WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

Skin Disease Detection using Deep Learning Techniques

B. Yaswanth, B. Vinay, Ch. Dhana Sri, K. Akhil, Dr. T. Geetamma, Dr. S. Shanamthi

GMR Institute of Technology Deemed to be University

A B S T R A C T

Skin diseases vary widely in appearance, making accurate classification from clinical images challenging. Conventional diagnosis relies primarily on clinical expertise and visual examination, which can be time-consuming and may be affected by similarities among different disease conditions. This study presents a deep learning-based framework for automated skin disease classification using transfer learning. Different convolution neural network architectures, namely EfficientNetV2B0, MobileNetV3-Large, ResNet50, and DenseNet121, SE-ResNet50 are investigated and comparatively evaluated under a common experimental framework. The proposed approach involves image pre processing, data augmentation, transfer learning, model training, and performance evaluation across multiple classification metrics. The pre trained models are adapted to extract different visual features from skin images and classify them into the corresponding disease categories. Model performance is assessed using accuracy, precision, recall, F1-score, and confusion matrix analysis. The comparative evaluation provides insights into the effectiveness of different CNN architectures for skin disease classification and identifies the architecture that provides the most suitable performance for the considered data-set. The proposed framework demonstrates the potential of transfer learning as a computer-aided approach for supporting preliminary skin disease identification, while emphasizing that automated predictions should be complemented rather than replacing professional clinical assessment.

Keywords: Automated skin disease classification, deep learning, transfer learning, convolution neural networks, medical image analysis, EfficientNetV2B0, comparative model analysis.

Introduction

Dermatological conditions represent prevalent health concerns affecting individuals throughout various age groups and demographic populations. These conditions may develop as a consequence of multiple underlying causes, encompassing infectious agents, hypersensitivity reactions, climatic influences, hereditary predispositions, and sustained contact with external substances. Precise classification of dermatological conditions is essential because comparable visual presentations may correspond to distinct pathological entities necessitating divergent therapeutic approaches. Traditional diagnostic methodology predominantly depends on clinical observation and specialist expertise, which may prove inadequate when confronted with morphologically indistinguishable disease presentations, insufficient access to qualified professionals, or substantial patient volumes.

The expanding accessibility of medical imaging repositories and technological progress in machine learning have facilitated the development of automated diagnostic systems for disease categorization. Notably, deep learning methodologies have demonstrated substantial proficiency in identifying significant patterns within image data. Convolutional neural networks (CNNs) possess the capacity to autonomously identify hierarchical characteristics including contours, surface properties, geometric configurations, and sophisticated visual attributes, thereby diminishing reliance upon manually constructed feature descriptors. These capabilities position CNN-based methodologies as particularly advantageous for image-centered dermatological assessment.

Nevertheless, constructing a comprehensive deep neural network typically necessitates substantial quantities of annotated training data and intensive computational infrastructure. Transfer learning constitutes a pragmatic solution by leveraging feature knowledge acquired from extensive image collections and customizing such knowledge for a designated classification objective. Distinct CNN frameworks incorporate varying mechanisms for feature identification and architectural configurations; consequently, their effectiveness fluctuates according to dataset characteristics and the nature of the classification assignment.

Related Work

Deep learning has emerged as a significant methodology for the automated examination of dermatological imagery, as convolutional neural networks (CNNs) possess the capability to acquire discriminative visual representations directly from image data. In contrast to earlier computer-aided methodologies that predominantly depended on manually engineered features and conventional classifiers, deep learning facilitates hierarchical feature extraction and comprehensive end-to-end learning. The growing accessibility of clinical and dermoscopic skin-image datasets has substantially promoted the advancement of CNN-based techniques for automated skin disease categorization.

Ahmad et al. [1] introduced a discriminative feature learning framework designed for skin disease categorization utilizing transfer learning methodology. Their approach involved fine-tuning pre-trained ResNet152 and InceptionResNet-V2 architectures while integrating a triplet loss function to acquire discriminative image embeddings. Rather than depending exclusively on conventional classification loss, the methodology projects images into a Euclidean feature space and promotes images corresponding to identical disease classifications to maintain reduced distances while maintaining separation between images from distinct categories. The investigation examined facial skin disease imagery and demonstrated that discriminative feature learning could enhance classification outcomes. The researchers additionally implemented layer-wise fine-tuning of pre-trained architectures, underscoring the significance of transfer learning when annotated medical-image datasets are constrained.

Wu et al. [2] conducted an investigation into the performance of various CNN architectures applied to facial skin disease classification utilizing clinical imagery. The researchers constructed a dataset comprising 2,656 facial images depicting six prevalent skin conditions and assessed five leading CNN architectures. Their experimental framework additionally incorporated an independent dataset sourced from alternative anatomical regions to examine the efficacy of transfer learning. The findings demonstrated that transfer learning enhanced mean precision and recall across the majority of examined architectures, underscoring the significance of pretrained feature representations in clinical dermatological image classification. This research holds considerable importance for comparative deep learning investigations as it illustrates that architectural design influences classification outcomes and that systematic evaluation across multiple CNN models yields valuable guidance for architecture selection.

Li et al. [3] tackled a distinct obstacle pertaining to rare skin disease classification, wherein a restricted quantity of annotated samples may exist for particular disease categories. The proposed Dynamic Subcluster-Aware Network (SCAN) employs a dual-branch structure to simultaneously acquire class-level representations and implicit subcluster configurations within disease categories. This methodology is grounded in the recognition that images corresponding to identical skin diseases frequently demonstrate considerable heterogeneity within the same class. A clustering-oriented loss function and a purity loss mechanism are integrated to enhance the quality of learned representations. The technique underwent evaluation on the SD-198 and Derm7pt datasets within few-shot learning frameworks and demonstrated enhancements of approximately 2% and 5% relative to previously established optimal methods on the respective datasets across multiple performance metrics.

Khalaf et al. [5] conducted an extensive examination of deep learning methodologies employed in skin cancer segmentation and classification. This investigation encompassed CNN-based techniques, transfer learning applications, data preprocessing strategies, composite architectural designs, available datasets, and obstacles related to clinical implementation. The examination identified multiple persistent challenges, including insufficient dataset diversity, unequal class distribution, limited model interpretability, elevated computational demands, inadequate generalization capabilities, and barriers to real-world clinical integration. Furthermore, the review underscored the necessity of constructing more inclusive datasets and enhancing the effectiveness and dependability of deep learning systems to facilitate broader clinical acceptance.

The examined literature indicates that skin disease classification has evolved from traditional feature-extraction methodologies toward progressively advanced deep learning frameworks. Targeted methodologies including discriminative metric learning, few-shot learning, feature combination, and attention-based mechanisms can effectively mitigate specific obstacles such as restricted training data, heterogeneous intra-class characteristics, resource limitations, and enhanced feature differentiation. Nevertheless, these methodologies simultaneously introduce heightened structural intricacy and are typically assessed utilizing disparate datasets and distinct experimental protocols, thereby complicating direct architectural evaluation. Concurrently, contemporary research continues to recognize dataset heterogeneity, model generalization, computational efficiency, and system interpretability as critical impediments to successful practical implementation.

Driven by these findings, the current investigation undertakes a systematic comparison of five transfer-learning-based convolutional neural network architectures: EfficientNetV2B0, MobileNetV3-Large, ResNet50, SE-ResNet50 and DenseNet121. Rather than proposing an additional bespoke network architecture, this research assesses well-established architectures within a unified experimental framework employing identical datasets, preprocessing methodologies, and assessment metrics. The comparative investigation aims to determine the comparative performance of the examined models for the specified skin disease classification application and to establish a foundation for determining the most suitable architecture based on classification accuracy and computational efficiency considerations.

Methodology

This investigation presents a systematic comparison of four convolutional neural network (CNN) architectures designed for skin disease classification. The examined models comprise EfficientNetV2B0, MobileNetV3-Large, ResNet50, SE-ResNet50, and DenseNet121. The methodological framework encompasses dataset preparation, image preprocessing, model architecture selection, transfer learning, model training, and performance evaluation. All models undergo assessment utilizing an identical classification dataset and standardized evaluation protocols to determine their efficacy for the specified application. The comprehensive procedural framework of the proposed methodology is depicted in Fig. 1.

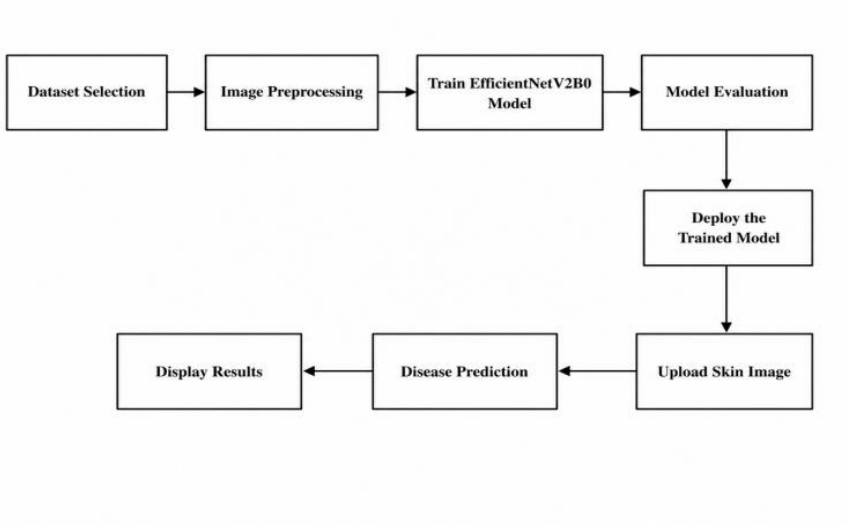


Fig.1. *Block diagram of the proposed skin disease classification system using EfficientNetV2B0.*

Dataset Description

The dataset employed in this investigation was procured from Kaggle and encompasses dermatological disease imagery accompanied by corresponding categorical annotations. The dataset annotations are maintained within a CSV file, which facilitates the association between images and their designated labels. The available categorical designations and the quantity of images corresponding to each category are ascertained during the dataset preparation phase.

Prior to model training, the dataset undergoes comprehensive examination to confirm image accessibility, label uniformity, and the proportional representation of samples across all categories. The images are subsequently partitioned into training, validation, and testing subsets according to the segmentation methodology implemented in the experimental design. The training subset serves to optimize model parameters, the validation subset functions to assess model performance throughout the training process, and the test subset is designated for comprehensive final assessment. The quantity of categories, aggregate image quantity, and subset allocation ratios are documented based on the empirical dataset and corresponding implementation. Illustrative examples from the dataset are presented in Fig. 2.



Fig. 2. *Representative images illustrating the eight skin disease classes considered in the proposed classification system: (1) Acne and Rosacea, (2) Atopic Dermatitis, (3) Eczema, (4) Nail Fungus and Other Nail Disease, (5) Seborrheic Keratoses and Other Benign Tumors, (6) Vascular Tumors, (7) Vasculitis, and (8) Warts, Molluscum and Other Viral Infections.*

Image Pre-Processing

Image preprocessing constitutes the foundational step that prepares the dataset for integration into the designated CNN architectures. All images undergo resizing to a standardized dimension of 300 × 300 pixels, ensuring uniform spatial characteristics throughout the experimental framework. The images are maintained as three-channel colour inputs.

Pixel values are processed in accordance with the specific input requirements of each selected architecture. Notably, the appropriate preprocessing methodology is implemented for the corresponding pretrained model rather than presuming that all four models necessitate identical input normalization. This distinction is critical when employing pretrained CNNs, as their anticipated input representations may vary substantially.

Data augmentation is implemented on the training images exclusively when activated within the implementation parameters. The augmentation operations employed in the experiment are systematically documented to establish a clear distinction between the original images and their transformed counterparts. The validation and test images are processed uniformly without the application of random training augmentation, thereby ensuring that model evaluation remains uncompromised by randomly transformed samples. The preprocessing workflow is presented in Fig. 3 should an augmentation or preprocessing diagram be included.

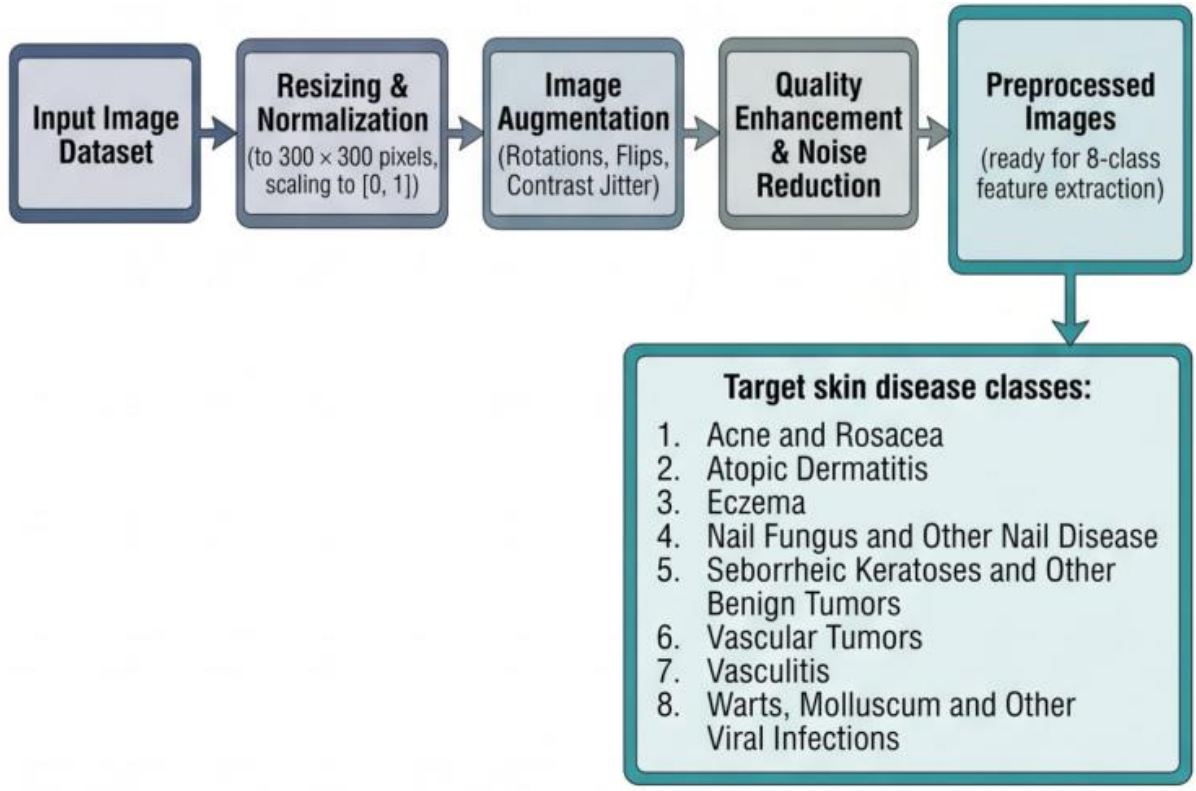


Fig. 3. *Image Preprocessing workflow designed to optimize clinical images of various dermatological conditions for robust 8 - class deep learning classification.*

CNN Architectures Used in the Study

Four convolutional neural network architectures have been selected to examine how distinct feature-extraction methodologies influence the classification of skin diseases. Each architecture accepts an input image and develops visual representations that facilitate categorization into the designated disease classifications.

1) EfficientNetV2B0

EfficientNetV2B0 represents a convolutional neural network engineered to achieve equilibrium between computational efficiency and feature-learning performance. The architecture integrates fused mobile inverted bottleneck convolution (Fused-MBConv) and mobile inverted bottleneck convolution (MBConv) components throughout its feature-extraction phases. These components sequentially extract visual features across multiple levels of conceptual abstraction. Within this investigation, EfficientNetV2B0 is assessed as a prospective architecture for deriving discriminative features from dermatological disease imagery.

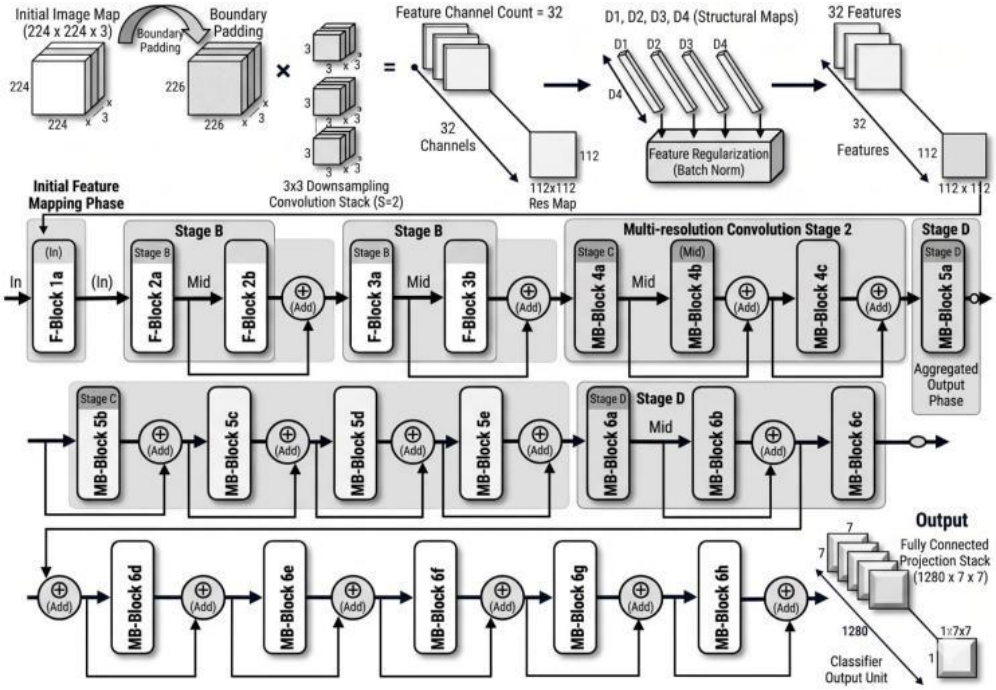


Fig. 4. *Architecture of EfficientNetV2-B0 model used for skin disease classification.*

2) MobileNetV3 - Large

MobileNetV3-Large is a streamlined convolutional neural network crafted to strike a balance between classification accuracy and computational efficiency. It employs depthwise separable convolutions along with inverted residual bottleneck blocks to minimize computational complexity while effectively extracting significant image features. Additionally, the architecture integrates Squeeze-and-Excitation (SE) modules in specific blocks for channel-wise feature recalibration and utilizes the hard-swish activation function to enhance feature representation. In this research, MobileNetV3-Large serves as a backbone for feature extraction, with its classification layer modified to predict eight categories of skin diseases.

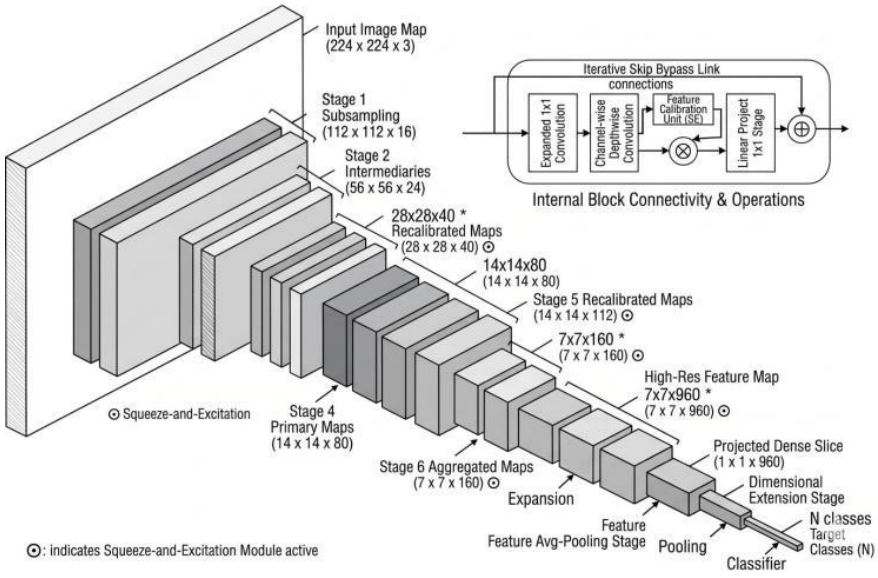


Fig. 5. *Architecture of MobileNetV3 - Large Model used for classification.*

3) ResNet50:

ResNet50 represents a profound convolutional neural network architecture that employs bypass connections to enhance the acquisition of visual feature representations across successive convolutional strata. The residual bottleneck components within this architecture enable the network to progressively capture increasingly sophisticated features while facilitating the transmission of information and gradients throughout the entire network structure. ResNet50 has been selected to assess the classification efficacy of a residual-learning framework when applied to the designated dataset.

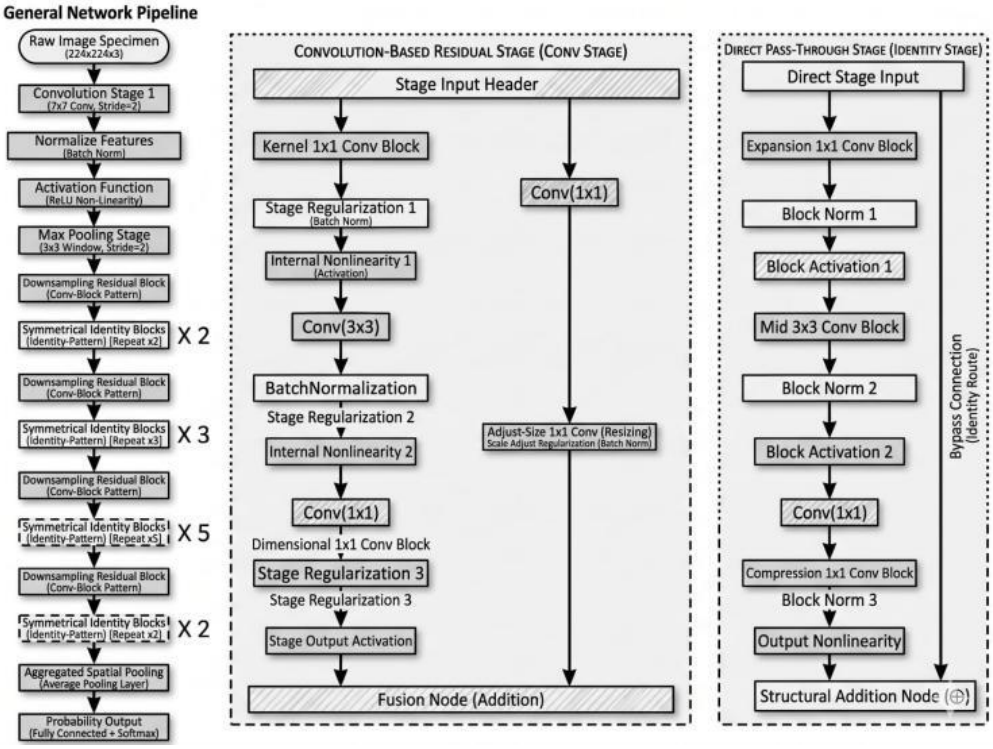


Fig. 6. *Architecture of ResNet50 Model based on residual learning.*

4) DenseNet121

DenseNet121 implements dense interconnectivity patterns within its dense blocks, wherein feature maps originating from antecedent layers are transmitted to subsequent layers contained within the identical block. This architectural configuration promotes the reuse of extracted features and facilitates comprehensive information dissemination across the network. DenseNet121 is incorporated to evaluate the performance characteristics of its densely interconnected feature-extraction methodology in relation to the alternative selected network architectures.

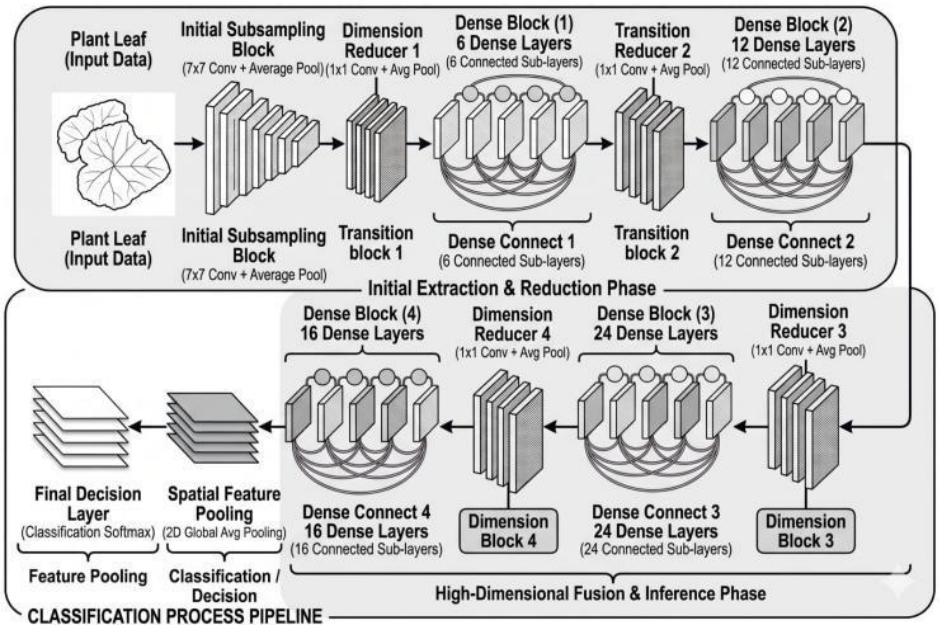


Fig. 7. *Architecture of DenseNet121 model illustrating dense feature connections.*

5) SE - ResNet50

SE-ResNet50 represents an advanced iteration of the ResNet50 architecture, integrating Squeeze-and-Excitation (SE) blocks to enhance feature representation. This model employs residual learning to facilitate the effective training of a deep network, while the SE blocks dynamically adjust channel-wise feature responses by allocating higher weights to the more informative channels. Each SE block executes global average pooling, followed by fully

connected layers and a sigmoid activation function to produce channel-wise weights. In this research, SE-ResNet50 is utilized to extract discriminative features from images of skin diseases and categorize them into eight distinct disease classification.

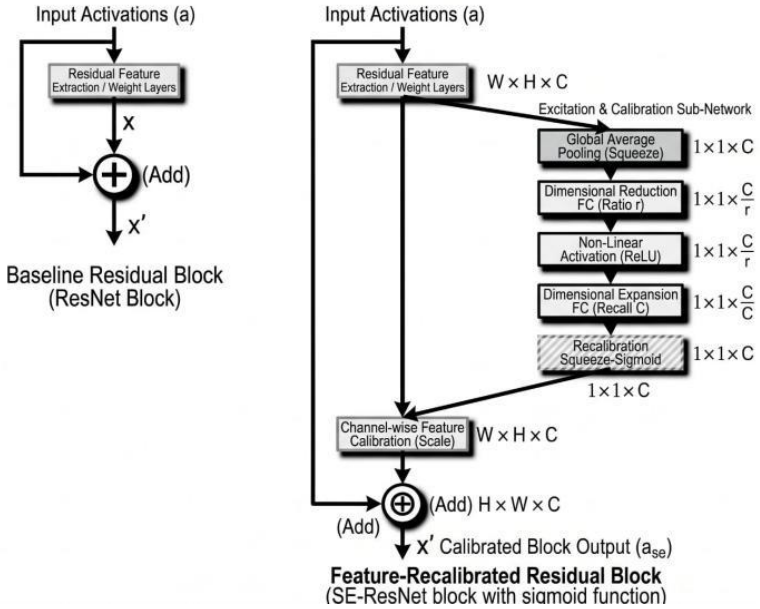


Fig. 8. *Architecture of SE- ResNet50 model incorporating squeeze and excitation blocks.*

Transfer Learning and Classification

The four selected architectures undergo adaptation to the target skin disease classification task through a transfer-learning methodology employed in this experiment. When pretrained ImageNet weights are utilized, the models initialize with features derived from a large-scale image dataset and subsequently adjust to accommodate skin disease imagery.

The original classification output layer is modified or substituted as necessary to enable the model to predict the classes represented in the target dataset. The resulting models generate class scores or probability distributions for each input image, from which the final predicted class is derived. The particular classification head configuration, pooling mechanism, trainable layer selection, and fine-tuning approach are contingent upon the implemented model architecture and are documented in detail.

The four models are trained and assessed independently. Their respective model-specific input preprocessing specifications are preserved, whereas the dataset and evaluation metrics remain standardized to facilitate a rigorous and meaningful comparison.

Model Training

The models were developed utilizing Python in conjunction with the selected deep learning framework for this investigation. The input images were standardized to dimensions of 300 × 300 pixels, with a batch size configured at 32 samples. All additional training hyperparameters—encompassing the optimization algorithm, learning rate, objective function, epoch count, and any adaptive learning rate schedules or premature termination mechanisms—were established in accordance with the designated training specifications.

Throughout the training phase, model weights were refined through application of the training dataset and the designated loss function. Validation metrics were continuously evaluated to track model progression and, when necessary, facilitate informed decisions regarding model selection or parameter adjustment. The test dataset remained isolated from both model optimization and validation-informed determinations, serving exclusively for comprehensive performance evaluation.

This standardized experimental methodology was consistently applied across EfficientNetV2B0, MobileNetV3-Large, ResNet50, SE-ResNet50 and DenseNet121 architectures. Model-specific variations in data preprocessing procedures, layer trainability configurations, or optimization methodologies were systematically documented to ensure accurate interpretation of comparative findings. The definitive configuration parameters for each model architecture are presented within the experimental setup section.

Performance Evaluation

The trained models undergo evaluation utilizing the held-out test subset. Classification performance is assessed through the application of accuracy, precision, recall, and F1-score metrics. These metrics deliver complementary insights regarding overall correctness and the recognition of distinct disease categories.

Accuracy quantifies the proportion of correctly classified samples relative to all evaluated samples. Precision quantifies the proportion of accurate predictions among samples classified as belonging to a particular class. Recall quantifies the proportion of actual class samples that are successfully identified. The F1-score represents the harmonic mean of precision and recall.

In multiclass classification scenarios, the averaging methodology employed for calculating precision, recall, and F1-score is explicitly documented within the experimental framework. A confusion matrix is additionally constructed to analyze the distribution of accurate predictions and misclassifications across the designated classes.

The outcomes derived from the five architectures are systematically compared utilizing an identical test set and standardized reporting metrics. Training and validation curves may further be employed to investigate the learning dynamics exhibited by the models. The quantitative findings and their analytical interpretation are presented within the Results and Discussion section.

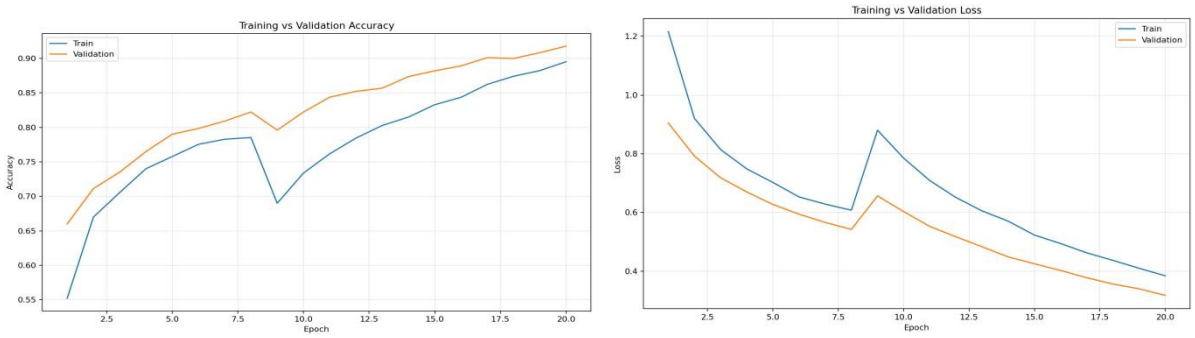
Results and Discussion

The effectiveness of the five chosen deep learning architectures was assessed through training and validation curves, confusion matrices, and quantitative classification metrics. The models evaluated include EfficientNetV2-B0 (Pretrained), ResNet50, DenseNet121, SE-ResNet50, and MobileNetV3-Large. Their classification performance was measured using test accuracy, macro precision, macro recall, macro F1-score, and weighted F1-score.

Training And Validation Performance

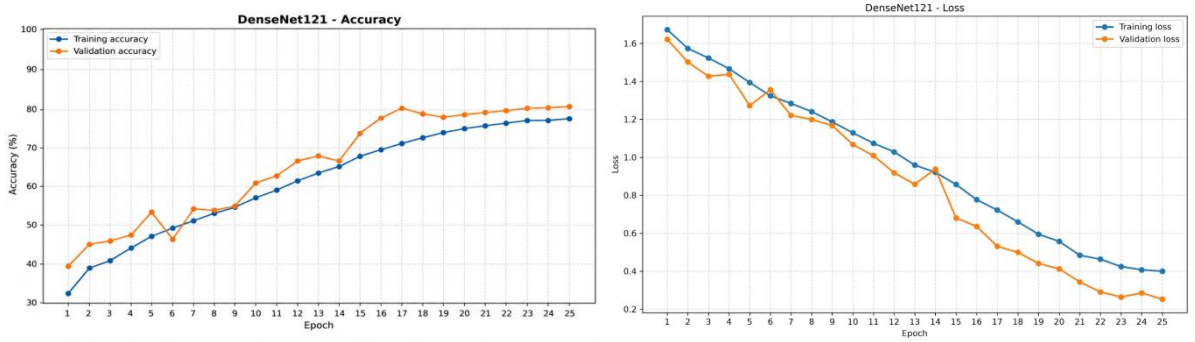
The training and validation accuracy and loss curves were examined to gain insights into the learning behavior of each model throughout the training epochs. The accuracy curves depict the fluctuations in classification accuracy during both training and validation phases, whereas the loss curves reflect the changes in prediction error.

Figure 1 displays the training and validation accuracy and loss curves for EfficientNetV2-B0 (Pretrained). Figures 2 through 5 illustrate the corresponding curves for ResNet50, DenseNet121, SE-ResNet50, and MobileNetV3-Large, respectively. These visual representations facilitate a comparative analysis of the learning patterns and convergence behaviors of the chosen architectures.



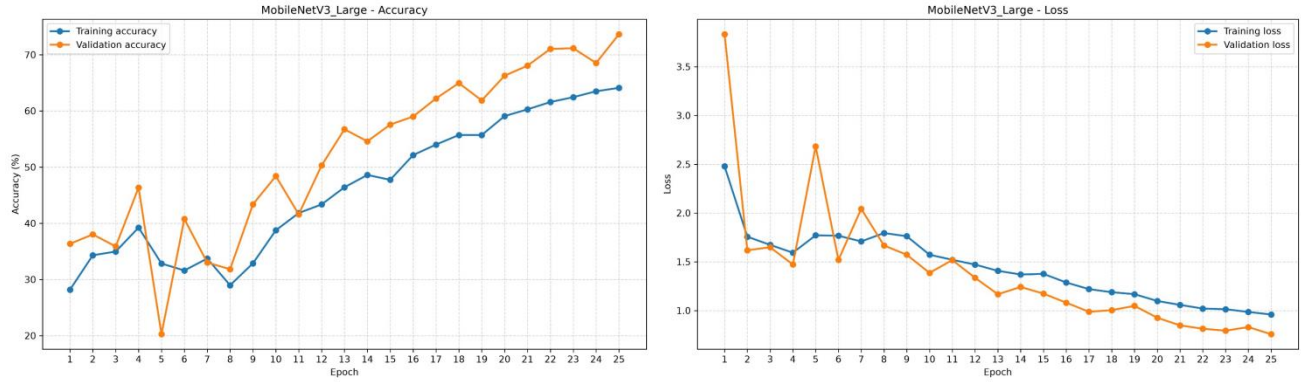
a) *b)*

Fig. 9a), b). *EfficientNetV2-B0 Accuracy and Loss Curves.*



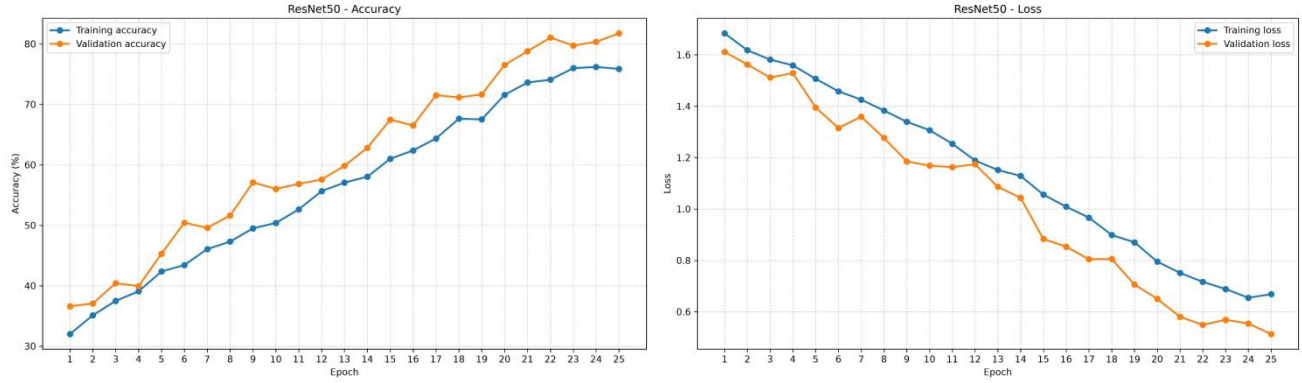
a) b)

Fig. 10a), b). *DenseNet121 Accuracy and Loss Curves.*



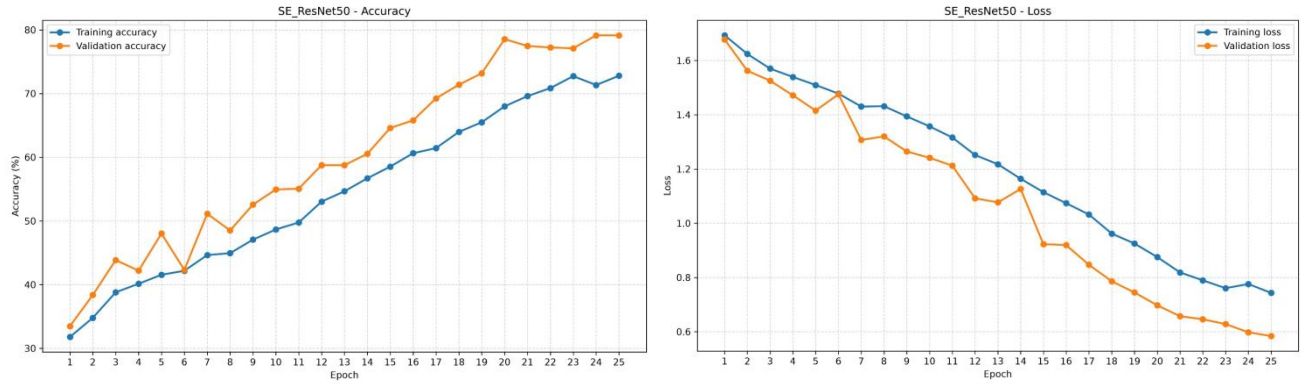
a) b)

Fig. 11a), b) *MobileNetV3-Large Accuracy and Loss Curves.*



a) b)

Fig. 12 a), b) *ResNet50 Accuracy and Loss Curves.*



a) b)

Fig. 13.a), b) *SE- ResNet50 Accuracy and Loss Curves.*

The accuracy and loss curves are analyzed in conjunction to assess the effectiveness of the models' learning and to observe the degree of similarity between their training and validation behaviors. Discrepancies in performance between training and validation may suggest differences in generalization capabilities. Additionally, these curves assist in detecting fluctuations and potential indicators of overfitting throughout the training process.

Confusion Matrix Analysis

Confusion matrices were utilized to assess the classification efficacy of each model across the eight categories of skin diseases. Each matrix juxtaposes the actual class labels against the corresponding predicted labels. Correct predictions are depicted along the main diagonal, while the off-diagonal entries signify misclassifications among the disease categories.

The figures below illustrate the confusion matrices for EfficientNetV2-B0 (Pretrained), ResNet50, DenseNet121, SE-ResNet50, and MobileNetV3-Large, respectively. These matrices offer a class-specific perspective on the classification outcomes and assist in pinpointing categories that are accurately recognized as well as those that may be mistaken for other categories.

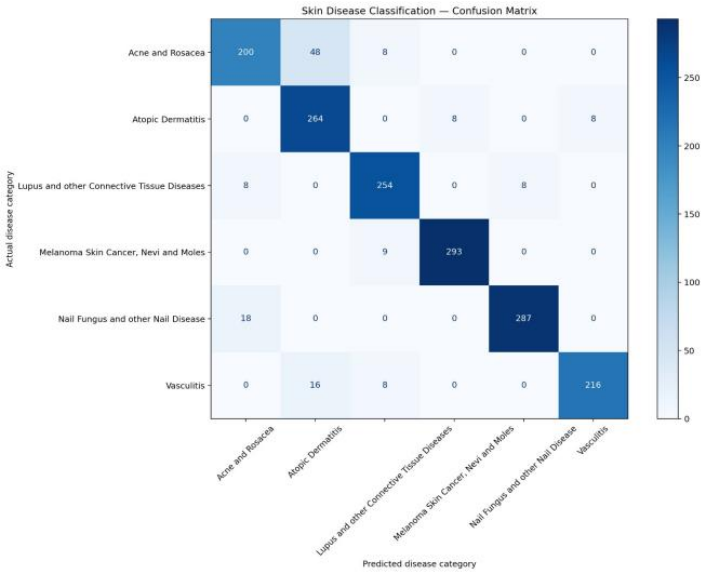


Fig. 13. *Confusion matrix for EfficientNetV2-B0 model.*

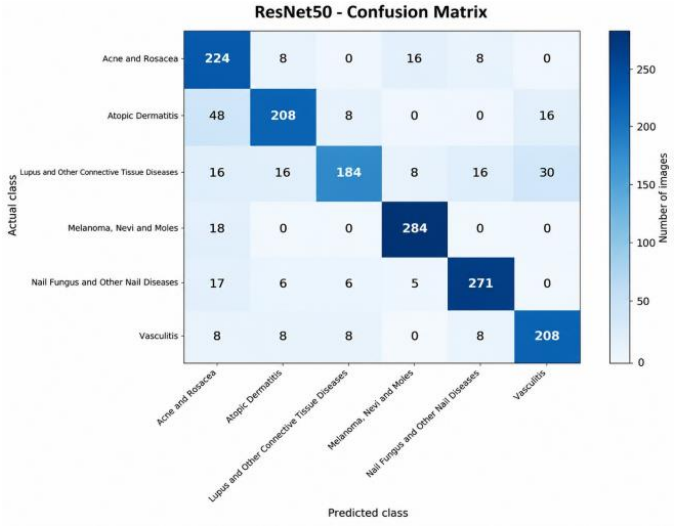


Fig. 14. *Confusion matrix for ResNet50 model.*

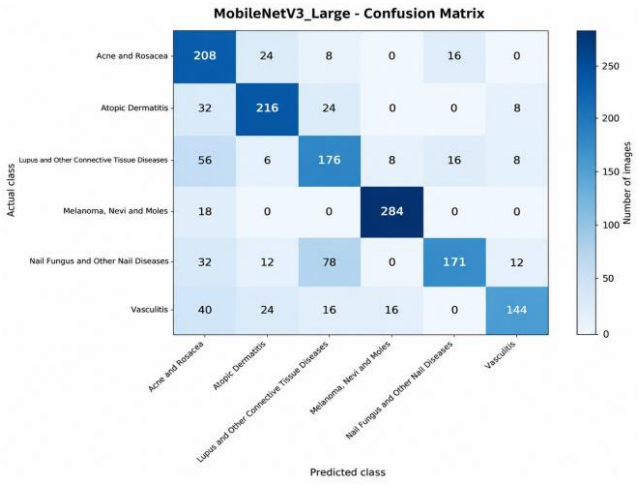


Fig. 15. *Confusion matrix for MobileNetV3-Large model.*

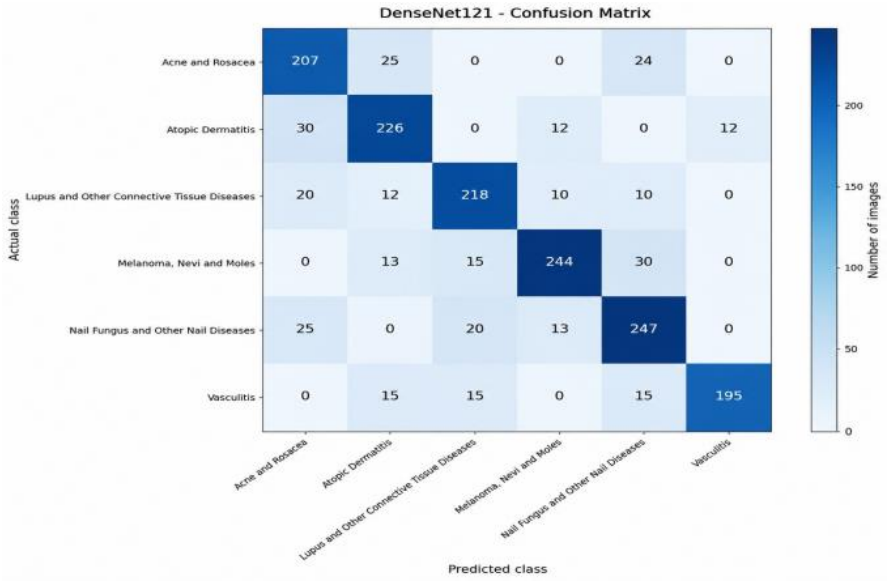


Fig. 16. *Confusion matrix for DenseNet121 model.*

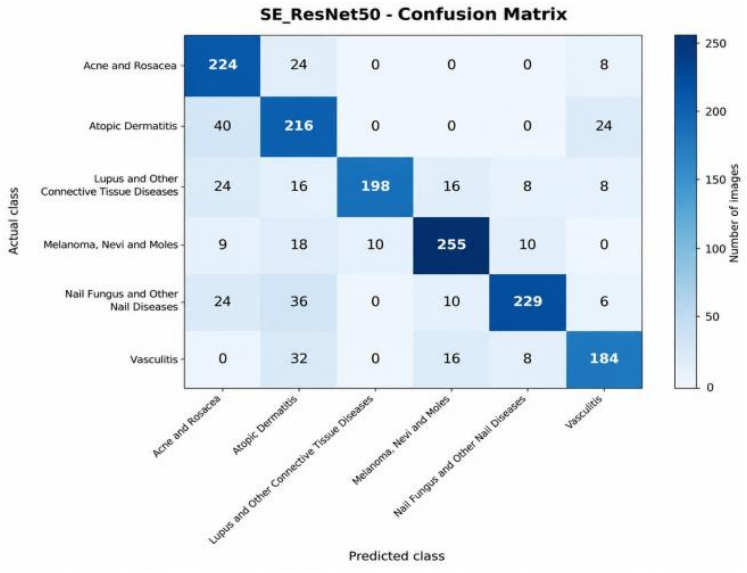


Fig. 17. *Confusion matrix for SE-ResNet50 model.*

Confusion matrices enhance the overall evaluation metrics by illustrating the distribution of accurate and inaccurate predictions for each class. They are especially beneficial for analyzing whether specific categories of skin diseases pose more significant classification difficulties than others.

Quantative Performance Comparison

Table I illustrates the comparative performance of five models according to the test evaluation metrics. EfficientNetV2-B0 (Pretrained) attained the highest test accuracy of 91.59%, in addition to a macro precision of 91.85%, macro recall of 91.27%, macro F1-score of 91.39%, and a weighted F1-score of 91.60%.

ResNet50 secured the second-highest test accuracy at 83.42%, followed by DenseNet121 with 80.87%, SE-ResNet50 at 79.01%, and MobileNetV3-Large at 73.53%. Furthermore, EfficientNetV2-B0 also exhibited the highest values in all reported precision, recall, and F1-score metrics.

Performance Metric	EfficientNetV2-B0 (Pretrained)	ResNet50	DenseNet121	SE-ResNet50	MobileNetV3-Large
Test Accuracy	91.59%	83.42%	80.87%	79.01%	73.53%
Macro Precision	91.85%	83.94%	82.10%	80.63%	74.82%
Macro Recall	91.27%	83.25%	81.99%	79.03%	72.28%
Macro F1-Score	91.39%	83.07%	81.86%	79.19%	72.26%
Weighted F1-Score	91.60%	83.39%	81.86%	79.34%	72.69%

Observation: EfficientNetV2-B0 (Pretrained) achieved the highest performance across all the reported evaluation metrics.

The findings reveal that EfficientNetV2-B0 (Pretrained) outperformed all other assessed architectures on the chosen test dataset.

ResNet50 exhibited the second-best performance, while MobileNetV3-Large showed the least favorable results across the metrics reported.

Overall Discussion

The comprehensive evaluation of the learning curves, confusion matrices, and assessment metrics offers a holistic view of the five deep learning models. The training and validation curves depict their learning patterns, while the confusion matrices reveal details regarding class-specific prediction results. The quantitative analysis further highlights the distinctions in their classification efficacy.

Among the architectures assessed, EfficientNetV2-B0 (Pretrained) exhibited the most robust overall performance, attaining a test accuracy of 91.59% and a macro F1-score of 91.39%. These findings indicate that it is the most proficient model among the five architectures for the skin disease classification task addressed in this research. However, conducting a class-wise analysis and evaluating on independent datasets remains crucial for further validating the reliability and generalization potential of the proposed classification method.

Limitations and Future Work

Limitation

While the assessed deep learning models exhibited encouraging classification results, the study is not without its limitations. The evaluation was performed using a chosen dataset that includes eight categories of skin diseases. Consequently, the reported performance may not encompass the entire spectrum of skin conditions that are typically encountered in real-world clinical environments. The ability of the models to generalize to independent datasets and images obtained under varying imaging conditions has not been verified in this study.

Moreover, the classification performance may fluctuate based on image quality, differences in skin appearance, and the similarities among disease categories. Although multiple classification metrics were utilized to evaluate the models, additional research is necessary to determine their robustness and appropriateness for practical use. Therefore, the proposed system should be regarded as an image-based classification method rather than a substitute for professional medical diagnosis.

Future Work

Although the deep learning models assessed demonstrated promising classification outcomes, the study has its limitations. The evaluation was conducted using a selected dataset comprising eight categories of skin diseases. As a result, the performance reported may not reflect the full range of skin conditions typically encountered in real-world clinical settings. This study has not verified the models' ability to generalize to independent datasets and images captured under different imaging conditions.

Furthermore, the classification performance may vary due to image quality, variations in skin appearance, and the similarities among disease categories. While various classification metrics were employed to assess the models, further research is required to establish their robustness and suitability for practical application. Consequently, the proposed system should be viewed as an image-based classification tool rather than a replacement for professional medical diagnosis.

Conclusion

This research conducted a comparative assessment of five deep learning architectures aimed at the automated classification of skin diseases across eight distinct categories. The models under evaluation included EfficientNetV2-B0 (Pretrained), ResNet50, DenseNet121, SE-ResNet50, and MobileNetV3-Large. Their performance was scrutinized through training and validation curves, confusion matrices, and various classification metrics, such as test accuracy, macro precision, macro recall, macro F1-score, and weighted F1-score.

Among the architectures assessed, EfficientNetV2-B0 (Pretrained) demonstrated the highest overall performance, achieving a test accuracy of 91.59%, a macro precision of 91.85%, a macro recall of 91.27%, a macro F1-score of 91.39%, and a weighted F1-score of 91.60%. ResNet50 followed with the second-best test accuracy of 83.42%, succeeded by DenseNet121, SE-ResNet50, and MobileNetV3-Large.

The comparative results indicate that the selection of deep learning architecture significantly impacts the performance of skin disease classification. Within the experimental framework of this study, EfficientNetV2-B0 (Pretrained) was identified as the most effective among the five models evaluated. These findings lay the groundwork for further exploration into deep learning approaches for skin disease classification, with future research aimed at broader dataset validation, enhanced model interpretability, and practical implementation.

References

[1] Z. Wu, S. Zhao, Y. Peng, X. He, X. Zhao, K. Huang, X. Wu, W. Fan, F. Li, M. Chen, J. Li, W. Huang, X. Chen, and Y. Li, "Studies on different CNN algorithms for face skin disease classification based on clinical images," IEEE Access, vol. 7, pp. 66505–66511, 2019, doi: 10.1109/ACCESS.2019.2918221.

[2] B. Ahmad, M. Usama, C.-M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, "Discriminative feature learning for skin disease classification using deep convolutional neural network," IEEE Access, vol. 8, pp. 39025–39033, 2020, doi: 10.1109/ACCESS.2020.2975198.

[3] Z. L. Fathonah, P. W. Adi, and H. A. Wibawa, "Optimized multiclass classification of skin lesions using a modified EfficientNetV2B0 with transfer learning," Discover Artif. Intell., vol. 6, Art. no. 775, 2026, doi: 10.1007/s44163-026-01545-0.

[4] R. A. Hafizh, R. Wijaya, A. I. Yafi', and S. A. P. Siwi, "Improving skin disease classification with EfficientNetB0: A deep learning approach to overfitting control," in Proc. Int. Conf. Intell. Cybern. Technol. Appl. (ICICyTA), 2024, pp. 513–518, doi: 10.1109/ICICYTA64807.2024.10913154.

[5] K. Xia, Y. Zheng, and Q. He, "Skin lesion recognition with compound scaling and coordinated attention mechanism," in Proc. 2nd Int. Conf. Health Big Data Intell. Healthc. (ICHIH), 2023, pp. 29–32, doi: 10.1109/ICHIH60370.2023.10396419.

[6] A. D. Khalaf, H. Hamdan, A. A. Halin, and N. Manshor, "Segmentation and classification of skin cancer diseases based on deep learning: Challenges and future directions," IEEE Access, vol. 13, pp. 90163–90184, 2025, doi: 10.1109/ACCESS.2025.3569170.

[7] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, Feb. 2017.

[8] DermNet skin disease image library (T. Habif). [Online]. Available: http://www.dermnet.com

[9] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," Sci. Data, vol. 5, no. 1, pp. 1–9, Aug. 2018.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, arXiv:1512.03385.

[11] M. Tan and Q. Le, "EfficientNetV2: Smaller models and faster training," in Proc. Int. Conf. Mach. Learn. (ICML), 2021, pp. 10096–10106.

[12] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," Proc. IEEE, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2009, pp. 248–255.

[14] S. Ibrahim, K. M. Amin, R. I. Alkanhel, H. A. Abdallah, and M. Ibrahim, "Soft attention based EfficientNetV2B3 model for skin cancer's disease classification using dermoscopy images," IEEE Access, vol. 12, pp. 161283–161295, 2024, doi: 10.1109/ACCESS.2024.3486153.

[15] F. Firdous, D. Yadav, S. Z. Rufai, S. Bashir, and S. S. Sarkar, "An ensemble deep learning framework for skin disease classification with explainable and generative AI support," Discover Artif. Intell., vol. 6, Art. no. 524, 2026, doi: 10.1007/s44163-026-01218-y.