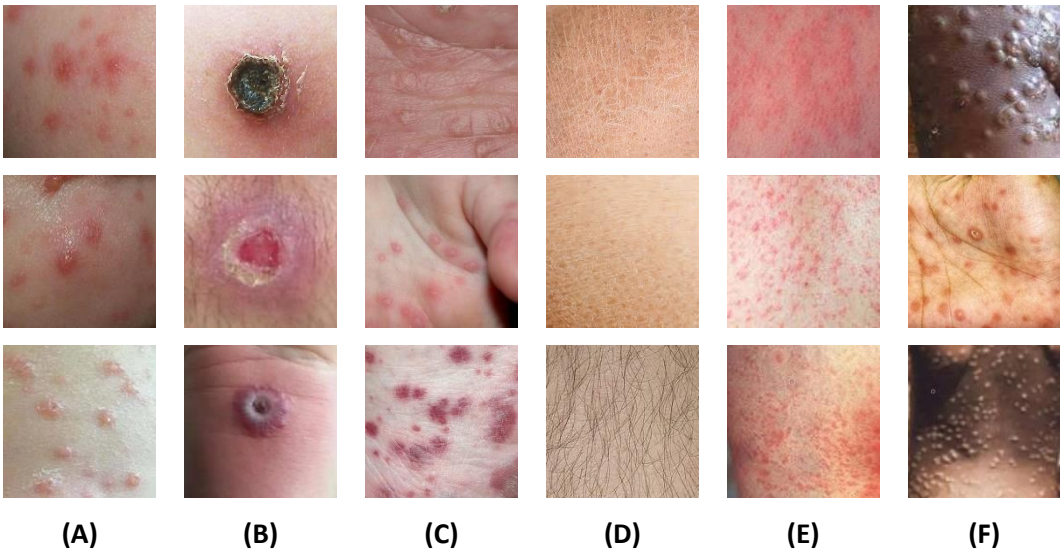

Figure 1. Sample images from each class: (a) chickenpox, (b) cowpox, (c) HFMD, (d) healthy, (e) measles, (f) mpox.

testing. This split ensures that the models have enough data for training while providing sufficient samples for validation and testing to avoid overfitting and ensure generalization [34]. The distribution of images across each of the six classes is shown in Table 1, with each class proportionally represented in all subsets.

Figure 1 displays sample images from each class to provide the visual characteristics of the various conditions. These samples reflect the diversity in skin lesion appearances across the six categories, providing insight into the model's learning process.

2.2. Deep Learning Models

This study employed three state-of-the-art deep learning architectures—ResNet50v2, EfficientNetB4, and DenseNet169—to develop an effective classification system for detecting mpox skin lesions and distinguishing them from other dermatological conditions with similar visual characteristics.

ResNet50v2 is a residual network architecture that utilizes skip connections to mitigate the vanishing gradient problem and enable deeper network training [35, 36]. With 50 layers, it captures both low- and high-level image features, making it suitable for complex image classification tasks like skin lesion detection. ResNet's residual blocks allow the model to learn hierarchical features more efficiently, which is essential for distinguishing visually similar diseases such as mpox, chickenpox, and measles.

EfficientNetB4 is part of the EfficientNet family, which optimizes accuracy and computational efficiency by scaling the network's depth, width, and resolution using a compound coefficient [37]. EfficientNetB4 is larger than the base version (B0) and provides enhanced feature extraction capabilities, making it particularly useful for medical image analysis where both precision and speed are crucial [38]. Its efficient scaling makes it well-suited for identifying complex patterns in skin lesions with minimal computational overhead, which is especially important for practical deployment in resource-limited healthcare settings.

DenseNet169, a dense convolutional network, connects each layer to every other layer, promoting feature reuse and efficient gradient flow [39]. This architecture is highly effective in medical imaging tasks because it propagates features across layers, enhancing the model's ability to capture subtle differences between skin conditions [40]. DenseNet169's dense connectivity enables it to maintain strong performance while using fewer parameters than traditional networks, making it a powerful tool for

Table 2. Hyperparameters used for training ResNet50v2, EfficientNetB4, and DenseNet169.

Hyperparameter	Value
Batch Size	32
Optimizer	Adam
Learning Rate	1e-5
Loss Function	Categorical Cross Entropy
Epochs	10
Activation Function	ReLU

accurately identifying skin lesions, even when presented with small dataset sizes.

To adapt the ResNet50v2, EfficientNetB4, and DenseNet169 architectures for mpox detection, we applied a consistent strategy across all models. We selectively froze the earlier layers of each network to retain the benefits of pre-trained weights while fine-tuning the deeper layers to focus on the specific features of skin lesions [41]. Custom layers were added on top of each model, including a GlobalAveragePooling2D layer for feature aggregation and a fully connected Dense layer with six output units and a softmax activation function for multiclass classification. This approach allowed the models to capture relevant features while minimizing the risk of overfitting and ensuring computational efficiency, making them well-suited for the task of distinguishing mpox from other dermatological conditions.

2.3. Model Training

To ensure each model's optimal performance, we applied the same set of hyperparameters during training. These hyperparameters were selected based on common best practices for image classification tasks in deep learning, ensuring consistency across the models for comparative evaluation. Table 2 summarizes the hyperparameters used for training all models.

The models were trained using a batch size of 32, which balances computational efficiency with convergence stability. We used the Adam optimizer, known for its adaptive learning rate, to achieve faster convergence [42]. The learning rate was set to 1e-5 to allow for gradual learning without significant oscillations in the loss function. The categorical cross-entropy loss function was employed since we are dealing with a multi-class classification problem. Finally, each model was trained for ten epochs to ensure sufficient learning without overfitting.

2.4. Explainability with Grad-CAM

In this study, we employed Gradient-weighted Class Activation Mapping (Grad-CAM) to enhance the explainability of our deep learning models. While deep learning models, such as ResNet50v2, EfficientNetB4, and DenseNet169, have demonstrated remarkable performance in image classification tasks, they often function as "black boxes," making it difficult to interpret how specific decisions are made [43]. In medical applications, where the consequences of misdiagnosis can be severe, model interpretability is crucial for building trust and enabling healthcare professionals to make informed decisions based on the model's outputs [44].

Grad-CAM provides a solution to this interpretability issue by creating visual explanations for the predictions made by the model [28]. Grad-CAM generates heatmaps that highlight the regions in an image that contributed most to the model's decision [45]. By visualizing these regions, clinicians can gain insight into which features of the skin lesion the model is focusing on when classifying a condition as mpox or another dermatological disease. The Grad-CAM technique works by computing the gradient of the predicted class concerning the feature maps of the final convolutional layer of the model. These gradients are then used to weight the importance of different features in the feature maps, combined to generate a heatmap over the input image. The resulting heatmap highlights the most relevant areas of the image, showing where the model "paid attention" during its decision-making process.

The use of Grad-CAM in our approach not only improves prediction accuracy but also provides clear visual explanations that align with the medical understanding of skin lesions. This is especially important for distinguishing mpox from other conditions, like chickenpox, cowpox, and measles, which may present similar visual symptoms. With Grad-CAM, we can ensure the model focuses on medically relevant features, such as the texture or shape of the lesion, rather than irrelevant background elements.

2.5. Performance Evaluation

In this study, the performance of the deep learning models was evaluated using a comprehensive set of metrics: weighted average accuracy, precision, sensitivity, specificity, and F1-score [19]. These metrics were selected to assess the models' ability to accurately classify mpox

Table 3. Performance comparison of deep learning models.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
ResNet50v2	99.33	99.34	99.33	99.69	99.32
EfficientNetB4	62.63	61.60	62.63	85.51	61.75
DenseNet169	93.94	94.11	93.94	97.81	93.89

while distinguishing it from other dermatological conditions in a multiclass setting.

Accuracy measures the overall proportion of correct predictions across all classes [46]. In a multiclass scenario, it is essential to calculate weighted accuracy to ensure that larger classes do not disproportionately influence the results. Precision focuses on the model's ability to correctly identify positive cases while minimizing false positives [47], and sensitivity evaluates the model's effectiveness in capturing all true positive cases, which is critical for minimizing false negatives [48].

On the other hand, specificity assesses how well the model identifies true negatives [49], ensuring that non-mpox cases are correctly classified, thus reducing false positives. Finally, the F1-score, as the harmonic mean of precision and sensitivity, provides a balanced measure of the model's performance, especially when an imbalance exists between classes [50]. The weighted F1-score ensures that the evaluation accounts for the varying representation of each class in the dataset.

The equation for weighted accuracy, precision, sensitivity, specificity, and F1-score were computed for the multiclass classification task represented in Equations 1-5, respectively:

$$Accuracy = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n}(TP_i + FP_i + FN_i)} \quad (1)$$

$$Precision = \frac{\sum_{i=1}^{n} w_i \cdot \frac{TP_i}{TP_i + FP_i}}{\sum_{i=1}^{n} w_i} \quad (2)$$

$$Sensitivity = \frac{\sum_{i=1}^{n} w_i \cdot \frac{TP_i}{TP_i + FN_i}}{\sum_{i=1}^{n} w_i} \quad (3)$$

$$Specificity = \frac{\sum_{i=1}^{n} w_i \cdot \frac{TN_i}{TN_i + FP_i}}{\sum_{i=1}^{n} w_i} \quad (4)$$

$$F1\text{-}Score = \frac{\sum_{i=1}^{n} w_i \cdot \frac{2 \cdot TP_i}{2 \cdot TP_i + FP_i + FN_i}}{\sum_{i=1}^{n} w_i} \quad (5)$$

where TP_i, FP_i, TN_i, and FN_i, represent the true positives, false positives, true negatives, and false negatives for class i, respectively, and w_i represents the weight for each class based on its proportion in the dataset.

2.6. Software and Tools

To implement the mpox skin lesion detection system, we used Python 3.10.9 along with several essential libraries. TensorFlow 2.10.1 was employed for deep learning model development, allowing efficient training and fine-tuning of ResNet50v2, EfficientNetB4, and DenseNet169. For data manipulation, Pandas 2.2.2 and NumPy 1.23.5 were used to handle datasets and perform numerical operations. Visualization of results, such as accuracy and loss, was done with Matplotlib 3.6.3, while Scikit-learn 1.5.2 was utilized for performance evaluation and metrics calculation. These tools ensured a smooth, efficient, and reproducible workflow.

3. Results and Discussion

The results of the deep learning models trained to classify mpox skin lesions and other similar dermatological conditions are presented and discussed here. These models were evaluated based on five key metrics: accuracy, precision, sensitivity, specificity, and F1-score, which collectively provide a comprehensive view of their performance in distinguishing between multiple skin conditions. The performance results for each model are summarized in Table 3.

The performance of ResNet50v2 stands out, achieving the highest scores across all evaluation metrics. With an accuracy of 99.33%, precision of 99.34%, sensitivity of 99.33%, specificity of 99.69%, and an F1-score of 99.32%, ResNet50v2 demonstrates exceptional performance in identifying mpox and distinguishing it from other similar skin conditions. This high performance can be attributed to the model's ability to effectively learn and utilize both low- and high-level features through its deep residual connections, which help prevent degradation while training deeper layers.

DenseNet169 also performed notably well, with an accuracy of 93.94%, precision of 94.11%, and F1-score of 93.89%. Its dense connectivity architecture, which enables feature reuse and efficient gradient flow, allowed it to capture significant patterns in the dataset, resulting in a solid performance. However, DenseNet169, while robust, falls short compared to ResNet50v2, particularly in accuracy and specificity. The lower specificity (97.81%) of DenseNet169 suggests that it may have slightly more

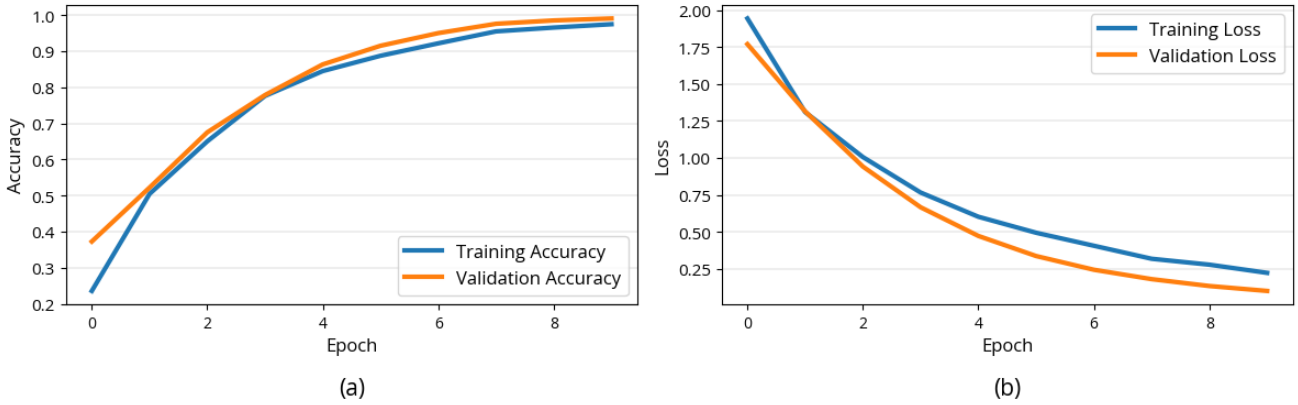

Figure 2. Visualization of the modified ResNet50v2 training and validation: (A) accuracy; (B) loss.

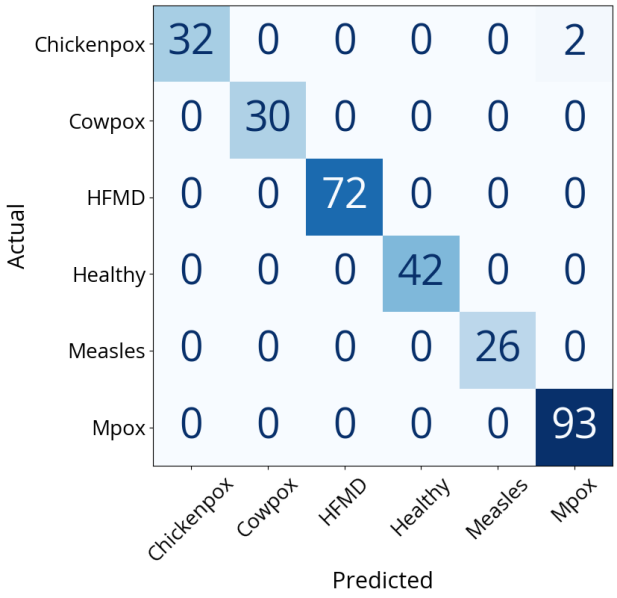

Figure 3. Confusion matrix of the testing set prediction from modified ResNet50 model.

difficulty differentiating non-mpox cases, which could lead to more false positives.

On the other hand, EfficientNetB4 performed significantly worse than the other two models, with an accuracy of 62.63%, precision of 61.60%, and an F1-score of 61.75%. Although EfficientNetB4 is designed to optimize performance while balancing computational efficiency, in this case, it was less capable of capturing the intricate features necessary for accurate classification. The relatively high specificity of 85.51% indicates that it was better at identifying true negatives than classifying positive cases, as reflected by its lower sensitivity (62.63%).

ResNet50v2 showed outstanding performance in the classification task, as demonstrated by the results in Table 3. To further illustrate the model's training performance, Figure 2 provides a visualization of the training and validation accuracy and loss curves for ResNet50v2. The model's accuracy improves steadily over the training epochs, with both training and validation accuracy nearing 100%. This indicates that the model is learning effectively, with minimal signs of overfitting. The corresponding loss curves also demonstrate a steady decline in training and validation, confirming that the model converges well. These visualizations affirm the robustness and reliability of the ResNet50v2 model in distinguishing mpox lesions from other conditions.

The classification performance of the modified ResNet50 model can be further analyzed using the confusion matrix shown in Figure 3. This matrix explains how well the model distinguishes between the six skin conditions in the testing set. As observed, the ResNet50 model demonstrates excellent classification accuracy, with nearly perfect results for most classes. For example, the model correctly classified all cowpox, HFMD, healthy, and measles cases without misclassifications. The mpox class was also well-identified, with 93 correct predictions and only two instances of chickenpox being misclassified as mpox. This misclassification could be due to the visual similarity between the lesions caused by these two diseases. Both conditions exhibit skin rashes and lesions that can appear similar, especially in early stages or mild cases, which might lead the model to misidentify chickenpox as mpox. Despite this misclassification, it is important to note that there were no instances of mpox being misclassified as any other condition, indicating that the model is highly sensitive to identifying true mpox cases. This is crucial for clinical applications, as the model's primary goal is to ensure that mpox cases are correctly identified to prevent mismanagement and control outbreaks.

In the next step of our analysis, we used Grad-CAM to visualize the classification results of the ResNet50v2 model on a sample of images from the test set. This technique helps to understand which regions of the skin

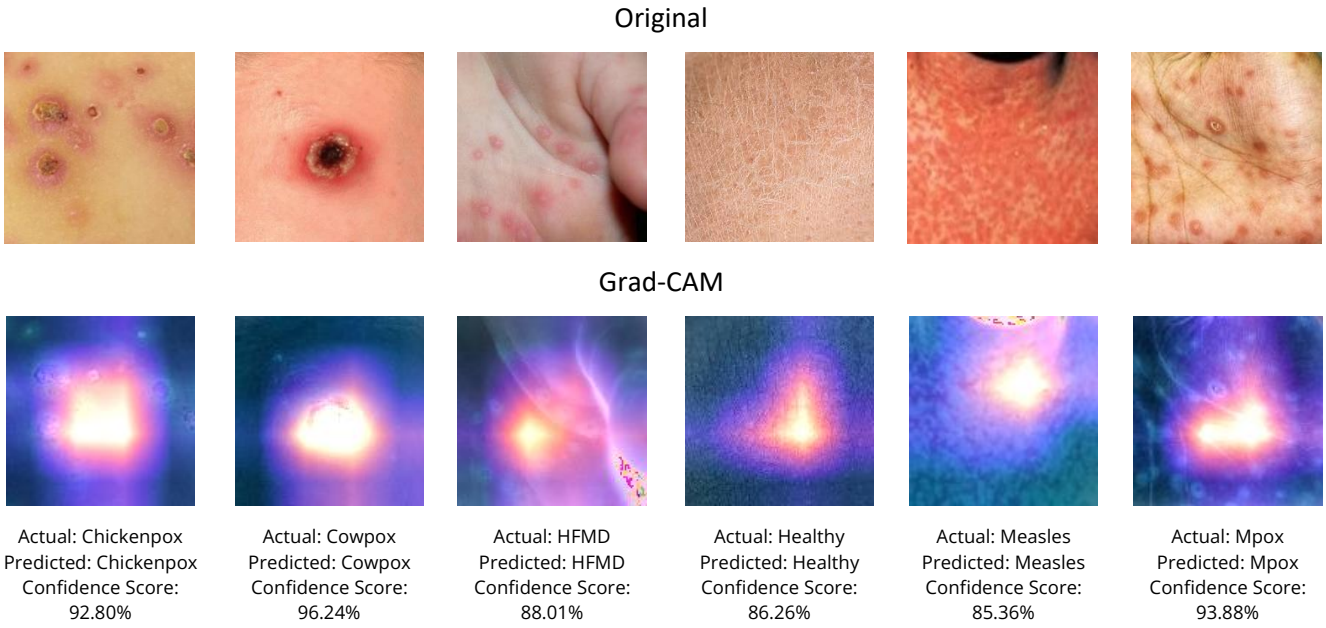

Figure 4. Grad-CAM visualization of model predictions. The top row shows the original images, and the bottom row shows the corresponding Grad-CAM heatmaps with confidence scores for each prediction. (a) chickenpox, (b) cowpox, (c) HFMD, (d) healthy, (e) measles, (f) mpox.

lesion images the model focused on while making its predictions. Figure 4 compares the original skin lesion images (top row) and their corresponding Grad-CAM heatmaps (bottom row), highlighting the areas that contributed most to the model's classification.

As seen in the figure, the Grad-CAM visualizations confirm that the model's attention is concentrated on the relevant regions of the lesions, particularly in correctly classified cases like cowpox, HFMD, and mpox. For example, in the mpox image, the heatmap emphasizes the central lesion, aligning with medical expectations for such a diagnosis. The confidence scores for each prediction are also included, indicating the model's certainty in its classifications, with confidence values ranging from 85.36% to 96.24%, further reinforcing the model's reliability in these predictions.

The findings of this study have important implications for the development of AI-based diagnostic tools, particularly in the context of mpox and other visually similar skin conditions. By demonstrating the effectiveness of the ResNet50v2 model, this study highlights the potential for deep learning models to serve as reliable, fast, and accessible diagnostic aids in clinical settings. The high accuracy, precision, and sensitivity achieved by ResNet50v2 indicate that AI can effectively support healthcare providers in distinguishing between diseases that present with overlapping symptoms, such as skin lesions, common in viral infections like mpox, chickenpox, and measles.

Grad-CAM also adds a layer of interpretability, which is crucial in medical applications. The ability to visually verify the regions of an image that the model used for its decision-making process helps clinicians understand and trust the AI's predictions. This transparency can enhance the model's integration into clinical workflows, providing healthcare professionals with predictive accuracy and explanations that align with medical understanding.

Despite the promising results, there are several limitations to this study. First, the dataset was relatively small, with 1,594 images distributed across six disease classes. Although the model performed well on this dataset, a larger, more diverse dataset with more variations in skin types, age groups, and disease stages would be necessary to fully assess the model's generalizability. Additionally, while strong, the model's performance was not flawless—some misclassifications occurred, particularly between chickenpox and mpox. This suggests that the model may struggle to distinguish between conditions with very similar visual features. Another limitation is the focus on image data alone. In real-world clinical settings, skin condition diagnoses often incorporate additional patient information, such as medical history, symptom progression, and laboratory results. Incorporating these contextual data into the model could improve its diagnostic accuracy and reduce the risk of misclassification.

Future research should address this study's limitations by expanding the dataset to include more diverse and comprehensive examples of skin lesions from various

populations and settings. This will help to improve the model's generalizability and robustness in real-world applications. Additionally, exploring the combination of image-based deep learning models with other forms of patient data, such as clinical history or laboratory results, could lead to more holistic and accurate diagnostic tools. Moreover, further work could enhance the model's ability to differentiate between similar conditions, such as Chickenpox and mpox. Techniques such as transfer learning or multi-task learning, which allow models to share knowledge between tasks, could be explored to improve the model's capacity to handle these challenging cases. Finally, integrating the model into mobile health applications could be a valuable direction for future research, as this would enable remote diagnostics in resource-limited settings where access to specialized healthcare providers may be limited.

4. Conclusions

This study demonstrates the effectiveness of the ResNet50v2 model in accurately detecting mpox and differentiating it from other dermatological conditions, achieving an impressive accuracy of 99.33%, precision of 99.34%, and F1-score of 99.32%. The model's performance, supported by Grad-CAM visualizations, focuses on medically relevant features, providing high diagnostic accuracy and interpretability. Despite its strong results, challenges remain in distinguishing between visually similar conditions, such as chickenpox and mpox, highlighting the need for further dataset expansion and model refinement. These findings underscore the potential of deep learning in supporting clinical diagnosis, particularly in resource-limited settings, and pave the way for future improvements in AI-driven healthcare tools.

Author Contributions: Conceptualization, G.M.I., T.R.N. and R.I.; methodology, G.M.I., T.R.N. and T.B.E.; software, G.M.I. and T.R.N.; validation, T.R.N., T.B.E. and R.I.; formal analysis, G.M.I. and T.B.E.; investigation, G.M.I. and T.B.E.; resources, T.R.N. and R.I.; data curation, T.R.N. and T.B.E.; writing—original draft preparation, G.M.I. and T.B.E.; writing—review and editing, T.R.N. and R.I.; visualization, G.M.I.; supervision, T.R.N., T.B.E. and R.I.; project administration, T.R.N.; funding acquisition, T.R.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset utilized in this study was obtained from the publicly available repository Kaggle. The "Mpox Skin Lesion Dataset Version 2.0 (MSLD v2.0)" was originally published by Joydip Paul and can be accessed via the following link: https://www.kaggle.com/datasets/joydippaul/mpox-skin-lesion-dataset-version-20-msld-v20. All data used in this research are freely available under the terms provided by the dataset's author.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Karagoz, A., Tombuloglu, H., Alsaeed, M., Tombuloglu, G., AlRubaish, A. A., Mahmoud, A., Smajlović, S., Ćordić, S., Rabaan, A. A., and Alsuhaimi, E. (2023). Monkeypox (Mpox) Virus: Classification, Origin, Transmission, Genome Organization, Antiviral Drugs, and Molecular Diagnosis, *Journal of Infection and Public Health*, Vol. 16, No. 4, 531–541. doi:10.1016/j.jiph.2023.02.003.

2. Ullah, M., Li, Y., Munib, K., and Zhang, Z. (2023). Epidemiology, Host Range, and Associated Risk Factors of Monkeypox: An Emerging Global Public Health Threat, *Frontiers in Microbiology*, Vol. 14. doi:10.3389/fmicb.2023.1160984.

3. Yu, X., Shi, H., and Cheng, G. (2023). Mpox Virus: Its Molecular Evolution and Potential Impact on Viral Epidemiology, *Viruses*, Vol. 15, No. 4, 995. doi:10.3390/v15040995.

4. Selvaraj, N., Shyam, S., Dhurairaj, P., Thiruselvan, K., Thiruselvan, A., Kancherla, Y., and Kandamaran, P. (2023). Mpox: Epidemiology, Clinical Manifestations and Recent Developments in Treatment and Prevention, *Expert Review of Anti-Infective Therapy*, Vol. 21, No. 6, 641–653. doi:10.1080/14787210.2023.2208346.

5. Maronese, C. A., Avallone, G., Aromolo, I. F., Spigariolo, C. B., Quattri, E., Ramoni, S., Carrera, C. G., and Marzano, A. V. (2023). Mpox: An Updated Review of Dermatological Manifestations in the Current Outbreak, *British Journal of Dermatology*, Vol. 189, No. 3, 260–270. doi:10.1093/bjd/ljad151.

6. van Nispen, C., Reffett, T., Long, B., Gottlieb, M., and Frawley, T. C. (2023). Diagnosis and Management of Monkeypox: A Review for the Emergency Clinician, *Annals of Emergency Medicine*, Vol. 81, No. 1, 20–30. doi:10.1016/j.annemergmed.2022.07.014.

7. Biswas, D., and Tešić, J. (2024). Binarydnet53: A Lightweight Binarized CNN for Monkeypox Virus Image Classification, *Signal, Image and Video Processing*, Vol. 18, No. 10, 7107–7118. doi:10.1007/s11760-024-03379-8.

8. Zhou, Y., and Chen, Z. (2023). Mpox: A Review of Laboratory Detection Techniques, *Archives of Virology*, Vol. 168, No. 8, 221. doi:10.1007/s00705-023-05848-w.

9. Chauhan, R. P., Fogel, R., and Limson, J. (2023). Overview of Diagnostic Methods, Disease Prevalence and Transmission of Mpox (Formerly Monkeypox) in Humans and Animal Reservoirs, *Microorganisms*, Vol. 11, No. 5, 1186. doi:10.3390/microorganisms11051186.

10. Maffert, P., Reverchon, S., Nasser, W., Rozand, C., and Abaibou, H. (2017). New Nucleic Acid Testing Devices to Diagnose Infectious Diseases in Resource-Limited Settings, *European Journal of Clinical Microbiology & Infectious Diseases*, Vol. 36, No. 10, 1717–1731. doi:10.1007/s10096-017-3013-9.

11. Musa-Booth, T. O., Medugu, N., Adegboro, B., and Babazhitsu, M. (2023). A Review of the Epidemiology, Diagnosis, Treatment, Vaccines and Economic Impact of Human Monkeypox (Mpox) Outbreaks, *African Journal of Clinical and Experimental Microbiology*, Vol. 24, No. 1, 1–8.

12. Noviandy, T. R., Nainggolan, S. I., Raihan, R., Firmansyah, I., and Idroes, R. (2023). Maternal Health Risk Detection Using Light Gradient Boosting Machine Approach, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 48–55. doi:10.60084/ijds.v1i2.123.

13. Mascarenhas, M., Afonso, J., Ribeiro, T., Andrade, P., Cardoso, H., and Macedo, G. (2023). The Promise of Artificial Intelligence

in Digestive Healthcare and the Bioethics Challenges It Presents, *Medicina*, Vol. 59, No. 4, 790. doi:10.3390/medicina59040790.

14. Noviandy, T. R., Idroes, G. M., Syukri, M., and Idroes, R. (2024). Interpretable Machine Learning for Chronic Kidney Disease Diagnosis: A Gaussian Processes Approach, *Indonesian Journal of Case Reports*, Vol. 2, No. 1, 24–32. doi:10.60084/ijcr.v2i1.204.
15. Hidayat, T., Hadinata, E., Damanik, I. S., Vikki, Z., and Irvanizam, I. (2023). Implementation of Hybrid CNN-XGBoost Method for Leukemia Detection Problem, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 15–21. doi:10.60084/ijds.v1i1.87.
16. Li, Z., Koban, K. C., Schenck, T. L., Giunta, R. E., Li, Q., and Sun, Y. (2022). Artificial Intelligence in Dermatology Image Analysis: Current Developments and Future Trends, *Journal of Clinical Medicine*, Vol. 11, No. 22, 6826. doi:10.3390/jcm11226826.
17. Yeasmin, M. N., Al Amin, M., Joti, T. J., Aung, Z., and Azim, M. A. (2024). Advances of AI in image-based computer-aided diagnosis: A review, *Array*, Vol. 23, 100357. doi:10.1016/j.array.2024.100357.
18. Suganyadevi, S., Seethalakshmi, V., and Balasamy, K. (2022). A Review on Deep Learning in Medical Image Analysis, *International Journal of Multimedia Information Retrieval*, Vol. 11, No. 1, 19–38. doi:10.1007/s13735-021-00218-1.
19. Maulana, A., Noviandy, T. R., Suhendra, R., Earlia, N., Bulqiah, M., Idroes, G. M., Niode, N. J., Sofyan, H., Subianto, M., and Idroes, R. (2023). Evaluation of Atopic Dermatitis Severity Using Artificial Intelligence, *Narra J*, Vol. 3, No. 3, e511. doi:10.52225/narra.v3i3.511.
20. Suhendra, R., Suryadi, S., Husdayanti, N., Maulana, A., Noviandy, T. R., Sasmita, N. R., Subianto, M., Earlia, N., Niode, N. J., and Idroes, R. (2023). Evaluation of Gradient Boosted Classifier in Atopic Dermatitis Severity Score Classification, *Heca Journal of Applied Sciences*, Vol. 1, No. 2, 54–61. doi:10.60084/hjas.v1i2.85.
21. Aijaz, S. F., Khan, S. J., Azim, F., Shakeel, C. S., and Hassan, U. (2022). Deep Learning Application for Effective Classification of Different Types of Psoriasis, *Journal of Healthcare Engineering*, Vol. 2022, 1–12. doi:10.1155/2022/7541583.
22. Okamoto, T., Kawai, M., Ogawa, Y., Shimada, S., and Kawamura, T. (2022). Artificial Intelligence for the Automated Single-Shot Assessment of Psoriasis Severity, *Journal of the European Academy of Dermatology and Venereology*, Vol. 36, No. 12, 2512–2515. doi:10.1111/jdv.18354.
23. Gao, W., Li, M., Wu, R., Du, W., Zhang, S., Yin, S., Chen, Z., and Huang, H. (2021). The Design and Application of an Automated Microscope Developed Based on Deep Learning for Fungal Detection in Dermatology, *Mycoses*, Vol. 64, No. 3, 245–251. doi:10.1111/myc.13209.
24. Xu, J., Luo, Y., Wang, J., Tu, W., Yi, X., Xu, X., Song, Y., Tang, Y., Hua, X., Yu, Y., Yin, H., Yang, Q., and Huang, W. E. (2023). Artificial Intelligence-Aided Rapid and Accurate Identification of Clinical Fungal Infections by Single-Cell Raman Spectroscopy, *Frontiers in Microbiology*, Vol. 14. doi:10.3389/fmicb.2023.1125676.
25. He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep Residual Learning for Image Recognition, *Computer Vision and Pattern Recognition*.
26. Noviandy, T. R., Idroes, G. M., Hardi, I., Afjal, M., and Ray, S. (2024). A Model-Agnostic Interpretability Approach to Predicting Customer Churn in the Telecommunications Industry, *Infolitika Journal of Data Science*, Vol. 2, No. 1, 34–44. doi:10.60084/ijds.v2i1.199.
27. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Enhancing Loan Approval Decision-Making: An Interpretable Machine Learning Approach Using LightGBM for Digital Economy Development, *Malaysian Journal of Computing (MJOC)*, Vol. 9, No. 1, 1734–1745. doi:10.24191/mjoc.v9i1.25691.
28. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 618–626. doi:10.1109/ICCV.2017.74.
29. Holzinger, A., Biemann, C., Pattichis, C. S., and Kell, D. B. (2017). What Do We Need to Build Explainable AI Systems for the Medical Domain?, *ArXiv Preprint ArXiv:1712.09923*.
30. Ali, S., Akhlaq, F., Imran, A. S., Kastrati, Z., Daudpota, S. M., and Moosa, M. (2023). The Enlightening Role of Explainable Artificial Intelligence in Medical & Healthcare Domains: A Systematic Literature Review, *Computers in Biology and Medicine*, Vol. 166, 107555. doi:10.1016/j.compbiomed.2023.107555.
31. Chen, H., Gomez, C., Huang, C.-M., and Unberath, M. (2022). Explainable Medical Imaging AI Needs Human-Centered Design: Guidelines and Evidence from a Systematic Review, *Npj Digital Medicine*, Vol. 5, No. 1, 156. doi:10.1038/s41746-022-00699-2.
32. Ali, S. N., Ahmed, M. T., Jahan, T., Paul, J., Sakeef Sani, S. M., Noor, N., Asma, A. N., and Hasan, T. (2024). A Web-Based Mpox Skin Lesion Detection System Using State-of-the-Art Deep Learning Models considering Racial Diversity, *Biomedical Signal Processing and Control*, Vol. 98, 106742. doi:10.1016/j.bspc.2024.106742.
33. Ali, S. N., Ahmed, M. T., Paul, J., Jahan, T., Sani, S. M., Noor, N., and Hasan, T. (2022). Monkeypox Skin Lesion Detection Using Deep Learning Models: A Feasibility Study, *ArXiv Preprint ArXiv:2207.03342*. doi:10.48550/arXiv.2207.03342.
34. Noviandy, T. R., Maulana, A., Zulfikar, T., Rusyana, A., Enitan, S. S., and Idroes, R. (2024). Explainable Artificial Intelligence in Medical Imaging: A Case Study on Enhancing Lung Cancer Detection through CT Images, *Indonesian Journal of Case Reports*, Vol. 2, No. 1, 6–14. doi:10.60084/ijcr.v2i1.150.
35. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Identity Mappings in Deep Residual Networks. doi:10.48550/arXiv.1603.05027.
36. Idroes, G. M., Maulana, A., Suhendra, R., Lala, A., Karma, T., Kusumo, F., Hewindati, Y. T., and Noviandy, T. R. (2023). TeutongNet: A Fine-Tuned Deep Learning Model for Improved Forest Fire Detection, *Leuser Journal of Environmental Studies*, Vol. 1, No. 1, 1–8. doi:10.60084/ljes.v1i1.42.
37. Tan, M., and Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, *International Conference on Machine Learning*, PMLR, 6105–6114.
38. Talukder, M. A., Layek, M. A., Kazi, M., Uddin, M. A., and Aryal, S. (2024). Empowering COVID-19 Detection: Optimizing Performance through Fine-Tuned EfficientNet Deep Learning Architecture, *Computers in Biology and Medicine*, Vol. 168, 107789. doi:10.1016/j.compbiomed.2023.107789.
39. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely Connected Convolutional Networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.
40. Ashwini, A., Purushothaman, K. E., Rosi, A., and Vaishnavi, T. (2023). Artificial Intelligence based real-time automatic detection and classification of skin lesion in dermoscopic samples using DenseNet-169 architecture, *Journal of Intelligent & Fuzzy Systems*, Vol. 45, No. 4, 6943–6958. doi:10.3233/JIFS-233024.
41. Gupta, N. (2021). A Pre-Trained Vs Fine-Tuning Methodology in Transfer Learning, *Journal of Physics: Conference Series*, Vol. 1947, No. 1, 012028. doi:10.1088/1742-6596/1947/1/012028.
42. Llugsi, R., Yacoubi, S. El, Fontaine, A., and Lupera, P. (2021). Comparison between Adam, AdaMax and Adam W optimizers to implement a Weather Forecast based on Neural Networks for the Andean city of Quito, *2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM)*, IEEE, 1–6. doi:10.1109/ETCM53643.2021.9590681.
43. Petch, J., Di, S., and Nelson, W. (2022). Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology, *Canadian Journal of Cardiology*, Vol. 38, No. 2, 204–213. doi:10.1016/j.cjca.2021.09.004.

44. Xu, H., and Shuttleworth, K. M. J. (2024). Medical Artificial Intelligence and the Black Box Problem: A View Based on the Ethical Principle of "Do No Harm", *Intelligent Medicine*, Vol. 4, No. 1, 52–57. doi:10.1016/j.imed.2023.08.001.

45. Noviandy, T. R., Maulana, A., Khowarizmi, F., and Muchtar, K. (2023). Effect of CLAHE-based Enhancement on Bean Leaf Disease Classification through Explainable AI, *2023 IEEE 12th Global Conference on Consumer Electronics (GCCE)*, IEEE, 515–516. doi:10.1109/GCCE59613.2023.10315394.

46. Noviandy, T. R., Maulana, A., Emran, T. B., Idroes, G. M., and Idroes, R. (2023). QSAR Classification of Beta-Secretase 1 Inhibitor Activity in Alzheimer's Disease Using Ensemble Machine Learning Algorithms, *Heca Journal of Applied Sciences*, Vol. 1, No. 1, 1–7. doi:10.60084/hjas.v1i1.12.

47. Noviandy, T. R., Maulana, A., Idroes, G. M., Maulydia, N. B., Patwekar, M., Suhendra, R., and Idroes, R. (2023). Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery, *Malacca Pharmaceutics*, Vol. 1, No. 2, 48–54. doi:10.60084/mp.v1i2.60.

48. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). An Interpretable Machine Learning Strategy for Antimalarial Drug Discovery with LightGBM and SHAP, *Journal of Future Artificial Intelligence and Technologies*, Vol. 1, No. 2, 84–95. doi:10.62411/faith.2024-16.

49. Noviandy, T. R., Nisa, K., Idroes, G. M., Hardi, I., and Sasmita, N. R. (2024). Classifying Beta-Secretase 1 Inhibitor Activity for Alzheimer's Drug Discovery with LightGBM, *Journal of Computing Theories and Applications*, Vol. 2, No. 2, 138–147. doi:10.62411/jcta.10129.

50. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Machine Learning Approach to Predict AXL Kinase Inhibitor Activity for Cancer Drug Discovery Using XGBoost and Bayesian Optimization, *Journal of Soft Computing and Data Mining*, Vol. 5, No. 1, 46–56.