International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5291 - 5297
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0692

Classification of Skin Diseases with Deep Learning Based Approaches

Satyajit Behera[1], Sri Bapuji Bismaya Kumar Giri[2], Subham Singh Mundari[3], Kamalakanta Pradhan[4]
[1,2,3]Post Graduate (MCA), Radhakrishna Institute of Technology and Engineering, Bhubaneswar, Odisha, India
Biju Patnaik University of Technology, Rourkela, Odisha, India
[4]Assistant Professor, Radhakrishna Institute of Technology and Engineering, Bhubaneswar, Odisha, India
Biju Patnaik University of Technology, Rourkela, Odisha, India
Emails: *satyajitbehera322@gmail.com[1], bapujigiri1@gmail.com[2], Kamalakanta_cse@riteindia.edu.in[3], mundarisubhamsingh@gmail.com[4]*

Abstract

Skin diseases represent a significant global health concern, requiring accurate and timely diagnosis for effective treatment. This paper presents a deep learning-based approach for classification of skin diseases using medical image analysis. The proposed system uses Convolutional Neural Networks (CNNs) and related deep learning techniques to automatically learn discriminative visual features from skin-lesion images and classify disease categories. Image preprocessing and data augmentation are used to improve consistency and robustness. The workflow covers image acquisition, preprocessing, augmentation, model training, validation, testing, and performance analysis. The system is intended to assist healthcare professionals by providing rapid image-based decision support while reducing dependence on purely manual screening. Standard measures such as accuracy, precision, recall, F1-score, and confusion-matrix analysis are considered for evaluation. The approach demonstrates the potential of artificial intelligence in dermatology while recognizing that dataset quality, external validation, interpretability, and clinical supervision remain essential for responsible deployment.

Keywords: *Skin Disease Classification, Deep Learning, Convolutional Neural Network (CNN), Medical Image Analysis, Artificial Intelligence, Dermatology*

1. Introduction

Skin disorders range from common benign conditions to diseases that may require rapid clinical attention. Visual examination can be difficult when different conditions have similar color, texture, border, shape, or pigmentation patterns. Conventional diagnosis relies heavily on clinical experience and may require additional examination. Computer-aided image analysis can provide consistent pattern recognition and support clinicians during screening. Deep learning has become an important approach for medical image analysis because CNNs can learn hierarchical visual representations directly from images. Early layers can learn edges and textures, while deeper layers combine these patterns into more complex lesion characteristics. This reduces dependence on manually designed image descriptors and makes CNNs suitable for multi-class skin-disease classification when appropriately labeled training data are available. The objective of this work is to describe a complete deep-learning workflow for skin disease classification, including preprocessing, augmentation, model development, evaluation, and interpretation. The system is designed as decision support; a predicted class should not be treated as a definitive medical diagnosis.

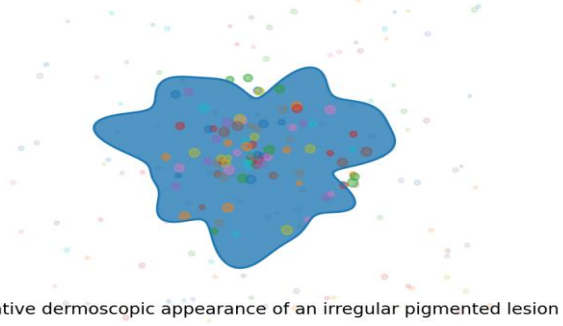


Figure 1 Illustrative dermoscopy-style representation of an irregular pigmented lesion

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5291 - 5297
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0692

1.1.System Architecture

The proposed architecture contains five main stages: image dataset, preprocessing, deep-learning feature extraction, classification, and decision support. Images are first collected with disease labels. Preprocessing standardizes image dimensions and pixel values and can reduce irrelevant artifacts. Augmentation creates realistic training variation. The processed images are supplied to a CNN or transfer-learning model. The final classifier produces probabilities for the target disease categories. As shown in Figure 1 Illustrative dermoscopy-style representation of an irregular pigmented lesion, Figure 2 Proposed deep-learning architecture for skin disease classification, Figure 3 Proposed Deep-Learning Architecture

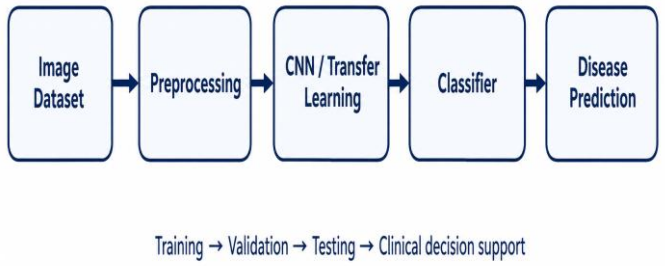


Figure 2 Proposed deep-learning architecture for skin disease classification

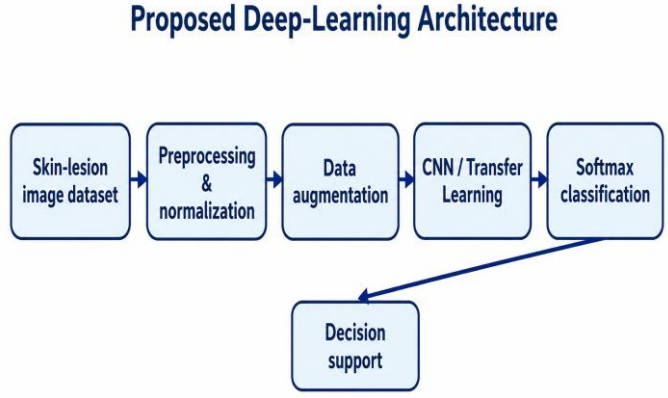


Figure 3 Proposed Deep-Learning Architecture

1.2.Machine Learning Integration

In practical dermatology applications, the learning process can be organized as a supervised multi-class problem. Each image is paired with a target label, and the network minimizes classification error over the training set. The learned representation is not limited to a single visible characteristic; it may combine color distribution, lesion symmetry, local texture, boundary irregularity, and other spatial patterns. This is useful because visually similar skin conditions may differ in several subtle features that are difficult to describe with fixed handcrafted rules. Machine learning is integrated through supervised training using labeled skin-lesion images. A CNN automatically extracts features and learns a mapping from image patterns to disease categories. Transfer learning can be used when the medical dataset is relatively small: a model pertained on a large image dataset provides initial representations that can be fine-tuned on dermatology images. Validation data are used to monitor generalization and select the best model configuration. A transfer-learning strategy can further reduce the amount of task-specific training required. A pertained backbone such as a residual network can be initialized with general visual representations and then fine-tuned using the dermatology dataset. During training, the learning rate, batch size, number of epochs, dropout, and early-stopping criteria should be selected using only the training and validation partitions. This separation helps prevent the test set from influencing model selection and provides a more reliable estimate of generalization. For multi-class classification, the final layer can use a softmax function to convert model outputs into class probabilities. Cross-entropy loss measures the difference between predicted probabilities and the true label. Optimizers such as Adam or stochastic gradient descent update model parameters during training. Class weighting or balanced sampling may be considered when disease classes are unevenly represented.

2. Method

The methodology follows image acquisition, preprocessing, augmentation, training, validation, prediction, and evaluation. Images are resized to a common input dimension and normalized. Training-only augmentation can include rotation, flipping, cropping, and controlled brightness variation. The dataset is divided into training, validation, and testing

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5291 - 5297
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0692

subsets while maintaining class representation as far as possible. The trained network is then used to generate disease-category predictions for unseen test images. As shown in Figure 4 hierarchical feature extraction and multi – class prediction, Figure 5 CNN Processing Pipeline for Hierarchical Feature Extraction[1].

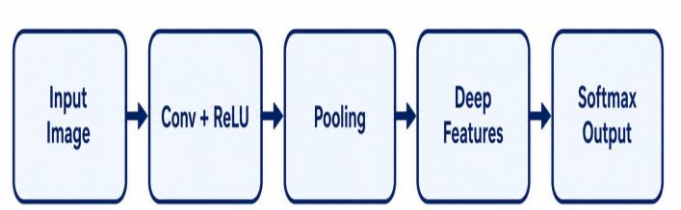


Figure 4 hierarchical feature extraction and multi – class prediction

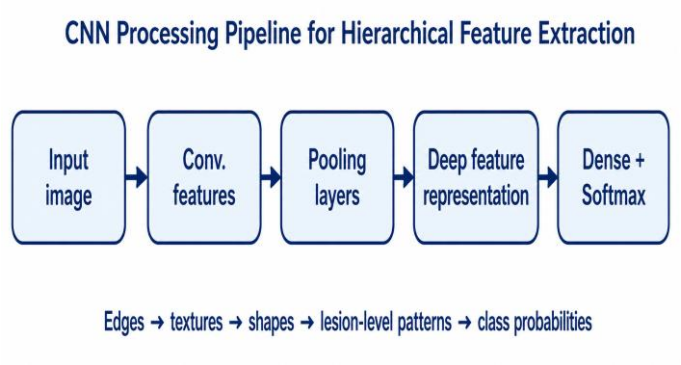


Figure 5 CNN Processing Pipeline for Hierarchical Feature Extraction

Preprocessing is important because image quality can strongly influence learned features. Images with severe blur, incorrect labels, duplicated samples, or unusable regions should be identified before model training. Data leakage must be avoided by ensuring that closely related images from the same case do not appear across training and testing subsets. The final test set should remain isolated until model selection is complete[3].

3. Experiment / Analysis

The experimental design should report both overall and class-specific performance. Accuracy gives the overall proportion of correctly classified images, while precision indicates how reliable positive predictions are for a class. Recall indicates the ability to identify images belonging to that class. The F1-score combines precision and recall and is useful when class frequencies differ. A confusion matrix provides a detailed view of which disease categories are confused with one another. As shown in Table 1 Experimental input and evaluation parameters for the proposed approach[2].

Table 1 Experimental input and evaluation parameters for the proposed approach

Parameter	Typical setting	Purpose	Analysis
Input size	224 × 224 pixels	Standardize CNN input	Consistent features
Normalization	Scaled pixel values	Stabilize training	Improved convergence
Augmentation	Flip, rotate, crop	Increase variation	Better robustness
Loss	Cross-entropy	Measure error	Guide learning
Evaluation	Accuracy, P, R, F1	Measure quality	Class-wise comparison

The supplied source material describes the classification approach but does not provide a measured experimental dataset, trained-model scores, or a numerical confusion matrix. Therefore, numerical accuracy values are not fabricated here. Once the model is trained on a defined dataset, the results should be reported using the above measures and supported by class-wise error analysis. For a complete experimental study, the dataset should be documented by reporting the number of images, disease categories[4], class distribution, image source, and partitioning strategy. The same patient or lesion should not contribute related images to different partitions because this can produce overly optimistic results. A stratified split is preferable when enough samples are available in every class. Repeated experiments or cross-validation can also be considered during development, while a final independent test set should remain untouched until the end[5].

4. Results And Discussion

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5291 - 5297
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0692

The integrated workflow is expected to provide a consistent method for converting skin-lesion images into disease-category predictions. Preprocessing improves input consistency, augmentation helps reduce sensitivity to limited visual variation, and CNN feature learning enables automatic extraction of complex image patterns. The final evaluation stage provides evidence of overall and class-specific behavior. Performance can vary substantially with image source, acquisition device, illumination, lesion location, skin characteristics, disease prevalence, and annotation quality. A model that performs well on a single curated dataset may not generalize to a new clinical environment. External validation on independent data is therefore important before practical deployment. As shown in Figure 3 Evaluation and error-analysis workflow, Figure 4 Evaluation and error[6].

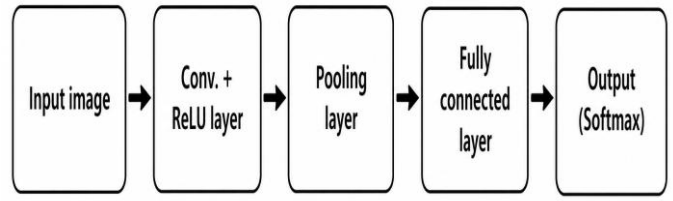


Figure 3 Evaluation and error-analysis workflow

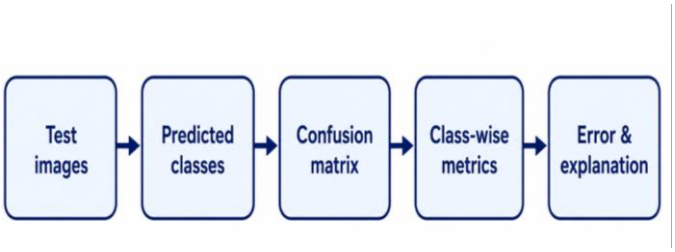


Figure 4 Evaluation and error

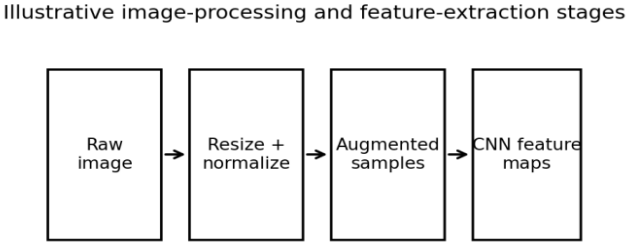
Interpretability is another important component. Visual explanation methods can indicate whether predictions are driven by the lesion region or by irrelevant background artifacts. Such analysis can reveal shortcut learning and guide additional preprocessing or data collection. In clinical use, the model should provide supportive information while the final decision remains with an appropriately qualified professional.

5. Clinical Deployment and Ethical Considerations

Privacy and governance are equally important. Medical images should be handled under applicable institutional and legal requirements, with appropriate de-identification, access control, secure storage, and documented data provenance. The model output should communicate uncertainty and should not discourage referral, biopsy, or further examination when clinical evidence indicates that these are necessary. A research[7] prototype should be separated from a clinical diagnostic product. Before deployment, the model should be evaluated on independent images collected under different conditions and, where possible, on data from multiple institutions. Monitoring is also necessary after deployment because camera upgrades, changes in clinical practice, and differences in patient populations can alter the input distribution[8].

5.1.Dataset Preparation and Quality Control

Figure 5 Illustrative image preprocessing and CNN feature-extraction workflow



Dataset preparation is a critical stage because the reliability of a classification model is strongly influenced by the consistency of its training examples. Each image should be associated with a verified disease label, and images with severe blur, unusable regions, incorrect annotations, or obvious duplication should be identified before model training. The dataset should also document the image source, acquisition conditions, disease categories, and class distribution so that later experiments can be reproduced and interpreted. To reduce information leakage, images originating from the same patient or lesion should not be distributed across training and testing subsets. A stratified partition is preferable when enough samples are available in every class because it preserves approximate class proportions. Augmentation should be applied only to the training

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5291 - 5297
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0692

data. Typical transformations include controlled rotation, horizontal or vertical flipping where medically appropriate, cropping, and moderate brightness variation. Excessive transformations should be avoided because they can create unrealistic lesion patterns.

5.2.Model Training and Hyperparameter Selection

During supervised training, the network receives an image and its corresponding class label and adjusts its parameters to reduce classification loss. For a multi-class problem, the soft ax output provides a probability distribution over the target categories. Cross-entropy loss is suitable for measuring the difference between predicted probabilities and the reference label. Adam or stochastic gradient descent can be used for parameter optimization, while dropout and early stopping can help control over fitting. Hyper parameters such as learning rate, batch size, number of epochs, dropout rate, and augmentation strength should be selected using the training and validation partitions only. The independent test set must remain untouched until the final model configuration is fixed. This separation is important because repeated decisions based on test performance can gradually make the test set part of the model-selection process and produce an overly optimistic estimate of generalization.

5.3.Error Analysis and Interpretability

Overall accuracy alone cannot describe the complete behavior of a medical image classifier. Class-wise recall is particularly important because missed cases in a less frequent category may be hidden by strong performance on common categories. A confusion matrix can reveal systematic confusion between visually similar disease classes and can guide targeted data collection or preprocessing improvements. Interpretability can further strengthen the evaluation process. Explanation techniques such as Grad-CAM can highlight image regions that contribute strongly to a prediction. If highlighted regions consistently correspond to the lesion rather than borders, rulers, skin markings, or background artifacts, the explanation provides useful evidence about the learned decision process. Conversely, attention to irrelevant regions may indicate shortcut learning and should trigger additional investigation.

5.4.Computational and Practical Considerations

The computational requirements of a deep-learning system depend on image resolution, network depth, batch size, and the chosen hardware. Larger models may provide greater representational capacity but require more memory and training time. Lightweight architectures can be investigated for resource-constrained environments, including systems where images are captured and analyzed near the point of care. Regardless of deployment hardware, preprocessing and inference should remain consistent with the conditions used during validation.A practical workflow can include image acquisition, automated quality checks, preprocessing, model inference, probability display, and explanation visualization. A confidence threshold or referral rule may be incorporated only after appropriate validation. The system should be presented as decision support rather than as an autonomous diagnostic authority, particularly when the predicted probabilities are uncertain or when the image falls outside the distribution represented in the training data.

Conclusion

Limitations

The proposed framework has several limitations. Its performance depends on the diversity and quality of the available images, and a model trained mainly on one acquisition environment may not transfer directly to another. Differences in cameras, dermatoscopes, illumination, lesion location, image resolution, and patient populations can create distribution shifts. Class imbalance can also make the network favor common categories. In addition, automated predictions cannot replace clinical examination, patient history, biopsy, or other investigations when these are medically indicated. These limitations motivate external validation and careful clinical oversight before deployment. This paper presents a deep-learning-based framework for classification of skin diseases from medical images. The proposed workflow combines image preprocessing, data augmentation, CNN-based feature extraction, multi-class classification, and multi-metric evaluation. The approach can support rapid and consistent image analysis and has potential for intelligent dermatology systems. However, reliable deployment requires

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5291 - 5297
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0692

representative data, rigorous validation, interpretability, privacy protection, and clinical oversight.

Future Work

Future work can investigate larger and more diverse datasets, transfer-learning comparisons, class-imbalance strategies, explainable AI, and external clinical validation. Lightweight models may also be explored for resource-constrained environments. Further research can examine multimodal systems that combine lesion images with relevant patient or clinical information while maintaining appropriate privacy and ethical safeguards.

Material And Method

The proposed system uses labeled skin-lesion images as the primary data source. The data are preprocessed, augmented, and divided into training, validation, and testing subsets. A CNN or transfer-learning model is trained with a suitable classification loss, and validation performance is monitored during model selection. Final testing is performed only after the model configuration is fixed. Predictions are summarized using standard classification metrics and a confusion matrix, followed by qualitative inspection of representative errors and explanation maps.

Acknowledgements

The authors acknowledge Radhakrishna Institute of Technology and Engineering, Bhubaneswar, Odisha, and Biju Patnaik University of Technology, Rourkela, for academic support and guidance during this work. As shown in Table 2 Comparative Analysis.

Table 2 Comparative Analysis

Approach	Feature Learning	Data Requirement	Computational Cost	Typical Use
Conventional CNN	Learned end-to-end	Moderate–high	Moderate	Baseline multi-class classifier
Transfer Learning	Pretrained + fine-tuned	Lower	Low–moderate	Small/medium datasets
Residual Network	Deep residual features	Moderate	Moderate–high	High-capacity classification
Lightweight CNN	Compact learned features	Moderate	Low	Mobile/edge deployment
CNN + Explainability	Features + visual explanation	Moderate	Moderate–high	Decision support and audit

For a fair comparison, the same training, validation, and independent test protocol should be used for every candidate model. Hyperparameters must be selected without using the final test set. Reporting macro-averaged precision, recall, and F1-score is especially important when some skin-disease categories contain fewer images than others. The choice of model should therefore consider more than overall accuracy. A clinically useful system should maintain balanced class-wise recall, produce reliable probabilities, tolerate changes in image acquisition, and provide explanations that focus on the lesion rather than background artifacts. Lightweight architectures may be preferable for mobile or resource-constrained deployments, while larger models may be suitable for research environments with stronger hardware. Deep-learning approaches differ in the way they represent visual information, computational cost, and their suitability for clinical decision support. A conventional CNN can learn task-specific features directly from the available training images, whereas transfer learning starts from a pretrained visual representation and adapts it to the target dermatology classes. In limited-data settings, transfer learning can reduce training time and improve the stability of feature extraction, provided that the target images are sufficiently compatible with the pertained representation.

Recommended Evaluation Checklist

Verify class labels and remove duplicated or unusable images; 2) split data at the patient/lesion

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5291 - 5297
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0692

level to reduce leakage; 3) use training-only augmentation; 4) tune hyperparameters on validation data; 5) report accuracy together with macro and class-wise precision, recall and F1-score; 6) inspect the confusion matrix and representative errors; 7) evaluate calibration and explanation maps where applicable; and 8) perform independent external validation before making clinical claims. As shown in Table 2 Classification metrics and their interpretation.

Table 2 Classification metrics and their interpretation

Metric	Formula / Definition	Interpretation
Accuracy	(TP + TN) / all samples	Overall correctness
Precision	TP / (TP + FP)	Reliability of positive predictions
Recall	TP / (TP + FN)	Ability to identify a class
F1-score	2PR / (P + R)	Balance of precision and recall
Macro-F1	Mean F1 across classes	Gives each class equal weight

Table 3 Comparison of classification approaches

Approach	Strength	Limitation	Best-fit scenario
CNN	Learns visual features directly	Needs sufficient labeled data	Baseline image classification
Transfer learning	Useful with smaller datasets	Needs suitable pretrained representation	Limited/medium dermatology datasets
Residual network	Supports deeper feature learning	Higher computational cost	Research-scale classification
Lightweight CNN	Lower memory and inference cost	May have lower capacity	Mobile or edge deployment
CNN + Grad-CAM	Adds visual explanation	Extra analysis and computation	Auditable decision support

References

[1]. A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017.
[2]. P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," Scientific Data, vol. 5, 2018.
[3]. N. C. F. Codella et al., "Skin lesion analysis toward melanoma detection: A challenge at the International Symposium on Biomedical Imaging," IEEE ISBI, 2018.
[4]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. IEEE CVPR, pp. 770–778, 2016.
[5]. R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," Proc. IEEE ICCV, pp. 618–626, 2017.
[6]. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," MICCAI, pp. 234–241, 2015.
[7]. M. E. Celebi et al., "A methodological approach to the classification of dermoscopy images," Computerized Medical Imaging and Graphics, vol. 31, no. 6, pp. 362–373, 2007.
[8]. L. G. F. D. Oliveira et al., "Deep learning for skin lesion classification: A review,"