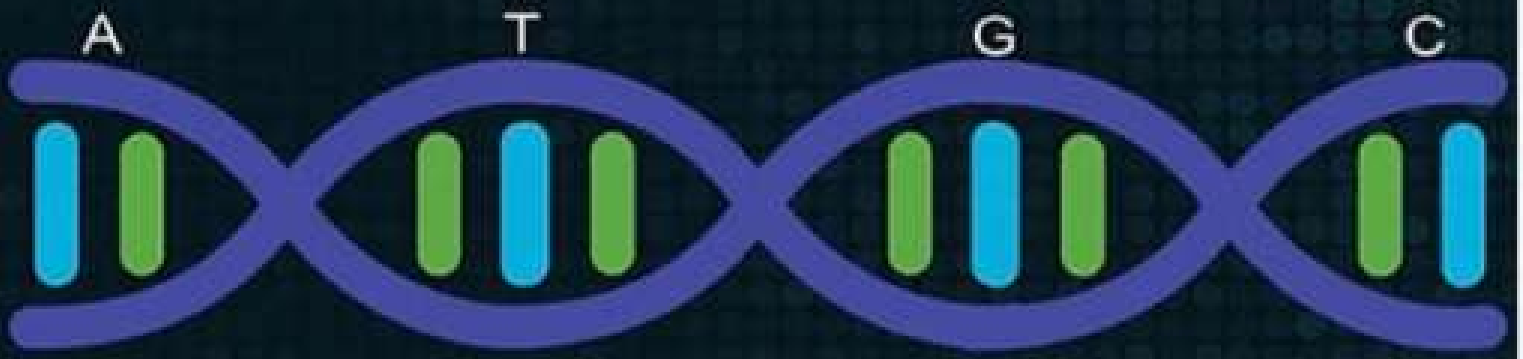
EDITED BY
SAURAV MALLIK, ZHONGMING ZHAO,
SOUMITA SETH, AIMIN LI,
KASMIKA BORAH AND
HIMANISH SHEKHAR DAS



FEATURE SELECTION AND FEATURE EXTRACTION ON OMICS DATA

A Chapman & Hall Book

CRC Press
Taylor & Francis Group

Designed cover image: Editor Ms. Kasmika Borah.

First edition published 2026
by CRC Press
2385 NW Executive Center Drive, Suite 320, Boca Raton FL 33431

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

ISBN: 978-1-032-96767-7 (hbk)
ISBN: 978-1-032-97000-4 (pbk)
ISBN: 978-1-003-59165-8 (ebk)

DOI: 10.1201/9781003591658

Typeset in Minion
by KnowledgeWorks Global Ltd.

CHAPTER 10

Analysis of Skin Diseases Using Deep Learning Techniques

Atikul Islam, Kalyani Mali, Mohit Kumar Halder, and Suvobrata Sarkar

10.1 INTRODUCTION

Millions of people around the world suffer from various skin diseases, ranging from mild infections to chronic and life-threatening conditions such as eczema, psoriasis, melanoma, and dermatitis [1]. Accurate diagnosis and classification are essential for effective treatment and management. There are, however, many traditional methods of diagnosing skin conditions that are subject to subjective judgement, are time-consuming, and can be inconclusive [2]. During the past few years, deep learning and machine learning have revolutionized medical image analysis, enabling automatic and highly accurate classification of diseases [3]. A neural network model known as a convolutional neural network (CNN) is particularly good at recognizing complex patterns in skin lesion images-sometimes more accurately than a dermatologist [4]. This study leverages a Skin Problem Dataset containing ten distinct categories of skin diseases collected from reliable sources. After preprocessing to enhance data quality, various deep learning models are applied to identify the most effective approach for disease classification. The main objective is to explore how deep learning can improve diagnostic accuracy and efficiency, ultimately facilitating early detection and better healthcare outcomes.

The process of classifying data points involves putting them in predefined categories based on their characteristics. The technology has been used for a wide range of applications, such as medical diagnosis, image recognition, natural language processing, and fraud detection [5]. In healthcare, classification techniques support disease identification, patient outcome prediction, and clinical decision-making [6]. There are a variety of traditional classification methods, including decision trees, support vector machines, and k-nearest neighbors (KNN) [7]. A CNN or recurrent neural network (RNN), however, offers superior performance when handling complex data like images and speech [8]. In this study, classification techniques are applied to the Skin Problem Dataset, with preprocessing and feature extraction steps improving model

DOI: 10.1201/9781003591658-10

accuracy and reliability. Pre-processing addresses data issues such as noise, missing values, and irrelevant features through techniques like normalization and transformation [9, 10]. Feature extraction further aids in identifying discriminative patterns essential for accurate classification [11, 12]. With CNNs, manual feature selection is minimized, as these networks learn hierarchical representations from images, capturing spatial and textural details automatically [13]. Exploratory Data Analysis (EDA) was conducted on the 6 GB dataset to understand its structure and distribution. Given its size and complexity, traditional machine learning methods showed limitations in efficiency and accuracy. Deep learning models, implemented using Visual Studio 2022, C programming, and Microsoft ML Builder Tool, significantly outperformed conventional approaches. According to the study, deep learning can be used to classify skin diseases in a reliable and scalable manner, allowing it to be applied to computer-aided dermatology and telemedicine systems.

10.2 LITERATURE SURVEY

The prevalence of skin diseases around the globe is among the highest, affecting millions of people across countless regions. According to the American Cancer Society [14], there were approximately 100,350 new cases of melanoma diagnosed in the US, resulting in 6,850 deaths. World Health Organization (WHO, 2001) data indicate that 20,000 deaths in Sub-Saharan Africa are caused by skin diseases [15]. In many developing regions, skin diseases are often transmissible and contagious, yet they remain treatable with proper medical intervention. In Western Ethiopia, 47% to 53% of respondents self-reported having a skin condition, but 67% of those who denied having a condition were in fact diagnosed with a treatable infection based upon clinical examination [16]. Likewise, a community-based survey in Sumatra, Indonesia, concluded that skin diseases were prevalent by 28%, indicating how widespread dermatological issues are. Several studies have examined how deep learning could improve the accuracy and timeliness of diagnosing and detecting skin diseases. According to Figueroa et al. [17] in their study, "The deadliest disease in the world is skin cancer, which can be classified using deep learning with a systematic review".

- **Imbalanced Training Data:** The number of images available for different skin diseases varies significantly, leading to biased model performance.
- **Cross-Domain Adaptability:** Models trained on specific datasets often struggle to generalize across different populations and imaging conditions, limiting their effectiveness in real-world applications.
- **Model Robustness:** The accuracy and reliability of deep learning models depend on their ability to cope with differences in image quality, lighting conditions, and skin tone.

To address these challenges, various deep learning methodologies have been employed. CNN-based architectures such as ResNet, DenseNet, and EfficientNet have been extensively used for skin disease classification, demonstrating promising results in improving diagnostic accuracy and reducing misclassification errors [17]. As a result of advancements in transfer learning and data augmentation, the model's ability to generalize across multiple datasets has been further improved over the past few years.

This literature survey highlights the growing impact of deep learning in automated skin disease classification, emphasizing its potential in aiding dermatological diagnosis and improving patient outcomes.

10.3 MATERIAL AND METHODS

10.3.1 Deep Neural Network

Deep neural networks (DNNs) extend artificial neural networks (ANNs) by incorporating additional hidden layers between input and output layers, thus increasing the depth of the model [18]. An important aspect of deep learning is that rather than explicitly programming algorithms, it optimizes performance through the process of learning. Using multiple hidden layers, deep learning models construct non-linear representations that enhance classification capabilities, particularly for complex data [19]. As data complexity grows, optimal classification may require the integration of multiple non-linear models. It is possible to achieve this by increasing the number of hidden layers, in order to create complex classification architectures that can deal with complex tasks [20].

It is demonstrated in this experiment that DNNs with sufficient hidden layers are capable of effectively classifying datasets with high non-linearity. The lowest loss indicates strong generalization capabilities. Future work may explore the effects of regularization techniques and alternative activation functions on classification performance. This type of data is more complex to classify. Therefore, achieving a model that accurately classifies it requires multiple hidden layers with appropriately adjusted weights. The general architecture of deep learning model as shown in Figure 10.1.

10.3.2 Understanding ResNet50 Architecture

A ResNet50 model consists of 48 convolutional layers, one using maximum pooling and another using average pooling. It performs 3.8×109 floating-point operations and is widely used in deep learning applications.

ResNet50's architecture is explored in depth due to its significant impact on computer vision. AlexNet won first place in the LSVRC2012 classification contest, marking a major milestone in deep learning. Following this, ResNet emerged as a groundbreaking development, demonstrating that ultradeep neural networks, with hundreds or even thousands of layers, could be effectively trained to achieve exceptional performance. The table illustrates several ResNet architectures, such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152. There is a description of each model's configuration of layers, including the convolutional layers, filter sizes, and feature map dimensions, as well as its computation complexity in Floating Point Operations (FLOPs). The architecture of ResNet50, as shown in Figure 10.2.

ResNet-18 and ResNet-34 utilize basic residual blocks, which involve fewer parameters and computations, making them more efficient for tasks requiring lower complexity [21]. In contrast, ResNet-50, ResNet-101, and ResNet-152 incorporate bottleneck layers, enabling the construction of deeper networks while maintaining computational feasibility [22].

In proportion to the depth of a network, FLOPs also increase, indicating a higher computational cost associated with deeper models. The table serves as a comprehensive

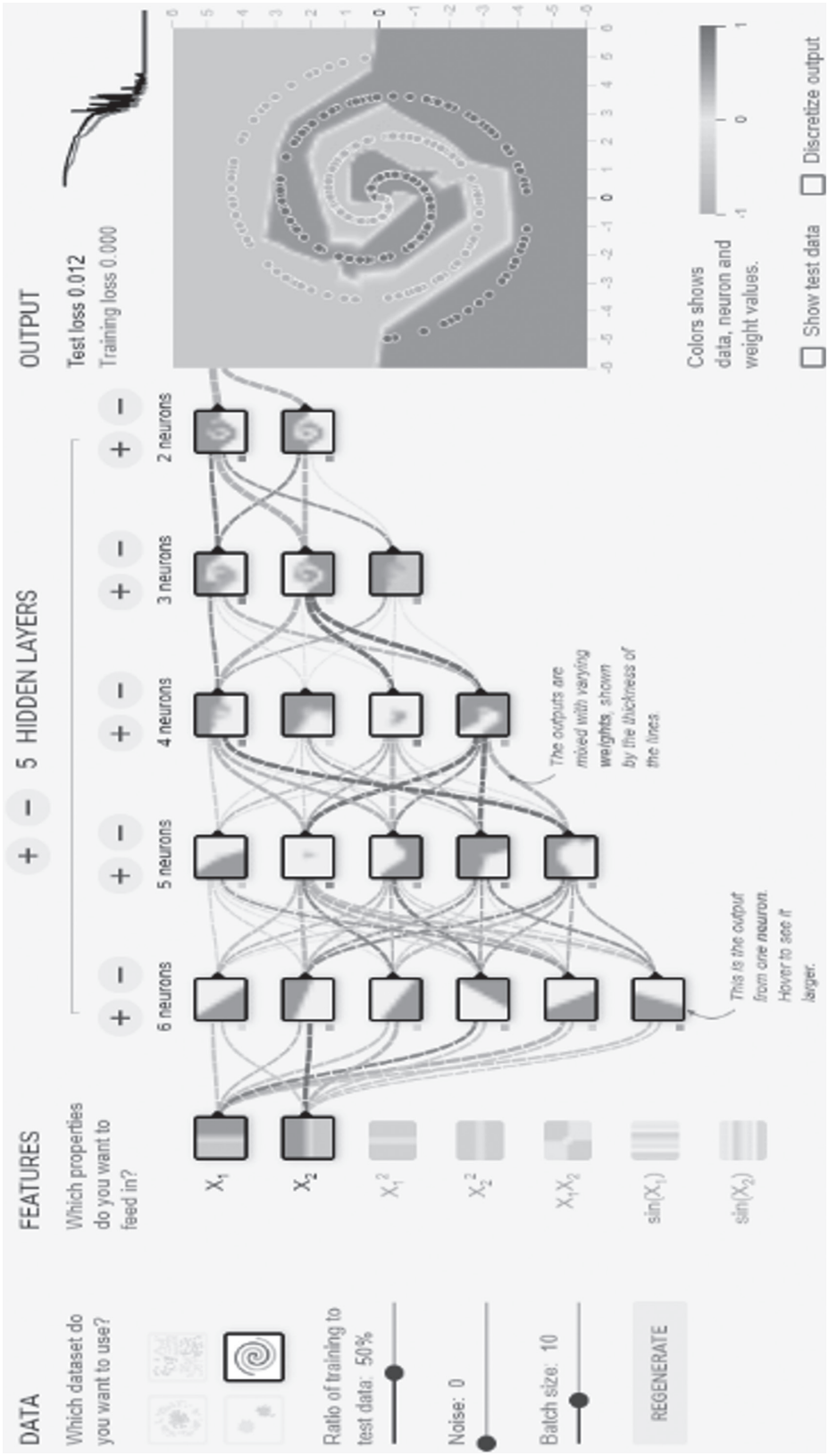


FIGURE 10.1 A deep learning model.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

FIGURE 10.2 Architecture of ResNet50.

guide for selecting an appropriate ResNet variant, balancing the trade-off between model accuracy and computational efficiency [23, 24].

10.4 PROPOSED METHOD

10.4.1 Dataset

First, we utilized the Skin Problem Dataset, which comprises ten distinct classes of images. A subsequent step was to enhance the data's quality by applying data preprocessing techniques. A subset of the preprocessed clean images is presented below. The dataset shown in the image as Figure 10.3 is a Skin Disease Classification Dataset, containing a total of

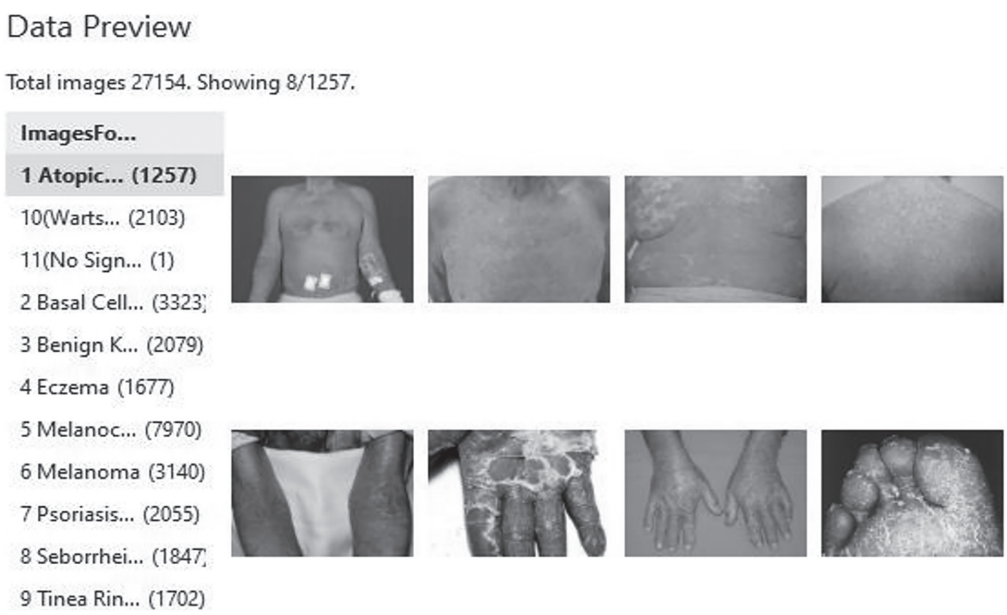


FIGURE 10.3 A sample data preview of the dataset.

27,154 images categorized into ten distinct skin disease classes. It provides a rich collection of dermatological images useful for developing machine learning models in medical image analysis. The preview displays a subset of 1,257 images from the dataset. The listed skin conditions include [25]:

- Atopic Dermatitis (1257 images)
- Warts (2103 images)
- Eczema (1677 images)
- Melanocytic Nevus (7970 images)
- Melanoma (3140 images)
- Benign Keratosis (2079 images)
- Psoriasis (2055 images)
- Seborrheic Keratosis (1847 images)
- Tinea Ringworm (1702 images)
- Basal Cell Carcinoma (3323 images)

This dataset is a valuable resource for deep learning and computer vision research in dermatology. Preprocessing techniques, including image augmentation, normalization, and noise reduction, are essential for optimizing model performance. CNNs and other deep learning architectures have proven to be effective in automating the diagnosis of skin conditions [26].

10.4.2 Proposed Method

We propose a deep learning-based approach combining DNNs and ResNet50 for accurate classification of skin diseases. The method leverages ResNet50's deep feature extraction capabilities alongside DNN's classification efficiency. The workflow from data preprocessing to classification is illustrated in Figure 10.4 using a visual representation. The image visually represents the proposed Deep Learning-based Skin Disease Classification workflow using DNN + ResNet50. The key steps in the diagram include:

1. Dataset Preparation:
   - Skin disease images are collected.
   - Data preprocessing techniques are applied to enhance image quality.
2. Model Selection:
   - ResNet50, a DNN, is chosen for feature extraction.
   - A DNN is integrated with ResNet50 for classification.

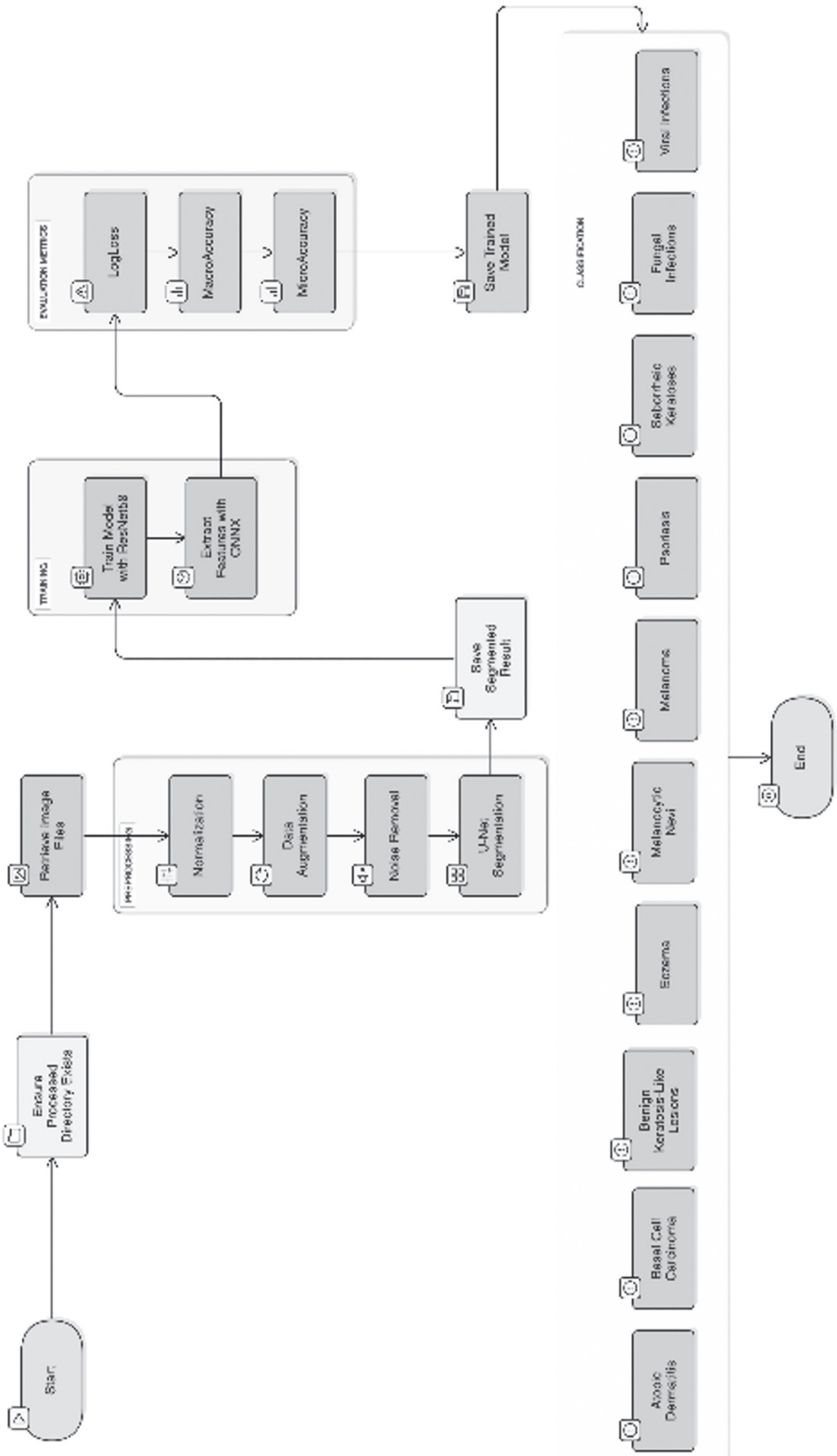


FIGURE 10.4 Workflow of the proposed method integrating ResNet50 and DNN for skin disease classification.

3. Training Phase:
   - The model undergoes backpropagation and optimization.
   - Normalization and augmentation techniques are applied to improve accuracy.
4. Evaluation and performance analysis:
   - The best accuracy achieved is 0.9205.
   - The training time recorded is 21,259 seconds.
5. Results and Predictions:
   - Classified images are categorized based on the model's predictions.
   - The final output provides detailed analysis and classification accuracy.

10.5 RESULT

The proposed model achieved an accuracy of 0.9205, demonstrating its effectiveness in classifying skin disease images. Additionally, the training time was recorded as 21,259 seconds, indicating the computational complexity of the model.

Initially, we utilized the given dataset and applied pre-processing techniques to enhance data quality. Various DNN models were then implemented and evaluated. Among them, the best-performing model was identified as DNN + ResNet50, achieving an accuracy of 0.9205 with a training time of 21,259 seconds. This approach yielded highly accurate results for the given inputs. A selection of the results obtained is presented below.

Figure 10.5 shows a skin lesion with a classification result on the right side. According to the model's prediction, the lesion has been identified as Melanocytic Nevus with 100% confidence. Other possible classifications, including Benign Keratosis, Melanoma, Basal Cell Carcinoma, and Seborrheic Keratosis, each have probabilities of less than 1%, indicating a very low likelihood of those conditions.

Figure 10.6 shows a skin lesion along with its classification results. The model predicts that the lesion is most likely a Melanocytic Nevus (82%), followed by Basal Cell Carcinoma

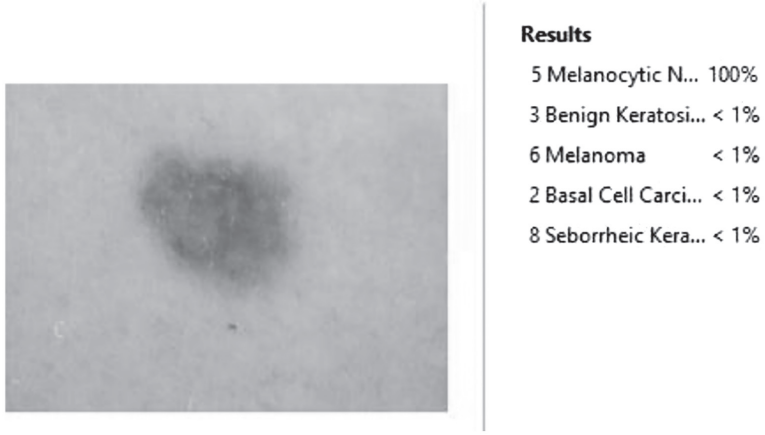


FIGURE 10.5 Image of a melanocytic disease.

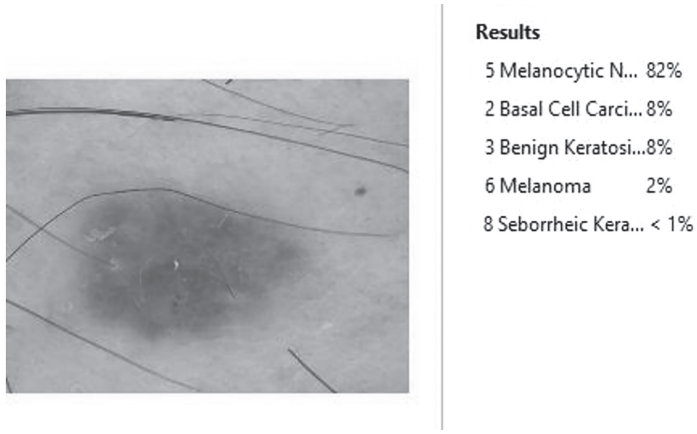


FIGURE 10.6 Image of a melanocytic disease with classification results.

(8%) and Benign Keratosis (8%). The probability of Melanoma is 2%, while Seborrheic Keratosis is less than 1%, indicating a low likelihood of malignancy.

Figure 10.7 shows a skin lesion with classification results. The model predicts a probability 96% of melanoma, indicating a high probability of malignancy. Furthermore, there is a probability 4% of Melanocytic Nevus, while Benign Keratosis, Basal Cell Carcinoma, and Seborrheic Keratosis each have less than a probability of 1%, suggesting that they are unlikely diagnoses.

The Figure 10.8 image shows a child's face with multiple skin lesions. The classification results indicate a 98% probability of Warts (Molluscum Contagiosum) as the most likely diagnosis. Furthermore, there is a probability 2% of atopic dermatitis, while Psoriasis, Eczema, and Tinea Ringworm each have less than 1%, making them less likely conditions.

A deep learning model was trained on a dataset consisting of ten distinct categories of skin diseases, resulting in this confusion matrix: Atopic Dermatitis, Basal Cell Carcinoma, Benign Keratosis-like Lesions, Eczema, Melanocytic Nevi, Melanoma, Psoriasis, Seborrheic Keratoses, Tinea Ringworm, and Warts/Molluscum as depicted in Figure 10.9. Matrix rows correspond to actual classes, while columns correspond to predicted classes. Incorrectly

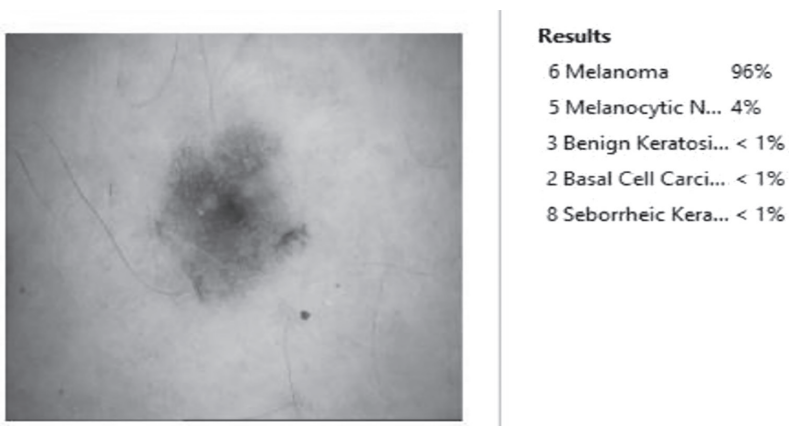


FIGURE 10.7 Image of a melanoma disease.

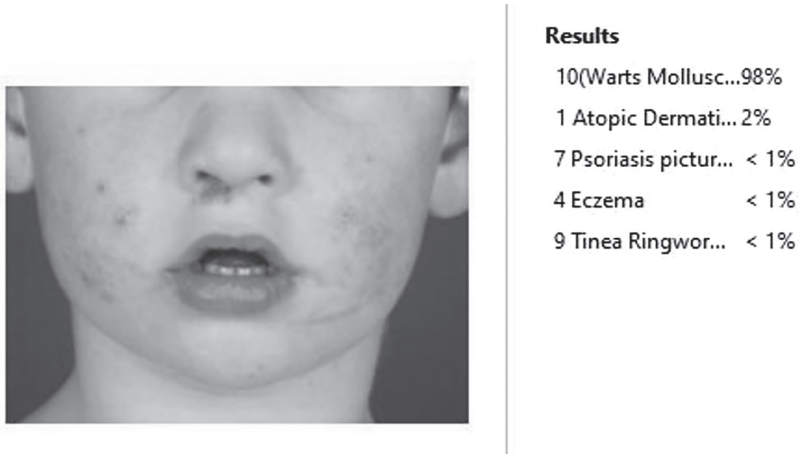


FIGURE 10.8 Image of a WartsMolluscum and other viral infections disease.

	AtopDemalitis	Basal Cell Carcinoma	Benlan Keratosis	Eczema	Melanonyic	Psoriasis	Seborrheic Keratosis	Tinea Ringwor
Atopic Dermat	179	0	0	3	2	1	5	50
Basal Cell Carcino	253	253	12	0	9	6	5	30
Eczema	36	260	190	1	11	1	0	30
Melano ma	4	1	1	1	290	290	7	28
Psorias	43	0	17	1	28	28	180	20
Sebor bheic K	24	0	13	0	0	32	189	20
Tinea Ringwo	48	1	1	1	29	22	180	20

FIGURE 10.9 Confusion matrix analysis for skin disease classification using deep learning.

classified instances are indicated by diagonal values, while misclassifications are highlighted by off-diagonal values.

Model	Accuracy
ResNet-52	92.05%
ResNet-32	74.00%
ResNet-18	71.00%

The performance of different deep residual networks was evaluated based on their classification accuracy. The results show that ResNet-52 achieved the highest accuracy of 92.05%, outperforming both ResNet-32, which scored 74.00%, and ResNet-18, which reached 71.00%. This comparison highlights the ability of a deeper network to learn more complex patterns and perform more accurately on the given skin lesion datasets. The improved accuracy of ResNet-52 suggests its greater depth and representational power may be more effective for this particular application.

10.6 DISCUSSION

Using deep learning techniques, especially DNN + ResNet50, to accurately classify skin diseases has been demonstrated in this study. A performance of 92.05% proved the model to be more accurate than traditional machine learning methods, indicating that deep learning can be used in dermatological diagnostics as well. A significant advantage of this approach is the utilization of ResNet50 for feature extraction, which enables the model to capture complex hierarchical patterns in skin disease images. DNNs improve classification accuracy and allow the model to distinguish between visually similar skin conditions when they are integrated into the model. They must, however, overcome a few challenges before they can be applied in real-world situations. The dataset used in this study exhibited an imbalance in class distribution, which may introduce bias in predictions. Future research should explore data augmentation or synthetic image generation to create a more balanced dataset.

Additionally, the computational complexity of training, which required 21,259 seconds, emphasizes the need for optimization, possibly by implementing lighter deep learning models without compromising accuracy. While the model performed well on the given dataset, its generalization to real-world scenarios requires further validation with diverse datasets encompassing various skin tones, imaging conditions, and demographic variations. In this study, deep learning is demonstrated to have potential in automating skin disease classification, especially for the application of telemedicine and computer-aided dermatology, offering a potentially valuable tool for early diagnosis. Future work could focus on incorporating explainable AI techniques, which would enhance model interpretability, making AI-driven dermatological tools more reliable and clinically acceptable.

10.7 CONCLUSION

The analysis of the Skin Problem Dataset is highly complex yet crucial for accurate skin disease prediction. By leveraging DNN + ResNet50, we can effectively classify newly provided skin disease images. As demonstrated in the previous section, this model has the

ability to correctly predict the class of inputs given. The DNN + ResNet50 methodology has proven to be a highly efficient deep learning technique for this task. The outcomes align with our expectations, reinforcing the reliability and effectiveness of this approach. Due to its superior performance, DNN + ResNet50 stands out as a preferable deep learning model for skin disease classification compared to other methods.

REFERENCES

1. Gupta, R., Tripathi, R., & Gupta, M. (2020). Recent advancements in skin disease classification using deep learning: A review. Biomedical Signal Processing and Control, 60, 101951.
2. Esteva, A., Kuprel, B., & Novoa, R. A. et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.
3. Litjens, G., Kooi, T., & Bejnordi, B. E. et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.
4. Brinker, T. J., Hekler, A., & Enk, A. H. et al. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. European Journal of Cancer, 113, 47–54.
5. Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Artificial Intelligence Review, 26(3), 159–190.
6. Sinz, C. et al. (2017). Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin. Journal of the American Academy of Dermatology, 77(6), 1100–1109..
7. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Elsevier.
8. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.
9. Garcıa, S., Luengo, J., & Herrera, F. (2015). Data Preprocessing in Data Mining. Springer International Publishing.
10. Pathan, S. et al. (2018). Techniques and algorithms for computer aided diagnosis of pigmented skin lesions? Review. Biomedical Signal Processing and Control. 39, 237–262.
11. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157–1182.
12. Mahbod, A. et al. (2019). Fusing fine-tuned deep features for skin lesion classification. Computerized Medical Imaging and Graphics, 71, 19–29..
13. Badrinarayanan, V. et al. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495.
14. American Cancer Society (2020). Skin Cancer Facts and Figures.
15. Mathers, C. D. (2006). Global Burden of Disease 2001. Data Sources.
16. Methods and Results. World Health Organization (WHO).
17. Figueroa, J. I., Fuller, L. C., Abraha, A., & Hay, R. J. (1998). Dermatology in southwestern Ethiopia: Rationale for a community approach. International Journal of Dermatology, 37(10), 752–758.
18. Wu, B., Chen, B., Zeng, A., & Pan, D. (2021). Skin problem classification with deep learning: A systematic review. Journal of Medical Imaging and Health Informatics, 11(2), 456–472.
19. An, Z., Pan, D., & Zeng, A. (2021). Deep learning-based skin disease classification: Challenges and solutions. Biomedical Signal Processing and Control, 68, 102731.
20. Popescu, D. M., Shade, J. K., Lai, C., Aronis, K. N., Ouyang, D., Moorthy, M. V., Cook, N. R., Lee, D. C., Kadish, A., Albert, C. M., Wu, K. C., Maggioni, M., & Trayanova, N. A. (2022). Arrhythmic sudden death survival prediction using deep learning analysis of scarring in the heart. Nature Cardiovascular Research, 1(4), 334–343. https://doi.org/10.1038/s44161-022-00041-9

21. Choy, B., Greenfield, S., & Saffari, A. (2023). Deep learning models across the range of skin disease: A systematic review. NPJ Digital Medicine, 6, Article 33. https://doi.org/10.1038/s41746-023-00914-8
22. Aellen, F. M., Alnes, S. L., Loosli, F., Rossetti, A. O., Zubler, F., De Lucia, M., & Tzovara, A. (2023). Auditory stimulation and deep learning predict awakening from coma after cardiac arrest. Brain, 146(2), 778–788. https://doi.org/10.1093/brain/awac340
23. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778).
24. Goyal, M. et al. (2020). Artificial intelligence-based image classification for diagnosis of skin cancer: Challenges and opportunities. Computers in Biology and Medicine, 127, 104065.
25. Ismailpromus, Skin Diseases Image Dataset, Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/ismailpromus/skin-diseases-image-dataset. [Accessed: 01-Jan-2025].
26. Pathan, S. et al. (2018). Techniques and algorithms for computer-aided diagnosis of pigmented skin lesions: A review. Biomedical Signal Processing and Control, 46, 1–12.