COMSERVA
JURNAL PENELITIAN DAN PENGABDIAN MASYARAKAT
Volume 4 No. 10 Februari 2025 (3732-XXX)
e-ISSN: 2798-5210
p-ISSN: 2798-5652

Development of Computational Models in Creating Skin Disease Classifiers Using CNN and ResNet Architecture

Elaine Faythe Hartono[1*], Putranegara Riauwindu[2]
University Pelita Harapan, Indonesia
*Email: Hartonoelainr@gmail.com
*Correspondence

ABSTRACT

This research paper delves into the exploration of various computational models utilized in creating skin disease classifiers, aiming to enhance diagnostic accuracy and optimize potential accessibility to isolated patients. By investigating the different types of computational models employed in this context, this paper aims to shed light on how these models contribute to more accessible and efficient diagnoses of skin conditions, enabling healthcare professionals to provide widespread patient care. Furthermore, the paper will address the challenges encountered in the development of effective skin disease classifiers using computational models, such as data quality issues, model interpretability, and generalizability across imbalanced datasets. Through a comprehensive analysis of these aspects, this research endeavors to advance our understanding of the potential of computational models in improving skin disease diagnosis and ultimately enhancing healthcare outcomes for individuals affected by dermatological conditions. Results showed that ResNet18 demonstrated higher overall accuracy on HAM10000 when compared to the confusion matrix and Grad-CAM visualizations of both models.

Kata kunci: Computational models; Skin disease classifiers; CNN; ResNet; Python.

INTRODUCTION

Skin diseases, such as acne, alopecia, bacterial skin infections, decubitus ulcers, fungal skin diseases, pruritus, psoriasis, scabies, urticaria, viral skin diseases, and skin cancer lesions, pose a significant public health concern worldwide, with recent years showing a rise in such diseases due to worsening living conditions and lack of access to healthcare (Yakupu et al., 2023). Skin diseases are the fourth most common cause of all human diseases, affecting nearly one-third of the world's population; however, their burden is often underestimated (Flohr & Hay, 2021). Skin and subcutaneous diseases lead to profound long-term alterations even after the disease has resolved, affecting not only the physical health but also the mental health and quality of life of the patient, placing a high burden on patients' families and national healthcare systems globally. The burden of skin conditions was high in both high- and low-income countries, indicating that prevention of skin diseases should be prioritized.

There are different types of skin lesions: Actinic keratosis (AK), Basal cell carcinoma (BCC), Benign keratosis (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic nevus (NV), Squamous cell carcinoma (SCC), and Vascular lesion (VASC), are shown in Fig. 1. The lesions are different in terms of their symptoms and severity. Some are permanent, and some are temporary and may be painless or painful. Among these skin diseases, melanoma is the most deadly and dangerous type. However, about 95% of skin disease patients can be recovered if identified at an initial state (Ahammed et al., 2022). Hence, widening the knowledge of skin disease epidemiology is critical for policy development and resource allocation, which ultimately leads to disease prevention (Yakupu et al., 2023).

However, in this digital age, the development of computational models has innovated the field of dermatology by offering new approaches to skin disease classification. Through the potential concept of integrating A.I with skin disease classification, the prevention of common skin diseases may be accomplished. Many researchers proposed several approaches for the classification of skin diseases. Related works are categorized into different types based on datasets, feature extraction techniques, feature selection techniques, and classification models. This section examines several relevant research articles to discover previous studies' tools and procedures and identify research gaps:

Jagdish (Jagdish et al., 2022) proposed a model for skin disease detection using image processing methods. They applied fuzzy clustering on 50 sample images with KNN and SVM classification algorithm with wavelet analysis. They showed that the K-Nearest Neighbor classification algorithm works well compared to the Support vector machine (SVM) classification technique with an accuracy of 91.2% and the algorithm identified the type of skin disease using classification methods. But they worked with only 50 sample images containing two classes (basal and squamous disease).

AlDera (AlDera & Othman, 2022) presented a skin disease diagnosis model that take an affected skin image and diagnose acne, cherry angioma, melanoma, and psoriasis. They applied Otsu's method for image segmentation and Gabor, Entropy and Sobel techniques for feature extraction on the dermnet NZ and atlas dermatologico. Finally, they applied K-Nearest Neighbor (K-NN) classifiers for classification, and achieved 67.1% accuracy.

Janney (Janney et al., 2018) applied the Support Vector Machine (SVM), the Artificial Neural Network (ANN), and the Naive Bayes Classifier. The features were extracted from the dermoscopic images using the Gray Level Co-occurrence Matrix (GLCM), texture, and wavelet features. Finally, the performance of the classification models was compared based on the accuracy, precision, and recall values. The system achieved an average accuracy of 89%, 71%, and 71% for ANN, SVM, and Naïve Bayes classifiers, respectively. But, the system has a poor performance on SVM and Naïve Bayes Classifiers. It is a vital drawback of the study.

Hameed (Hameed et al., 2018) presented a novel Computer-aided Diagnosis (CAD) system for diagnosing the most common skin lesions (acne, eczema, psoriasis, benign and malignant melanoma). The applied Dull Razor method, Gaussian filtering for preprocessing, and Otsu's thresholding method for extracting the region of interest from which the features are extracted. Several colors and texture features were extracted and applied to the Support Vector Machine (SVM) with the quadratic kernel. The experiment was performed on 1800 images, and 83% accuracy was achieved for six-class classification. However, the dataset was small, which is a limitation of the system.

Sinthura (Sinthura et al., 2020) proposed an advanced skin disease diagnosis model utilizing image processing. First, they applied Otsu's method to segment the disease portion. Next, some GLCM features such as area, perimeter, mean, and entropy were extracted. Finally, they applied the SVM classifier to identify the four classes of diseases: Acne, Psoriasis, Melanoma, and Rosacea. The method identifies skin diseases with an accuracy of 89%. They applied the model to a small dataset containing 100 images and used only a small number of features. These are the drawbacks of the system.

Ubale (Ubale & Paikrao, 2019) proposed a color phase model for detecting and classifying different skin diseases. They applied the HSV color phase model and LAB color phase model to extract features from the images. Finally, they employed the K Nearest Neighbor (KNN) classification method to classify skin diseases (Acne, Papillomas, Psoriasis, Melanoma, Mycosis, Vitiligo, Warts) that achieved 91.80% accuracy for HSV Color Phase Model and 81.60% accuracy for LAB Color Phase Model. Unfortunately, the preprocessing is not good enough, and there is no verified dataset used in this study.

From the above discussions, it can be concluded that most existing research has been conducted to detect skin diseases of a few classes. Most of the works are based on a single disease. Only a few that have been done are inadequate for classifying multiple classes. Most of them are performed on small size datasets. The preprocessing techniques are not good enough, and thresholding segmentation, edge-based segmentation, clustering-based segmentation are used in most of the studies. This study attempts to solve this limitation by examining the HAM10000 HAM10000 (“Human Against Machine with 10000 training images”) dataset, which consists of seven skin diseases with 10,015 images, and utilizes Convolutional Neural Networks (CNN) and an SVM model-Residual Networks (ResNet) to achieve its findings. CNNs are deep learning architectures that learn directly from data. CNNs are useful for detecting patterns in images and recognizing objects, classes, and categories (Ahammed et al., 2022). Residual Network (ResNet) is a deep learning model designed for computer vision applications. It is a CNN architecture capable of supporting hundreds or thousands of convolutional layers. ResNet uses "skip connections," which allows it to stack multiple identity mappings, skip those layers, and reuse the activations from the previous layer, resulting in improved performance for a larger number of layers [4].

METHOD

Data Collection and Analysis. The case data and images for skin lesions were collected from the HAM10000 using the Harvard Dataverse online query tool. The dataset comprises of 10,015 dermatoscopic images that was publicly released by the Harvard database in June 2018 to supply reliable training data for the automation of skin cancer lesion classification. In addition to the 10,015 images, a metadata file containing demographic information for each lesion was included, which were verified through histopathology (histo), while the ground truth for other cases was determined through follow-up examination (follow_up), expert consensus (consensus), or confirmation by in-vivo confocal microscopy (confocal). (AlDera & Othman, 2022) The selection of skin lesions and diseases reflects the prevalence, the case definition and the availability of the data used for the development of the models.

To begin, the HAM10000 Image Dataset was first imported into the code, which contains seven classes of skin cancer lesions: Melanocytic nevi, Melanoma, Benign keratosis-like lesions, Basal, cell carcinoma, Actinic keratoses, Vascular lesions, Dermatofibroma. As numerical representation is required to be usable for machine learning/artificial intelligence model, the classes must first be encoded into categorical outcomes.

Table 1
Statistical measures after applying Gaussian filtering

Lesion	Image	dx	dx_type	age	sex	localization	dataset	label
HAM_0004859	ISIC_0033653	mel	histo	70.0	female	neck	vidir_modern	4
HAM_0002856	ISIC_0028154	nv	histo	70.0	male	upper extremity	vidir_modern	5
HAM_0005112	ISIC_0024386	df	consensus	40.0	female	lower extremity	vidir_molemax	3
HAM_0005431	ISIC_0026413	nv	follow_up	65.0	male	lower extremity	vidir_molemax	5
HAM_0002796	ISIC_0025261	nv	histo	20.0	female	abdomen	vidir_modern	5
HAM_0002096	ISIC_0029518	bkl	consensus	50.0	male	face	vidir_modern	2
HAM_0002524	ISIC_0026184	nv	Follow_up	55.0	male	trunk	vidir_molemax	5

Standardization and Optimization of Dataset. To overcome the challenge of limited medical data and improve our model's learning, a technique called data standardization and data augmentation was used. Through PyTorch's function that combines multiple transformations into a single callable object, we adjusted the images' brightness, size, and colors using standard values from a large image database. Since all our images follow a similar pattern, the model will be able to learn at a faster rate with a simpler architecture. Next, data augmentation was applied to the training images. This involved creating variations of our images by flipping them horizontally and rotating them, which assists our model in becoming better at recognizing different skin conditions.

Other than augmenting and standardizing the image, the dataset was carefully organized into distinct groups to ensure effective learning. This process involves creating three essential sets: the 'train' set, the 'validation' set, and the 'test' set for evaluating its ability to recognize new, unseen images. By defining a custom sampler, StratifiedSampler, used for stratified sampling of data, it takes class_vector as input, which represents the class labels of the data. The test_size parameter determines the proportion of data to be held out for testing. The gen_sample_array method generates stratified train-test splits based on the class distribution. The sampler ensures that each split has approximately the same distribution of classes as the original data. It then splits the dataset into training, validation, and test sets using these generated indices. Finally, it prints the number of images in each set, fully rotated and loaded into memory.

RESULTS AND DISCUSSION

The results were measured using three parameters, which were using the validation accuracy, confusion matrix and Grad-Cam visualization. It can be concluded from the data collected that the SVM-ResNet18 produced higher overall accuracy on HAM10000 than CNN LeNet when used to create skin disease classifiers, making it a more ideal architecture to use when under budget or time constraints.

Accuracy and Training Loss. The results of both the CNN LeNet and ResNet 18 can be seen in Table 3. The two models' training and accuracy loss reveals the advantages and disadvantages of each model. The obtained overall accuracy surpasses chance, which is approximately 14% accuracy when randomly selecting a class out of the 7 available classes. This is calculated based on the expectation that random chance accuracy can be expressed as:

Accuracy by Chance = 1 / Number of Classes

In this case, with 7 classes, accuracy by chance would be 14.29%. Both models exceeded this and therefore fulfilled adequate benchmarks for AI learning. As the epochs increased, the respective accuracy rates increased proportionally, with SVM-ResNet showing a faster growth when provided the same amount of time and sample data.

Table 5
Performance metrics of our proposed models

Models	**Accuracy Rate**	
	10 Epochs	50 Epochs
CNN LeNet	33 %	54 %
SVM-ResNet18	67 %	79 %

Confusion Matrixes. The confusion matrices of the CNN LeNet and ResNet18 models can be seen in Figures 14 and 15. A confusion matrix is a table that summarizes the accuracy of the model's predictions, showing the number of true positive, true negative, false positive, and false negative predictions for each class. This gives a detailed breakdown of the model's classification accuracy and potential areas for misclassification across classes. The confusion matrix consists of diagonal and off-diagonal cells. The diagonal cells show the number of correct predictions in each class. The lighter or higher the value in these cells, the more accurately the model classifies instances from that class. Off-diagonal cells represent misclassification. Darker or lower values in these cells indicate instances where the model made errors [14]. To generate each confusion matrix, both CNN LeNet (net) and ResNet18 (net) are applied to a test dataset (*test_data_loader*), and the resulting predictions are compared with the actual labels.

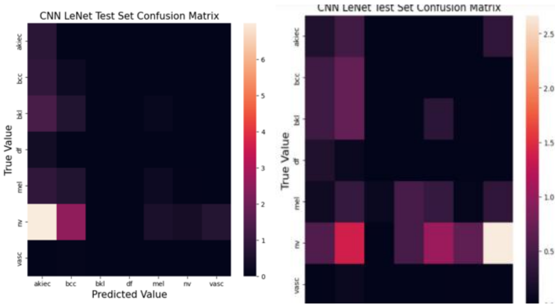

Figure 1
Confusion matrix of the CNN LeNet model

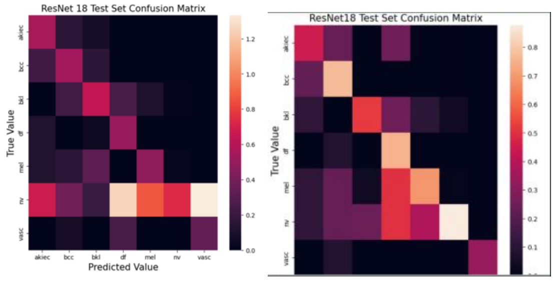

Figure 2
Confusion matrix of the ResNet18 model

Through the confusion matrix, it's clear that the ResNet18 model achieved more accurate results compared to its CNN LeNet model counterpart. Figure 14 shows that the CNN LeNet model achieved the highest performance value at identifying classes melanocytic nevi and basal cell carcinoma because they have the highest values in their diagonal cells, as indicated by their bright color in comparison to the rest of the classes. However, many of the off-diagonal cells are lower values, indicating that the model made mistakes. As illustrated in Figure 15, ResNet18 outperforms the initial CNN LeNet model, as evidenced by brighter diagonal cells. This demonstrates how the model more accurately classifies the image provided to it based on the actual conditions. The ResNet18 model achieved the highest performance in classifying the nv class, as evidenced by reaching the highest value. The brighter off-diagonal cells also indicate fewer mistakes were made in comparison to the CNN LeNet model, mel and bcc. Overall, this shows that the ResNet18 model works more efficiently with fewer errors compared to the CNN LeNet model when classifying skin conditions.

Grad-Cam Visualizations. The confusion matrices of the CNN LeNet and ResNet18 models can be seen in Figures 9 and 10. GradCAM is a visualization technique for interpreting the decision-making process of CNNs, particularly in image classification tasks. It generates heatmaps, highlighting regions of an input image that strongly influence the model's predictions. It helps highlight which features the model focuses on during classification. GradCAM enhances the explainability and trustworthiness of CNN-based image classification models [15].

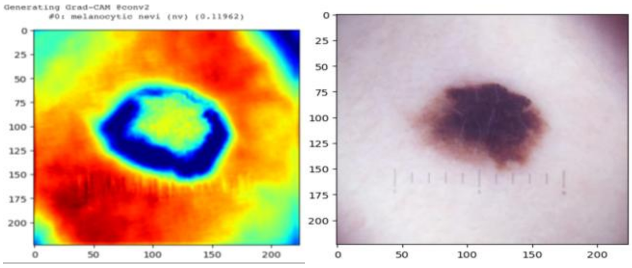

Figure 3
Grad-CAM visualizations of the CNN LeNet model

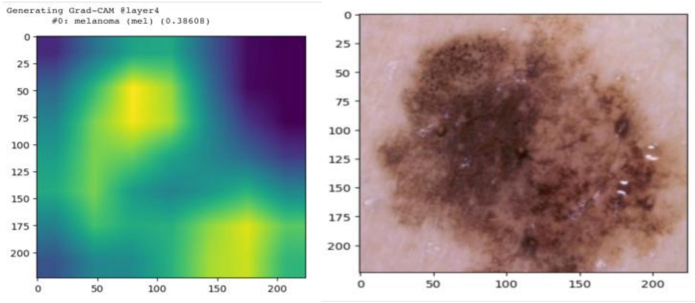

Figure 4
Grad-CAM visualizations of the ResNet18 model

The images on the left in both figures are Grad-CAM visualizations created for the specified target layer ("conv2") and target class. This visualization is created using the Grad-CAM technique on a heat map-style visual. It highlights the regions of the original image that made the most significant contribution to the model's prediction for the specified class. The warmer the color, the more important that region is to the model's ability to predict the image's classification.

The image on the right in both figures is the original image of the skin diseases used. Through comparing the ground truth label of the actual image with the prediction made in the Grad-CAM in both the CNN LeNet and ResNet18 models, both models can correctly predict the skin condition. As seen in both figures, the Grad-CAM can represent the shape of the skin conditions through the differences in colors. When comparing the Grad-CAM visualizations of both models, they performed similarly, as both models correctly predicted the skin conditions. However, ResNet18 was able to highlight the center parts of the image in warmer colors, showing that it can identify important ‘zones’ more efficiently and precisely, as seen in Figure 13. The CNN model on the other hand highlighted the surrounding skin, rather than the actual lesion, demonstrating a less precise identification. Therefore, ResNet18 has shown to have more accurate results overall on HAM10000.

Comparative Analysis. We have compared our result with some well-established recent methods including (Janney et al., 2018), (Hameed et al., 2018), (Sinthura et al., 2020) and (Ubale & Paikrao, 2019) shown in Table 5. Our technique performed on equal levels against these state-of-the-art methods but also required less time and costs. Complexity of a convolutional neural network model is O (k*n*d2), where, n is the sequence length, d is the representation dimension and k is the kernel size of convolution. On the other hand, we have used CNN and SVM. Each of these methods have less computational complexity than deep learning-based models. Moreover, our model has shown equal accuracy than the mentioned methods while spanning versatility.

Table 6
Comparative Analysis of the proposed model

Model	Dataset	Classifier	Classes	Accuracy
Proposed Method	HAM10000	CNN, SVM	7	54%, 79%*
Jagdish et al.	ISIC	KNN, SVM	2	91.2%

AlDera et al.	Other Datasets	CNN, SVM	4	90.7%, 84.2%, 67.1%
Janney et al.	ISIC	SVM, ANN, Naive Bayes	3	71%
Hameed et al.	ISIC	SVM	6	83%
Sinthura et al.	Unverified	SVM	4	89%
Ubale et al.	Unverified	KNN	7	91%

CONCLUSION

Skin disease is currently a global problem. People of many countries or regions suffer from different types of skin diseases. We can fight against these diseases by developing various techniques and processes. In this research, we have performed the work with several phases. We developed a skin classifier that combines traditional classifiers such as CNN and SVM, and programming by Python to help developers of all backgrounds potentially build upon it. We used a benchmark publicly available dataset, which is HAM10000. As, these datasets are somehow imbalanced, so we performed data balancing using random over sampling technique. We obtained an average accuracy of 79% and 54% for the HAM10000 dataset using SVM and CNN classifiers, respectively. It indicates that our code performs better using the SVM-ResNet18 model. We can use this model for other skin disease classification tasks.

But still, there is some scope to improve the classification performance. We had limited samples of medical data on the skin illnesses, which may lead to a narrower and less accurate scope of identifiable diseases by the training model. As seen in Figure 7 and Figure 9, both models experienced fluctuating rates of validation accuracy, despite the training loss decreasing. The lack of variety in the samples provided may have led to the models learning the samples, rather than the characteristics of identifying the illnesses. Furthermore, being able to achieve an adequate accuracy with the models would require a substantial amount of time. Hopefully, SVM-ResNet-18 architecture could be used more often as a learning model to assist healthcare professionals in a way that the public can always access anywhere (i.e. an app, etc.). This way, we can gain more information on skin diseases to tackle the public health concerns they cause worldwide. The code of this project is made public to encourage further improvements.

BIBLIOGRAPHY

Ahammed, M., Al Mamun, M., & Uddin, M. S. (2022). A machine learning approach for skin disease detection and classification using image segmentation. *Healthcare Analytics*, *2*, 100122.

AlDera, S. A., & Othman, M. T. Ben. (2022). A model for classification and diagnosis of skin disease using machine learning and image processing techniques. *International Journal of Advanced Computer Science and Applications*, *13*(5).

Flohr, C., & Hay, Rjbj. (2021). Putting the burden of skin diseases on the global map. In *British Journal of Dermatology* (Vol. 184, Issue 2, pp. 189–190). Blackwell Publishing Ltd Oxford, UK.

Hameed, N., Shabut, A. M., & Hossain, M. A. (2018). Multi-class skin diseases classification using deep convolutional neural network and support vector machine. *2018 12th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*, 1–7.

Jagdish, M., Gualán Guamangate, S. P., López, M. A. G., De La Cruz-Vargas, J. A., & Camacho, M. E. R. (2022). *Advance study of skin diseases detection using image processing methods.*

Janney, B. J., Roslin, S. E., & Shelcy, M. J. (2018). A comparative analysis of skin cancer detection based on svm, ann and naive bayes classifier. *2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE)*, 1694–1699.

Sinthura, S. S., Sharon, K. R., Bhavani, G., Mounika, L., & Joshika, B. (2020). Advanced skin diseases diagnosis leveraging image processing. *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 440–444.

Ubale, A. V, & Paikrao, P. L. (2019). Detection and classification of skin diseases using different color phase models. *International Research Journal of Engineering and Technology (IRJET)*, *6*(7).

Yakupu, A., Aimaier, R., Yuan, B., Chen, B., Cheng, J., Zhao, Y., Peng, Y., Dong, J., & Lu, S. (2023). The burden of skin and subcutaneous diseases: findings from the global burden of disease study 2019. *Frontiers in Public Health*, *11*, 1145513.