International Journal of Innovative Trends in Engineering Science and Technology
ijitest.org | E-ISSN: 3139-6887

Automated Skin Cancer Detection and Lesion Segmentation using CNN-Based Deep Learning

[1]CH. Swapna Priya, [2]K. Suseela, [3]T. Dansy Sri Abeesh
[1]Associate Professor, [2]Department of School education [3]Department of Computer Science Engineering,
[1]swapnachsp@gmail.com, [2]jackbenison12@gmail.com,[3]dansysriabeesh@gmail.com

Abstract -- Skin cancer is a significant health concern, and its early identification is essential for supporting timely medical intervention. However, distinguishing malignant skin lesions from benign conditions through visual examination can be challenging because of similarities in their appearance, colour and texture. This research presents a deep learning-based framework for automated skin lesion classification using dermoscopic images. The proposed system investigates the performance of five convolutional neural network architectures, namely InceptionV2, InceptionV3, DenseNet, MobileNet and GoogleNet. The HAM10000 skin lesion dataset is considered for model development and evaluation. Initially, the images are processed and organized into training and testing datasets. Transfer learning techniques are then employed to train the selected deep learning models for multiclass skin disease classification. Their performance is compared using classification accuracy, precision, recall, F1-score and confusion matrices. The experimental observations indicate that DenseNet provides the strongest classification performance among the evaluated architectures, achieving approximately 99% accuracy in the reported experiment. Furthermore, a U-Net-based segmentation module is incorporated to identify and highlight the affected lesion regions within the input images. The combined classification and segmentation approach provides both the predicted disease category and its corresponding affected region. The proposed framework can serve as a computer-aided screening tool to support dermatological assessment, although final diagnosis must remain under the supervision of qualified medical professionals.

Keywords: Skin cancer detection, melanoma classification, deep learning, convolutional neural network, DenseNet, HAM10000, transfer learning, U-Net, medical image segmentation.

I. INTRODUCTION

Skin cancer is one of the major dermatological health concerns worldwide, and its occurrence is influenced by several factors, including ultraviolet radiation exposure, genetic characteristics and environmental conditions. Among the different forms of skin cancer, melanoma requires particular attention because of its potential to spread to other parts of the body. Identifying suspicious skin lesions at an early stage can improve the possibilities of effective treatment and reduce the risk of serious complications.

Conventional skin lesion assessment generally involves visual examination by dermatologists, dermoscopic inspection and, when necessary, histopathological evaluation. Although these approaches are important in clinical practice, accurate identification can be difficult when different skin conditions exhibit similar visual characteristics. Variations in lesion shape, colour, texture, border irregularity and pigmentation further increase the complexity of the diagnostic process. Therefore, automated image analysis techniques have become an important research direction for supporting skin lesion assessment.

Recent developments in artificial intelligence and computer vision have introduced new possibilities for medical image classification. In particular, Convolutional Neural Networks (CNNs) can automatically learn relevant image representations through multiple convolutional and pooling operations. Unlike traditional approaches that depend heavily on manually selected features, deep learning models can identify complex visual patterns directly from training images. Transfer learning further improves their practical usefulness by allowing pretrained networks to be adapted to medical image classification tasks.

This research investigates the application of five established CNN architectures: InceptionV2, InceptionV3, DenseNet, MobileNet and GoogleNet. The models are trained and evaluated using dermoscopic skin lesion images from the HAM10000 dataset. The objective is to examine their classification capabilities and identify the architecture that provides the most effective results for the considered experimental setup.

An additional segmentation component is incorporated using the U-Net architecture. While classification identifies the predicted disease category, segmentation highlights the

DOI: 10.5281/zenodo.24468878
IJITEST-2026-030 | Volume 1, Issue 3, May 2026 | ISSN: 3139-6887
Creative Commons Attribution 4.0 International (CC BY 4.0)
Article Received: November 7, 2026 | Accepted: November 23, 2026

corresponding lesion region within the input image. This provides more visual information about the affected area and makes the system output easier to interpret.

The principal contributions of this research are as follows:

1. Development of an automated framework for multiclass skin lesion classification using deep learning.
2. Comparative evaluation of five CNN architectures to identify their classification performance.
3. Application of transfer learning to improve feature extraction and classification.
4. Integration of U-Net segmentation to identify the affected region of the skin lesion.
5. Presentation of disease classification and lesion segmentation outputs within a unified framework.

The remainder of this paper discusses related research, the proposed methodology, experimental observations, performance analysis and concluding remarks.

II. Related Work

Automated skin lesion analysis has been investigated using conventional image processing, machine learning and deep learning approaches. Early research primarily concentrated on image enhancement, lesion segmentation and the extraction of descriptive characteristics. Asha Gnana Priya et al. [1] investigated melanoma identification using dermoscopic images and traditional image processing techniques. Their approach employed Otsu thresholding and morphological operations to separate the lesion from the surrounding skin. Features extracted from the segmented region were subsequently used to calculate the Total Dermatoscopy Score (TDS) for melanoma identification.

Bakheet [2] introduced an SVM-based framework for malignant melanoma detection using optimized Histogram of Oriented Gradients (HOG) descriptors. The approach extracted important gradient information from skin lesion images and supplied the resulting feature vectors to a Support Vector Machine classifier. The reported experimental performance included 98.21% sensitivity, 96.43% specificity and 97.32% accuracy, demonstrating the potential of optimized feature extraction combined with machine learning.

Devassy et al. [3] examined the effectiveness of combining manually designed image descriptors with deep representations for melanoma classification. Their approach utilized a ResNet architecture for deep feature extraction and SIFT descriptors for conventional feature representation. The experimental findings indicated that combining SIFT with deep features did not necessarily improve classification accuracy. Their observations highlighted the importance of selecting appropriate feature representations rather than assuming that combining multiple feature types always produces better results.

Gao et al. [4] proposed a Privileged Modality Distillation framework for vessel border detection in intracoronary imaging. Their approach incorporated complementary image information during training to improve the learning process. The knowledge obtained from multiple modalities was subsequently transferred to a single-modality network, allowing the final model to operate without requiring the additional modality during testing. This research illustrates the potential of knowledge transfer and complementary information in medical image analysis.

Gao et al. [5] developed a privileged-information-based deep learning framework for implicit strain reconstruction in ultrasound elastography. Their method incorporated additional information during training to guide the neural network towards more meaningful intermediate representations. A physics-based data generation strategy was also introduced to establish relationships between training samples, privileged information and expected outputs. The experimental results demonstrated the effectiveness of this approach for ultrasound strain reconstruction.

Although these studies demonstrate the usefulness of conventional image processing, machine learning and advanced deep learning techniques, challenges remain in achieving reliable multiclass skin lesion classification and identifying the corresponding affected regions. Motivated by these observations, the present research evaluates multiple CNN architectures and incorporates a U-Net segmentation component to provide classification and visual localization within a common framework.

III. Proposed Methodology

The proposed framework is designed to classify different types of skin lesions using deep learning techniques and identify the corresponding affected regions through image segmentation. The overall workflow consists of dataset collection, image preprocessing, dataset preparation, CNN model development, performance evaluation and lesion segmentation.

3.1 Dataset Collection

The HAM10000 (Human Against Machine with 10,000 training images) dataset is used for the experimental investigation. It contains dermoscopic images representing seven categories of pigmented skin lesions. These categories include:

- Actinic keratoses and intraepithelial carcinoma (AKIEC)
- Basal cell carcinoma (BCC)
- Benign keratosis-like lesions (BKL)
- Dermatofibroma (DF)
- Melanocytic nevi (NV)
- Melanoma (MEL)
- Vascular lesions (VASC)

The dataset provides diverse skin lesion samples with variations in colour, shape, texture and appearance. Such

variations make it suitable for evaluating deep learning architectures in multiclass skin lesion classification.

3.2 Image Preprocessing

Image preprocessing is performed to prepare the dermoscopic images for CNN-based classification. The input images are loaded into the processing environment and converted into a suitable numerical representation. Image dimensions are standardized according to the input requirements of the selected network architectures.

The preprocessing stage also involves organizing the images according to their corresponding disease categories. These operations ensure consistency in the input representation and facilitate efficient model training.

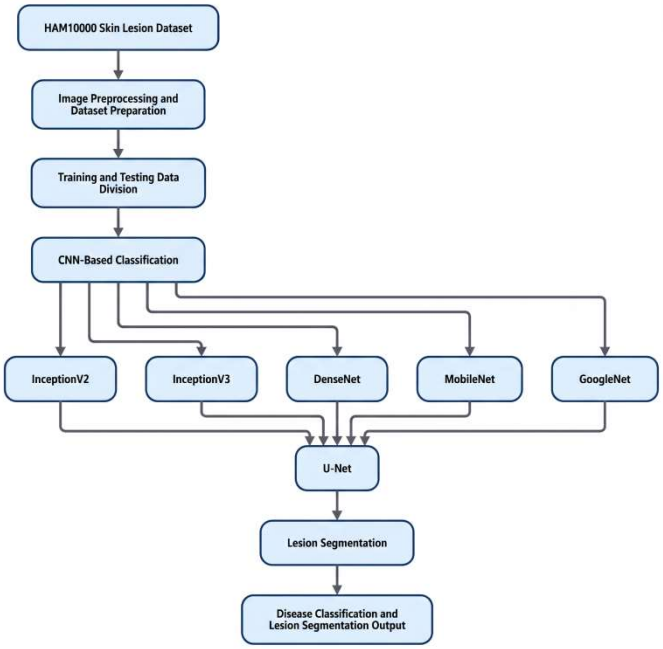

Figure 1: Overall workflow of the proposed CNN-based skin lesion classification and segmentation framework.

3.3 Dataset Division

The prepared image dataset is divided into training and testing subsets. The training samples are used to learn the visual characteristics of different skin lesion categories, while the testing samples are utilized to evaluate the classification performance of the trained models.

Maintaining separate training and testing datasets is important for assessing the ability of the models to classify previously unseen images. Class distribution and potential data imbalance should also be considered during dataset preparation.

3.4 Deep Learning-Based Classification

Five CNN architectures are investigated to identify their suitability for skin lesion classification. Transfer learning is employed to utilize previously learned image representations and adapt the networks to the target classification task.

InceptionV2: This architecture uses multiple convolutional operations to capture visual information at different scales. Batch normalization helps improve the stability of the learning process.

InceptionV3: InceptionV3 extends the Inception architecture through improved convolutional arrangements and computationally efficient feature extraction. It is used to investigate the classification capability of deeper network representations.

DenseNet: DenseNet introduces direct connections between successive layers, allowing feature information to be reused throughout the network. These connections support effective feature propagation and improve gradient flow during training. The architecture demonstrated the highest classification accuracy in the reported experiments.

MobileNet: MobileNet is a computationally efficient CNN architecture designed primarily for resource-constrained environments. It employs depthwise separable convolutions to reduce computational requirements while maintaining useful image representations.

GoogleNet: GoogleNet utilizes Inception modules containing parallel convolutional operations with different filter sizes. This structure enables the network to capture image characteristics at multiple spatial scales.

3.5 Model Training and Classification

Each selected CNN architecture is trained using the prepared skin lesion images. During the training process, the network learns discriminative features that represent the visual differences among the seven disease categories.

The final classification layer is adapted to generate predictions corresponding to the target skin lesion classes. After training, the models are evaluated using the testing dataset. Their predicted categories are compared with the actual labels to examine classification performance.

3.6 Performance Evaluation

The performance of the developed classification models is evaluated using commonly employed classification metrics.

Accuracy: Measures the proportion of correctly classified samples among all evaluated samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision: Measures the proportion of correctly predicted positive samples among all samples predicted as positive.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall: Determines the proportion of actual positive samples correctly identified by the classifier.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score: Represents the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Confusion matrices are also generated to examine the relationship between actual and predicted disease categories. These evaluation measures provide a broader understanding of model performance beyond overall classification accuracy.

3.7 Skin Lesion Segmentation Using U-Net

In addition to classification, U-Net is incorporated to identify the affected lesion region in dermoscopic images. U-Net is an encoder-decoder-based convolutional neural network designed for image segmentation tasks.

The encoder progressively extracts spatial and contextual information through convolutional and pooling operations. The decoder reconstructs the spatial representation to generate a segmentation mask. Skip connections transfer relevant feature information from the encoder to the corresponding decoder layers, helping preserve spatial details.

The resulting segmentation mask highlights the lesion region within the original image. By combining the classification result with the segmentation output, the system provides two forms of information: the predicted skin disease category and the corresponding affected region.

IV. Results and Discussions

The experimental evaluation was conducted to compare the classification performance of five CNN architectures on dermoscopic skin lesion images. The models considered in this investigation were InceptionV2, InceptionV3, DenseNet, MobileNet and GoogleNet. Their performance was examined using classification accuracy and confusion matrix representations.

4.1 Comparative Classification Performance

Table 1: Summarized reported experimental results.

CNN architecture	Reported accuracy
InceptionV2	28%
InceptionV3	28%
DenseNet	99%
MobileNet	78%
GoogleNet	66%

Reported accuracy values from the uploaded project experimental screenshots.

The experimental observations indicate considerable differences in the classification performance of the evaluated architectures. InceptionV2 and InceptionV3 achieved approximately 28% accuracy in the reported implementation. Their corresponding confusion matrices showed substantial disagreement between actual and predicted categories, indicating difficulties in distinguishing the different lesion classes under the considered experimental conditions.

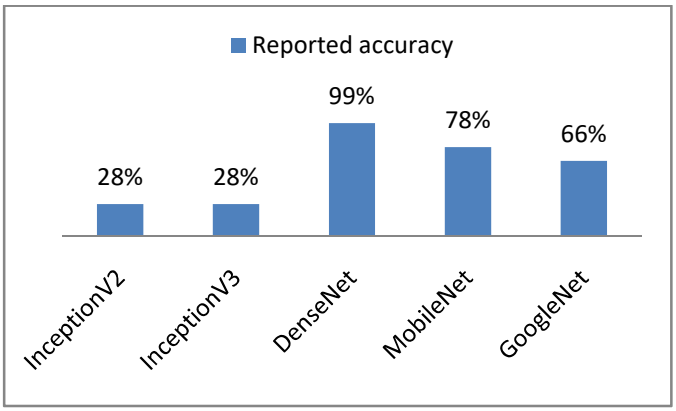

Figure 2: Comparison of CNN classification accuracy

MobileNet achieved 78% accuracy, demonstrating better classification performance than the two Inception models. Its computationally efficient architecture offers an advantage for applications where processing resources are limited. GoogleNet achieved 66% accuracy, indicating a moderate classification capability compared with the other evaluated models. DenseNet obtained the highest reported accuracy of approximately 99%. Its dense connectivity structure enables information to flow across multiple network layers, supporting the reuse of learned features. The corresponding confusion matrix showed stronger agreement between actual and predicted labels than the other evaluated architectures.

4.2 Analysis of Confusion Matrices

Confusion matrices were used to examine the classification behaviour of individual CNN models. In the InceptionV2 and InceptionV3 experiments, the predicted labels showed considerable deviations from the actual categories. This suggests that these models did not adequately distinguish the different skin lesion classes in the reported implementation.

In contrast, the DenseNet confusion matrix demonstrated a stronger diagonal distribution, indicating a higher number of correctly classified samples. This observation supports the superior overall accuracy reported for DenseNet. The confusion matrices of MobileNet and GoogleNet showed intermediate classification performance. These observations suggest that network architecture and feature representation have an important influence on the classification of visually similar skin lesions.

4.3 Skin Disease Prediction

The developed application was also evaluated using individual test images. The classification model generated predicted disease labels corresponding to the seven categories of the HAM10000 dataset. The reported demonstration included predictions such as AKIEC, BKL and MEL. These outputs illustrate the ability of the trained model to generate disease-category predictions from dermoscopic images.

4.4 Lesion Segmentation Results

The U-Net segmentation model was incorporated as an extension to the classification framework. Unlike the classification component, which identifies the disease

category, the segmentation component focuses on locating the affected lesion region. The segmentation output produces a mask corresponding to the identified skin lesion. This mask is presented alongside the original input image, allowing the affected region to be visually examined. The combined framework therefore generates two outputs: the predicted skin lesion category and its corresponding segmented region. This additional visual information can support the interpretation of the model's predictions during preliminary assessment.

4.5 Overall Discussion

The experimental comparison indicates that DenseNet achieved the strongest classification performance among the five evaluated CNN architectures. MobileNet and GoogleNet provided moderate results, whereas InceptionV2 and InceptionV3 demonstrated comparatively lower accuracy in the reported implementation. The integration of U-Net further extends the framework by providing lesion localization information in addition to disease classification. Nevertheless, the reported 99% accuracy should be interpreted in the context of the experimental dataset and evaluation procedure. Further testing with patient-independent data splits, class-wise sensitivity, specificity, precision, recall, F1-score and external datasets is necessary to establish the model's generalization capability. The system should be considered a research-oriented decision-support approach rather than an independent diagnostic solution.

V. Conclusion and Future Scope

This research presents a deep learning-based framework for automated skin lesion classification and affected-region identification using dermoscopic images. Five CNN architectures, namely InceptionV2, InceptionV3, DenseNet, MobileNet and GoogleNet, were investigated to compare their classification capabilities. The experimental results demonstrated that DenseNet achieved the highest reported accuracy of approximately 99%, outperforming the other evaluated architectures. MobileNet and GoogleNet showed moderate performance, whereas InceptionV2 and InceptionV3 obtained comparatively lower accuracy in the implemented experiments. Furthermore, the integration of U-Net provided an additional segmentation capability to identify and highlight the affected skin lesion regions. The combined classification and segmentation framework offers both disease-category prediction and visual information about the corresponding lesion area. These findings demonstrate the potential of deep learning techniques in developing computer-aided skin lesion assessment systems. However, further validation using independent clinical datasets and comprehensive evaluation metrics is required before considering the framework for practical clinical applications. Future research should focus on improving classification reliability through attention-based CNN architectures, vision transformers, explainable artificial intelligence, and larger, balanced clinical datasets, alongside enhanced lesion segmentation and external validation.

References

[1] Asha Gnana Priya, H., Anitha, J., and Poonima Jacinth, J. (2018). Identification of melanoma in dermoscopy images using image processing algorithms. Proceedings of the International Conference on Control, Power, Communication and Computing Technologies (ICCPCCT), pp. 553–557.

[2] Bakheet, S. (2017). An SVM framework for malignant melanoma detection based on optimized HOG features. Computation, 5(1), 1–13.

[3] Devassy, B., Yildirim-Yayilgan, S., and Hardeberg, J. (2019). The impact of replacing complex hand-crafted features with standard features for melanoma classification using both hand-crafted and deep features. Advances in Intelligent Systems and Computing, 868, 150–159.

[4] Gao, Z., et al. (2019). Privileged modality distillation for vessel border detection in intracoronary imaging. IEEE Transactions on Medical Imaging, 39(5), 1524–1534.

[5] Gao, Z., Wu, S., Liu, Z., Luo, J., Zhang, H., Gong, M., and Li, S. (2019). Learning the implicit strain reconstruction in ultrasound elastography using privileged information. Medical Image Analysis, 58, 101534.

[6] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4700–4708.

[7] Hussain, Z., Gimenez, F., Yi, D., and Rubin, D. (2017). Differential data augmentation techniques for medical imaging classification tasks. AMIA Annual Symposium Proceedings, 2017, 979.

[8] Jafari, M. H., Karimi, N., Nasr-Esfahani, E., Samavi, S., Soroushmehr, S. M. R., Ward, K., and Najarian, K. (2016). Skin lesion segmentation in clinical images using deep learning. 23rd International Conference on Pattern Recognition (ICPR), pp. 337–342.

[9] Jafari, M. H., Nasr-Esfahani, E., Karimi, N., Soroushmehr, S. M. R., Samavi, S., and Najarian, K. (2017). Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma. International Journal of Computer Assisted Radiology and Surgery, 12(6), 1021–1030.

[10] Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

[11] Tschandl, P., Rosendahl, C., and Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5, 180161.

Funding Declaration

The authors declare that no funds, grants, or other forms of financial support were received from any organization or institution for the conduct of this research.