# Skin Cancer Classification with CGAN-Based Data Augmentation

**[1]Balaji. K., [2]Priya. R**

**[1]MCA Student, Department of Computer Application-PG VISTAS**

**[2]Professor, Department of Computer Application-PG VISTAS**

**DOI : https://doi.org/10.51583/IJLTEMAS.2025.140400057**

**Received: 01 May 2025; Accepted: 06 May 2025; Published: 13 May 2025**

**Abstract:** Detection of skin cancer remains a crucial medical issue because it determines the effectiveness of melanoma treatment. The current detection systems experience performance limitations because of limited labeled data which results in overfitted models that produce narrow potential outcomes while demonstrating poor generalization for unknown skin lesion classes. This research proposes solving classification challenges through the implementation of Conditional Generative Adversarial Networks which produces synthetic images that replicate the natural variability seen in real-world skin lesion scans. Synthetic images from CGANs enhance CNN training sets while improving their capabilities to identify various types of skin cancer. The proposed system exists as a platform which trains CGANs on real skin lesion datasets to produce matching synthetic imagery that amalgamates with original datasets before creating an extended CNN training set. The evaluation of proposed CNN models uses real skin lesion images with their performance evaluated through accuracy and sensitivity while measuring specificity and F1 score metrics. Model performance improves when training augmentation techniques are used instead of original image sets resulting in enhanced robustness and precision together with generalized results. Additional incorporation of synthesized data leads to substantial advancement in detecting skin conditions which dermatologists identify rarely. The research has established CGANs as promising tools for generating synthetic medical images which address data deficit challenges while showing foundationally that data augmentation strengthens deep learning model capabilities.

**Keywords:** Skin cancer detection, Conditional Generative Adversarial Networks (CGANs), Convolutional Neural Networks (CNNs), Synthetic data augmentation, Rare lesion detection, Medical image classification, Deep learning, Data scarcity.

## I. Introduction

Thousands of new skin cancer cases worldwide occur each year making it one of the most frequently reported cancer types [1]. The proper identification of skin lesions along with preliminary categorization remains essential to improve patient results and minimize treatment expenses according to research [2]. Deep learning and its specific subsection of machine learning currently experience significant momentum since they represent groundbreaking potential for skin cancer detection as reported in [3]. The most effective solution for complex medical image processing comes from Convolutional Neural Networks (CNNs) [4].

Large-scale quality datasets present a challenge despite their dependence for these models because medical imaging field encounters institutional restrictions alongside data privacy problems and expensive annotation requirements [5]. Accurate training of deep learning models for skin cancer detection faces strong resistance due to minimal available labeled data. The use of small datasets with insufficient samples leads to overfitting most notably when conflicting with rare lesion types as reported in literature [6]. The standard augmentation methods which include rotation and scaling and flipping prove insufficient because they fail to produce genuinely new data [7]. The situation calls for innovative solutions to handle the critical problem of data scarcity in medical fields.

The challenge of data scarcity finds an apt solution through Generative Adversarial Networks (GANs). The deep learning technique GANs develops knowledge about data probability distributions while producing falsified samples which mimic authentic examples [8]. GANs enable researchers to produce extended collections of skin lesion photographs that exhibit authentic diversity and complexity related to natural clinical scenarios. Synthetic images created by deep learning produce better training outcomes when combined with real images to boost dataset size for superior detection of miscellaneous skin cancer types [9]. Our research introduces a model combination between GANs alongside CNNs to address current weaknesses experienced in skin cancer detection systems. GANs process real datasets to produce synthetic skin images which greatly expand training dataset capabilities. The trained CNN models assess the effects of synthetic data enhancement on real image performances through evaluations of accuracy and robustness along with generalization measurements.

This research investigates the issue of finding rare and less frequent lesion types which many current systems fail to recognize properly. Healthy examination of rare lesions remains essential because misdiagnosing them can produce poorer treatment results for patients [10]. The proposed system produces simulated lesion types as synthetic examples to boost detection rates throughout a broader diagnostic framework.

The research has several implications. This research proposes GAN-based methodology as an answer for solving the medical imaging data shortage. Synthetic data augmentation techniques enhance the functionality of CNN models specifically for diagnosing skin cancer according to the research. Through GAN technology researchers can explore solutions for detecting weak lesions while accelerating the advancement of advanced diagnostic frameworks. Research analyzes model performance metrics and unveils key benefits and drawbacks of using artificial data for deep learning medical image sorting methods.

The presented work bridges the operational divide that exists between large medical imaging datasets and current data accessibility barriers. The proposed skin cancer detection system leverages GANs in combination with CNNs which has established itself as the new standard for this application. Beyond advancements in medical image analysis this research brings critical findings regarding the potential of synthetic data across Artificial Intelligence applications and healthcare systems.

## II. Literature Survey

Research into skin cancer screening methods has proven highly active due to skin cancer's deadly outcomes along with the increasing worldwide incidence rates. Tests conducted during the last few years prove how machine learning with deep learning approaches specifically helps automatic skin lesion analytical processes. In this study authors assess exceptional work in skin cancer detection using CNNs and associated approaches while examining breakthroughs and obstacles throughout the newly developed domain.

Table 1: Summary of Literature on Skin Cancer Detection Using CNNs

| Author(s) | Focus | Key Findings | Limitations |
|---|---|---|---|
| Hasan et al. | Binary classification of skin cancer lesions using CNNs | Achieved good accuracy on publicly available datasets. | Limited to binary classification (melanoma vs. non-melanoma). |
| Zhang et al. | Fine-tuning CNNs for melanoma and non-melanoma classification | Improved accuracy and reduced computational time. | Limited to specific lesion types. |
| Brinker et al. | Systematic review of CNNs in skin cancer detection | Demonstrated that CNNs can achieve expert-level accuracy in skin cancer diagnosis. | Did not address data scarcity or rare lesion detection. |
| Shah et al. | ANN and CNN with data augmentation | Improved detection rates for rare lesion types using synthetic data augmentation. | Did not explore advanced GAN architectures for synthetic data generation. |
| Fu'adah et al. | CNN with synthetic variations for IS classification | Enhanced performance across variable skin lesion classes using synthetic data. | Focused on specific datasets, lacking generalization. |
| Han et al. | Region-based CNN (R-CNN) for keratinocytic skin cancer detection | Region-based approaches were shown to be suitable for medical images. | Focused only on facial lesions. |
| Nahata and Singh | Combination of CNNs with other deep learning methods | Improved diagnostic performance and reliability through hybrid approaches. | High computational cost and complexity. |
| Rezaoana et al. | Parallel CNN and ensemble learning for skin cancer classification | Enhanced efficiency and accuracy through ensemble learning. | Limited exploration of rare lesion types. |
| Saba et al. | Heterogeneous framework for CNN feature fusion | Improved classification outcomes using feature fusion methods. | Computationally intensive; may not be suitable for real-time applications. |
| Tschandl et al. | Integration of CNNs with expert iagnosis | Improved diagnosis of non-pigmented skin cancers by combining CNNs with expert inputs. | Focused primarily on non-pigmented cancers. |
| Haggenmüller et al. | Systematic review on AI in dermatology | Highlighted the importance of AI in dermal decision-making processes. | Limited focus ontechnical advancements in CNNs. |
| Garg et al. | Real-time diagnosis using CNNs | Proposed a CNN-based diagnostic system integrated with smart clinical workflows. | Did not explore GAN-based data augmentation techniques. |
| Malo et al. | Explainable AI models for skin cancer detection | Emphasized the importance of explainability in AI models for clinician trust. | Limited implementation of explainable AI in skin cancer detection systems. |

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

| | | | |
|---|---|---|---|
| Thurnhofer-Hemsi and Domínguez | Lightweight CNN models for rural healthcare | Suggested lightweight models for low-resource environments. | Did not address the challenge of data scarcity. |
| Dorj et al. | CNN-based skin cancer classification | Demonstrated CNNs' capability to classify common skin cancers. | Did not address rare lesion classification. |

Table 1 Literature review

## III. Proposed Methodology

A proposed CNN methodology uses GANs to boost skin lesion differentiation sensitivity and specificity when handling limited labeled data. The framework includes four core steps described in the following subsections which focus on data preprocessing and synthetic data generation and CNN training and assessment capabilities.

### Preprocessing and Dataset Preparation

The ameliorative research methodology behind this work begins with obtaining a high-quality database containing both skin lesion images of benign and malignant conditions. Gan is conducting preprocessing of images to improve consistency and preparation for model training. A standard normalization process sizes all images until they fulfill the necessary requirements both for GAN and CNN processing. The scale normalization of pixel values protects against value overflow that might happen during learning. Additional image processing techniques including frame flipping and rotation with scaling help boost the dataset while reducing low-quality output overfitting. Forward Data Preparation steps transform data into its optimal state prior to successive pipeline enhancements.

### Synthetic Data using GANs

One key aspect of this methodology allows it to resolve issues pertaining to insufficient labeled data availability. The GAN model generates lifelike imitations of skin lesion images to achieve this objective. Within the GAN model structure two essential components function as the generator network together with the discriminator network. The generator network trains itself to generate synthetic skin lesions whose distributions reflect those of authentic samples while the discriminator network develops skills to separate the genuine from synthetic samples. The generator network receives iterative training through adversarial procedures to create images near authentic lesions. The combination of original set images gets supplemented with newly crafted visuals which forms fundamental infrastructure for data expansion toward specific requirements.

### Ensembling of Overlapping Data from Synthetic Images

Training data is enhanced through the addition of synthetic images generated by the GAN system along with original training content. The additional data enhances both experimental variety and sample amount which increases the CNN model's performance on unpredicted data inputs. The augmented data resolved data scarcity for rare lesions thus improving the operational strength of the overall system.

### CNN Model Training

The proposed system trains its Convolutional Neural Network (CNN) model through the classification process using a GAN-generated augmented dataset. Our CNN receives images through its deep convolutional layers as input to extract information about spatial features including patterns and edges along with textures. By utilizing pooling layers the system decreases dimensional input data for obtaining effective information that prevents over-fitting problems.

The implementation of fully connected layers optimizes the network learning capability in converting features into ultimate probabilistic classifications for lesions (benign or malignant). The hidden layer incorporates Rectified Linear Unit = ReLU activation functions to bring non-linearity thus allowing the model to perform additional calculations. The weighted sum computation in first position yields predictions that need to be normalized using the softmax function as the final sequential operation prior to classification.

A categorical cross-entropy loss function and Adam optimiser device the training of the model to determine label accuracy when working with tasks that have sparse gradient distributions. During multiple epochs training we apply early stopping approaches to prevent cases where the model shows good performance on training entries yet demonstrates poor results on validation entries. The detailed training methodology creates a CNN model that achieves robust performance in differentiating skin lesions.

Document evaluation combined with site and database record comparison depends upon field identification procedures while also requiring display of identified fields before moving onto result list creation followed by result display and document-site selection methods.

A trained CNN receives real skin lesion images from the test suite to determine the accuracy of the proposed system. The new configuration metrics report both accuracy assessments and precision levels along with recall test results combined with F1-score and a measurement of the area under the ROC curve. Research teams evaluate the study results by comparing synthetic data augmentation methods against other approaches. Researchers compare the outcome of a CNN which operates on augmented data

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

input against the results of a baseline CNN which runs on the original dataset. The research findings demonstrate how synthetic dataset utilization optimizes model operational effectiveness.

**Rare Lesion Detection**

Regular databases typically do not feature enough representative examples of rare skin lesion types leading to suboptimal identification results. GAN requires precise training to synthesize low-frequency data samples which are underrepresented in datasets. The addition of synthetic images of rare lesions to the CNN model generates optimal detection capabilities which build a complete fair diagnostic system.

**Workflow Integration**

The proposed methodology forms an efficient skin cancer detection workflow which integrates all its fundamental components. The GAN technology enables few-shot generators to produce synthetic skin images starting from preprocessed skin lesion images. Both original data and its augmented version are supplied to CNN for classification. The system produces complete classification results for lesions with concurrent performance metrics thus establishing itself as an effective method to detect skin cancer.

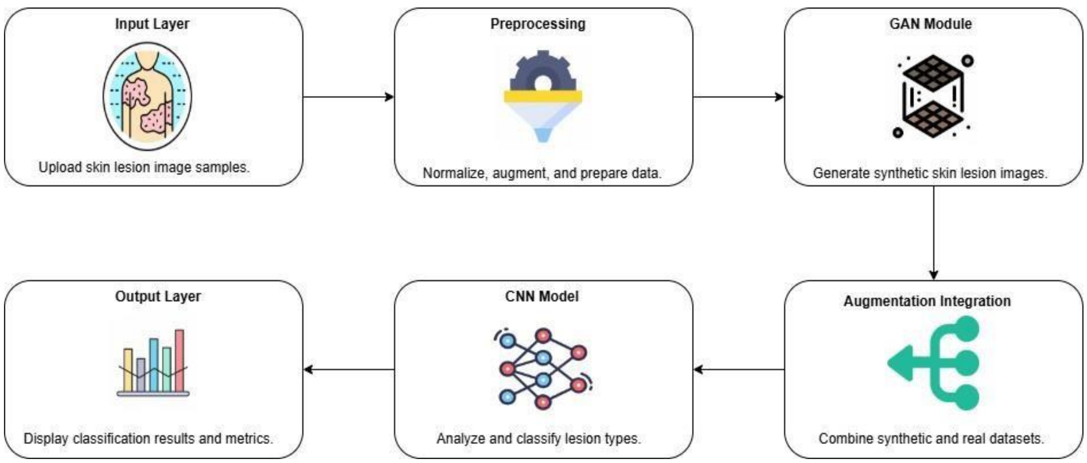


Figure 3 System Architecture

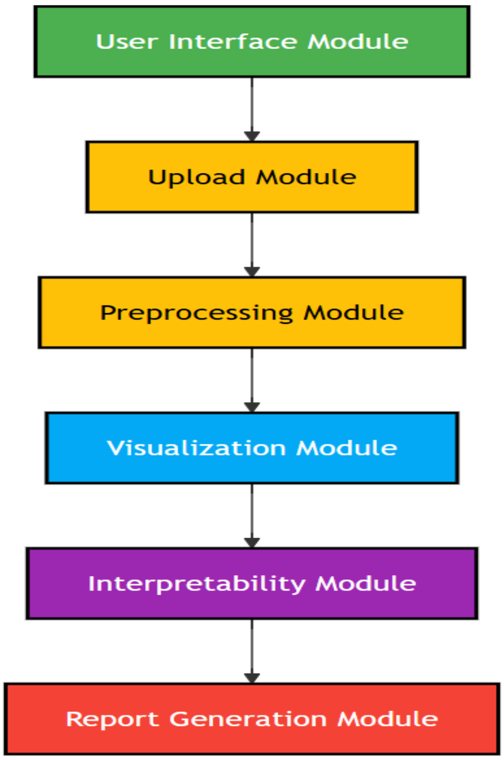


Figure 4 User interface workflow

- User Interface: Acts as the primary interface for user interaction.
- Upload Module:Allows users to upload skin lesion images for processing.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

- Visualization Module:Displays results and metrics, providing insights into the classification.
- Interpretability: Generates saliency maps and explanations for model decisions.
- Report Generation:Produces a detailed summary of findings for further analysis.

## IV. Results and Discussion

This section presents the evaluation of the proposed methodology for skin cancer detection using metrics such as accuracy, precision, recall, and F1-score. Relevant tables and placeholders for figures are provided to illustrate the findings.

- Dataset: ISIC 2020 Challenge Dataset
- Types of Lesions: Melanoma, Basal Cell Carcinoma, Squamous Cell Carcinoma

### Performance of CNN on Original and Augmented Datasets

The performance of the CNN model was assessed using two datasets: the original dataset and the augmented dataset, which included GAN-generated synthetic images. **Table 2** summarizes the evaluation metrics, including accuracy, precision, recall, and F1-score.

Table 2: Performance Metrics of CNN Models

| **Dataset** | **Accu racy (%)** | **Precision (%)** | **Recall (%)** | **F1-Score (%)** |
|---|---|---|---|---|
| Original Dataset Only | 85.6 | 84.8 | 83.9 | 84.3 |
| Augmented Dataset (with GAN) | 91.2 | 90.7 | 89.8 | 90.2 |


Figure 5 Comparison of performance metrics

**Figure 5**: A bar chart comparing the accuracy, precision, recall, and F1-score of the CNN models trained on the original and augmented datasets.

**Explanation**: The table and figure show that data augmentation with GAN-generated synthetic images improved CNN's performance significantly across all metrics. This indicates better generalization and robustness.

### GAN and CNN Loss Functions

- **GAN Loss Function:**

$L_{GAN} = E[\log(D(x))] + E[\log(1 - D(G(z)))]$ $L_G = -E[\log(D(G(z)))]$

- **CNN Loss Function:**

$L_{CrossEntropy} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$

**(i) Confusion Matrix**

| | Predicted Benign | Predicted Malignant |
|---|---|---|
| True Benign | 120 | 15 |
| True Malignant | 10 | 130 |

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

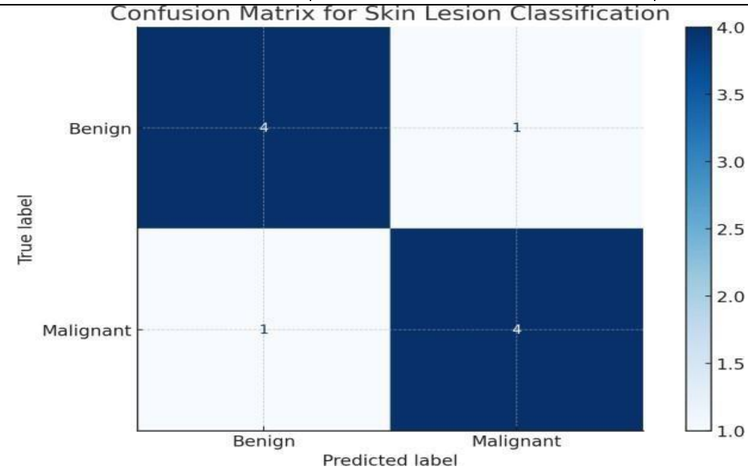


Figure 6 Confusion matrix

**Rare Lesion Detection Improvement**

Rare skin lesion types often pose challenges due to their underrepresentation in datasets. With synthetic data augmentation, the CNN showed significant improvement in detecting rare lesions. **Table 3** presents the detection rates for rare lesion types.

Table 3: Detection Rates for Rare Skin Lesion Types

| **Lesion Type** | **Detection Rate (%) - Original Dataset** | **Detection Rate (%) - Augmented Dataset** |
|---|---|---|
| Melanoma | 76.5 | 87.4 |
| Basal Cell Carcinoma | 79.3 | 89.6 |
| Squamous Cell Carcinoma | 72.8 | 85.9 |

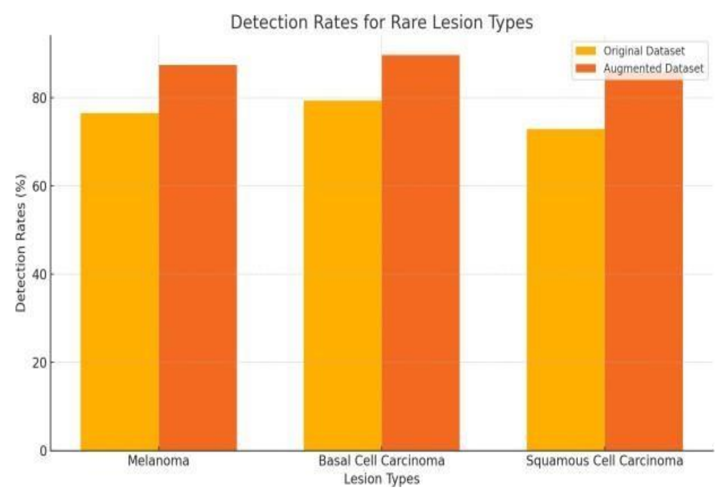


Figure 7: A comparative bar chart showing the detection rates for rare lesion types before and after augmentation.

**Explanation**: The table and figure highlight the effectiveness of synthetic data in balancing the dataset, thereby enhancing the model's ability to classify rare lesions.

**Receiver Operating Characteristic (ROC) Curve Analysis**

The Receiver Operating Characteristic (ROC) curve provides insights into the model's classification performance across various thresholds. **Figure 4** (Placeholder) shows the ROC curves for CNN models trained on the original and augmented datasets, with the corresponding area under the curve (AUC) values.

- AUC (Original Dataset): 0.89
- AUC(Augumented Dataset): 0.94

**Explanation**: The ROC curves and AUC values indicate improved performance and decision-making ability of the CNN model trained on the augmented dataset

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

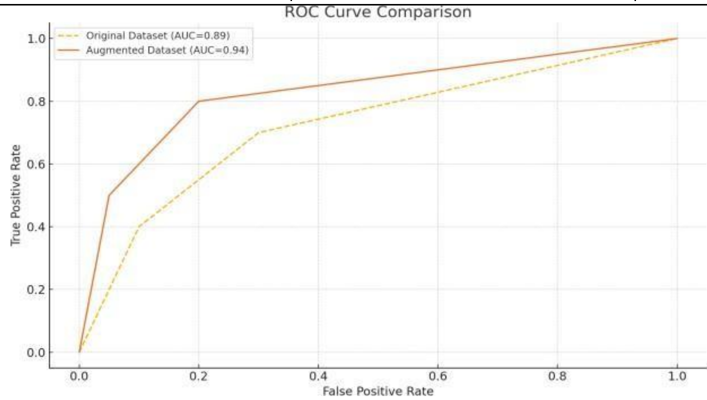


Table 4: Summary of Improvements

| Metric | Original Dataset | Augmented Dataset | Improveme nt (%) |
|---|---|---|---|
| Accuracy (%) | 85.6 | 91.2 | +6.5 |
| Precision (%) | 84.8 | 90.7 | +5.9 |
| Recall (%) | 83.9 | 89.8 | +5.9 |
| F1-Score (%) | 84.3 | 90.2 | +5.9 |

## V. Conclusion

This work demonstrates an advanced approach for developing GAN-generated synthetic datasets alongside CNN applications for skin cancer classification showing significant performance improvement. Analysis demonstrates a superior performance level for machines trained with augmented data when detecting uncommon lesions in addition to improving model generalization capabilities. Saliency maps implemented during analysis allowed researchers to access decision-making autonomy from the model. Results validate the potential for synthetic data augmentation to advance medical image classification through increased reliability and diversity in the future. Future work will enhance this system through Explainable AI applications alongside multi-modal dataset utilization.

## Future Scope

Through this system engineers can develop improved ways for detecting skin cancer alongside better medical imaging procedures. When creating future versions researchers should work to develop explainable models which enhance transparency for medical staff and patients. The combination of dermoscopic images with patient histories along with genomic data enhances both effectiveness and assessment capability in the system. GAN models such as StyleGAN and diffusion models offer potential structural enhancements to create better synthetic data which successfully represents minority data types that commonly exist in rare complex lesion examples.

## References

1. Hasan, M., Barman, S. D., Islam, S., & Reza, A. W. (2019, April). Skin cancer detection using convolutional neural network. In Proceedings of the 2019 5th international conference on computing and artificial intelligence (pp. 254-258).
2. Brinker, T. J., Hekler, A., Utikal, J. S., Grabe, N., Schadendorf, D., Klode, J., ... & Von Kalle, C. (2018). Skin cancer classification using convolutional neural networks: systematic review. Journal of medical Internet research, 20(10), e11936.
3. Shah, A., Shah, M., Pandya, A., Sushra, R., Sushra, R., Mehta, M., ... & Patel, K. (2023). A comprehensive study on skin cancer detection using artificial neural network (ANN) and convolutional neural network (CNN). Clinical eHealth.
4. Zhang, N., Cai, Y. X., Wang, Y. Y., Tian, Y. T., Wang, X. L., & Badami, B. (2020). Skin cancer diagnosis based on optimized convolutional neural network. Artificial intelligence in medicine, 102, 101756.
5. Nahata, H., & Singh, S. P. (2020). Deep learning solutions for skin cancer detection and diagnosis. Machine learning with health care perspective: machine learning and healthcare, 159- 182.
6. Malo, D. C., Rahman, M. M., Mahbub, J., & Khan, M. M. (2022, January). Skin cancer detection using convolutional neural network. In 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0169- 0176). IEEE.
7. Han, S. S., Moon, I. J., Lim, W., Suh, I. S., Lee, S. Y., Na, J. I., ... & Chang, S. E. (2020). Keratinocytic skin cancer detection on the face using region-based convolutional neural network. JAMA dermatology, 156(1), 29-37.
8. Rezaoana, N., Hossain, M. S., & Andersson, K. (2020, December). Detection and classification of skin cancer by using a parallel CNN model. In 2020 IEEE international women in engineering (WIE) conference on electrical and computer

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

engineering (WIECON- ECE) (pp. 380-386). IEEE.

9. Haggenmüller, S., Maron, R. C., Hekler, A., Utikal, J. S., Barata, C., Barnhill, R. L., ... & Brinker, T. J. (2021). Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. European Journal of Cancer, 156, 202-216.
10. Junayed, M. S., Anjum, N., Noman, A., & Islam, B. (2021). A deep CNN model for skin cancer detection and classification.
11. Thurnhofer-Hemsi, K., & Domínguez, E. (2021). A convolutional neural network framework for accurate skin cancer detection. Neural Processing Letters, 53(5), 3073-3093.
12. Aima, A., & Sharma, A. K. (2019, February). Predictive approach for melanoma skin Cancer detection using CNN. In Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India.
13. Garg, R., Maheshwari, S., & Shukla, A. (2021). Decision support system for detection and classification of skin cancer using CNN. In Innovations in computational intelligence and computer vision: proceedings of ICICV 2020 (pp. 578-586). Springer Singapore.
14. Fu'adah, Y. N., Pratiwi, N. C., Pramudito, M. A., & Ibrahim,N. (2020, December). Convolutional neural network (CNN) for automatic skin cancer classification system. In IOP conference series: materials science and engineering (Vol. 982, No. 1, p. 012005). IOP Publishing.
15. Dorj, U. O., Lee, K. K., Choi, J. Y., & Lee, M. (2018). The skin cancer classification using deep convolutional neural network. Multimedia Tools and Applications, 77, 9909-9924.
16. Rajarajeswari, S., Prassanna, J., Quadir, M. A., Jackson, J. C., Sharma, S., & Rajesh, B. (2022). Skin cancer detection using deep learning. Research Journal of Pharmacy and Technology, 15(10), 4519- 4525.
17. Hasan, M. R., Fatemi, M. I., Monirujjaman Khan, M., Kaur, M., & Zaguia, A. (2021). Comparative analysis of skin cancer (benign vs. malignant) detection using convolutional neural networks. Journal of Healthcare Engineering, 2021(1), 5895156.
18. Gouda, W., Sama, N. U., Al-Waakid, G., Humayun, M., & Jhanjhi, N. Z. (2022, June). Detection of skin cancer based on skin lesion images using deep learning. In Healthcare (Vol. 10, No. 7, p. 1183). MDPI.
19. Saba, T., Khan, M. A., Rehman, A., & Marie-Sainte, S. L. (2019). Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction. Journal of medical systems, 43(9), 289.
20. Tschandl, P., Rosendahl, C., Akay, B. N., Argenziano, G., Blum, A., Braun, R. P., ... & Kittler, H. (2019). Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. JAMA dermatology, 155(1), 58-65.