

## Hybrid Deep Learning for Skin Cancer Classification Using Dermatoscopic Images With Explainable AI

<sup>1</sup>Monday Simon, <sup>2</sup>Nasu Mbimi, <sup>1</sup>Umar Abdullahi, <sup>1</sup>Samuel Amachundi Adda, <sup>3</sup>Victoria Zevini Sabo, <sup>1</sup>Munu Stephen

<sup>1</sup>Department of Computer Science Federal University Wukari, Taraba State.

<sup>2</sup>Department of Mathematical Sciences, ATBU, Bauchi

<sup>3</sup>Department of Cyber security Federal University Wukari, Taraba State

### ABSTRACT

Skin cancer is one of the most prevalent and life-threatening diseases worldwide, with melanoma representing the most aggressive form due to its high metastatic potential and mortality rate. Early and accurate detection is essential for improving treatment effectiveness and patient survival outcomes. However, conventional diagnostic approaches based on visual inspection and dermatoscopy remain highly subjective and dependent on clinician expertise, which can lead to variability and misdiagnosis. To overcome these limitations, this study proposes an explainable hybrid deep learning framework that integrates Convolutional Neural Networks (CNNs) and Gated Recurrent Units (GRU) for automated skin cancer classification using dermatoscopic images. The proposed CNN-GRU model leverages the CNN's capability to extract fine-grained spatial features such as texture, color variation, and lesion boundaries, while the GRU component captures broader contextual dependencies across feature representations to enhance classification robustness. The model was trained and evaluated using publicly available datasets, including the ISIC Archive and HAM10000, which contain high-quality annotated dermatoscopic images of benign and malignant skin lesions. Standard preprocessing techniques, including image resizing, normalization, and data augmentation, were applied to improve dataset diversity and reduce overfitting. An explainable AI framework was incorporated to improve clinical interpretability, utilizing Gradient-weighted Class Activation Mapping (Grad-CAM) for visualizing CNN feature importance and attention-based visualization for interpreting GRU outputs. These methods provide transparent insights into the regions of interest influencing model predictions, thereby increasing clinical trust and interpretability. Experimental results show that the proposed CNN-GRU hybrid model achieved an overall classification accuracy of 85%, outperforming the standalone CNN model, which achieved 84%. The hybrid model also recorded improved macro and weighted average precision, recall, and F1-scores of 0.85 compared to 0.84 for the CNN baseline, indicating more balanced and reliable performance. At the class level, the model achieved a precision of 0.87 and recall of 0.84 for benign lesions, and a precision of 0.81 and recall of 0.86 for malignant lesions, demonstrating strong sensitivity in detecting malignant cases, which is critical for reducing false negatives in clinical diagnosis. Hence, the results confirm that the integration of CNN and GRU architectures enhances feature representation by combining local spatial feature extraction with global contextual learning. The hybrid model demonstrates improved generalization, stable convergence, and reduced misclassification across classes. When combined with explainable AI techniques, the framework not only delivers high classification performance but also provides interpretable decision support, making it a reliable and clinically applicable tool for early skin cancer detection and diagnosis.

### ARTICLE INFO

#### Article History

Received: February, 2026

Received in revised form: April, 2026

Accepted: May, 2026

Published online: June, 2026

### KEYWORDS

Deep Learning (DL), Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), Dermatoscopic Images.

### INTRODUCTION

Skin cancer is a significant global health issue, with millions of cases annually,

Melanoma, a severe form poses a high risk of metastasis and death if not diagnosed early (Tahir et al., 2023). The rise in cases is linked to

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.

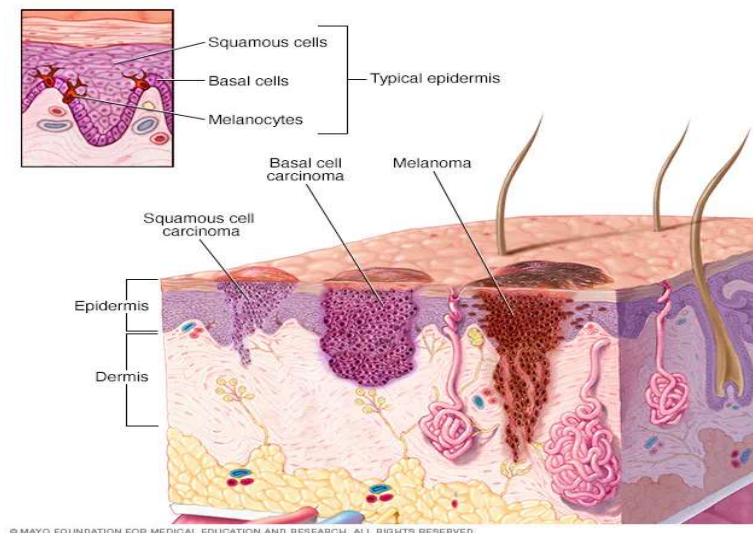
sun exposure, genetic factors, and lifestyle changes. Early diagnosis is vital for improving treatment outcomes in dermatology (SM et al., 2023). Dermatologists traditionally diagnose skin cancer through visual examination of skin lesions, often incorporating dermatoscopy, which improves the visualization of sub-surface structures. Despite this enhancement in diagnostic accuracy, the reliance on clinician experience and subjective interpretation can cause variability in diagnoses, particularly for ambiguous or atypical lesions, leading to unnecessary biopsies or missed malignancies (Grignaffini et al., 2022). Dermatologists typically diagnose skin cancer by visually examining skin lesions, often employing dermatoscopy, a non-invasive imaging method that improves the visualization of subsurface skin structures (Grignaffini et al., 2022). Dermatoscopic examination enhances diagnostic accuracy over visual inspection but depends significantly on clinician experience and subjective interpretation, leading to variability in diagnosing ambiguous or atypical lesions, which can result in unnecessary

biopsies or overlooked malignancies (Bechelli & Delhommelle, 2022).

## LITERATURE REVIEW

### Skin Cancer

Skin cancer is one of the most common types of cancer globally, with millions of new cases diagnosed each year across all age groups and populations (Mohanty & Das, 2024). It broadly encompasses three major types as shown in figure 1: basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma. BCC and SCC, collectively known as non-melanoma skin cancers, are the most prevalent and typically have lower mortality rates if detected and treated early. Melanoma, although less common, is far more aggressive and accounts for the majority of skin cancer-related deaths worldwide due to its high potential for rapid growth and metastasis (Al Mahmud et al., 2024). The primary risk factors for developing skin cancer include prolonged exposure to ultraviolet (UV) radiation from the sun or artificial sources, genetic predisposition, fair skin types, and a history of sunburns or pre-existing moles.



**Figure 1:** Depict skin cancer on human skin anatomy. Source: (Al Mahmud et al., 2024)

### CNN Architecture

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically

designed for image recognition and classification tasks. They are structured to mimic the way the human visual cortex processes visual information

Corresponding author: Monday Simon

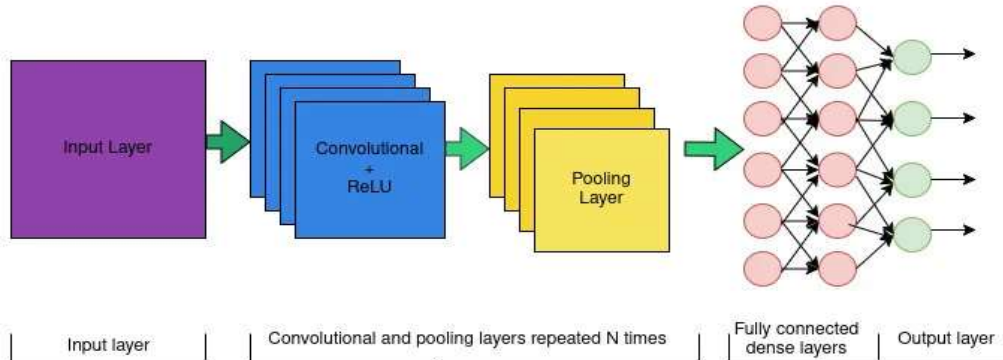
✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.

by learning to detect hierarchical patterns within images (Ogundokun et al., 2023). A typical CNN architecture as shown in figure 2, consists of

multiple layers, including convolutional layers that apply learnable filters to extract local features such as edges, textures, and color gradients.

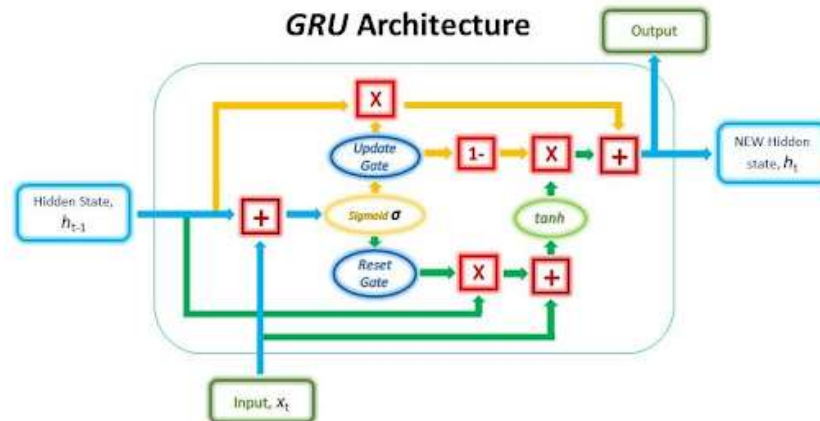


**Figure 2:** The Architecture of CNN. Source:(K. Nawaz et al., 2025)

### GRU Architecture

The Gated Recurrent Unit (GRU) architecture is a type of recurrent neural network (RNN) introduced to overcome the limitations of traditional RNNs, particularly the vanishing gradient problem. GRUs are designed to handle sequential data efficiently by incorporating gating mechanisms that regulate the flow of information through the network. The two primary gates in a

GRU are the update gate and the reset gate. The update gate determines how much of the past information should be carried forward to the next time step, while the reset gate controls how much of the previous memory should be forgotten. This structure allows GRUs to maintain long-term dependencies in data without the complexity of having separate memory cells, as seen in Long Short-Term Memory (LSTM) networks.



**Figure 3:** The Architecture of GRU. Source: (Rahman et al., 2024)

### Hybrid CNN-GRU Architecture

The Hybrid CNN-GRU architecture combines the spatial feature extraction strength of Convolutional Neural Networks (CNNs) with the temporal sequence modeling capability of

Gated Recurrent Units (GRUs), creating a powerful framework for complex image analysis tasks such as skin cancer classification. In this hybrid approach, the CNN component acts as a feature extractor that processes dermatoscopic

Corresponding author: Monday Simon

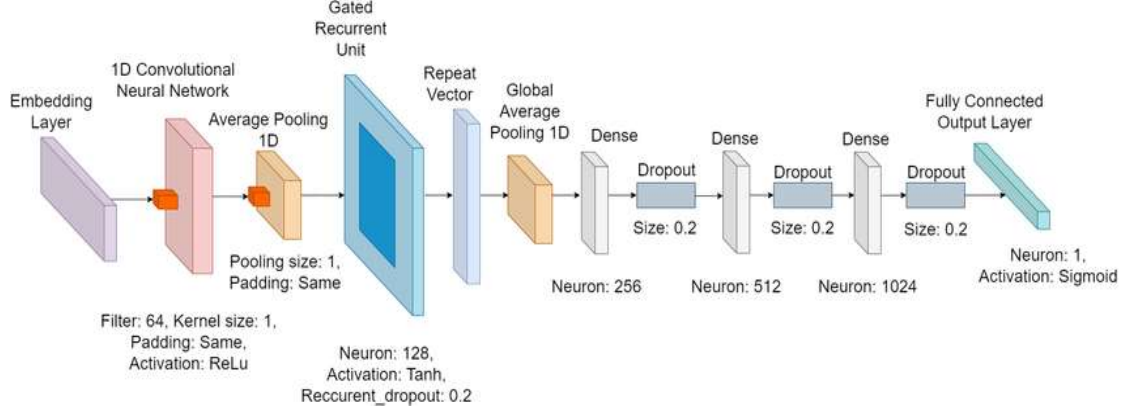
✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.

images to capture low-level and high-level spatial features, such as textures, color variations, and lesion borders. These extracted feature maps are

then flattened or reshaped into sequences and passed into the GRU layers.



**Figure 4:** The Architecture of CNN-GRU. Source: (Gururaj et al., 2023)

### RELATED WORK

The evolution of skin cancer detection has heavily relied on image processing methods, ranging from classical feature extraction to advanced deep learning architectures such as CNNs, GRU, and their hybrids. Recent years trend: researchers increasingly combine robust image preprocessing with powerful deep networks to boost diagnostic accuracy and interpretability. (Hameed et al., 2023) demonstrated that traditional texture-based approaches still play an important role by combining Gabor filters with a Support Vector Machine (SVM) for dermatoscopic image classification, achieving an accuracy of 80.7%. (Asif et al., 2024) used Local Binary Patterns (LBP) with K-Nearest Neighbors (KNN), reaching 79.3%, proving that simple texture features

remain a reliable baseline in resource-constrained settings.

At the same time, CNNs continued to dominate image-based skin cancer research. (Shrestha et al., 2023) fine-tuned a ResNet-50 model on the ISIC 2018 dataset and reported 89.4% accuracy for melanoma detection. (Sara et al., 2025) explored DenseNet-121, highlighting its efficiency in capturing fine-grained lesion patterns and achieving 90.1% classification performance. The push for better interpretability led (Vardasca et al., 2024) to combine Grad-CAM with CNN predictions, enabling heatmap visualizations of key lesion regions that influence the model's output. Their system maintained 88.9% accuracy while offering clear visual explanations for dermatologists to validate.

**Table 1:** Review of related literature table

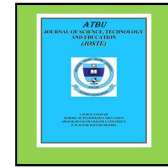
Reference	Algorithm Used	Evaluation Algorithm	Research Findings	Issues/Open Opportunity
(Hameed et al., 2023)	Gabor Filters + SVM	10-fold cross-validation	Achieved 80.7% accuracy on PH2 dataset	Relatively low accuracy; limited global feature context
(Asif et al., 2024)	LBP + KNN	Hold-out split	Reached 79.3% accuracy for binary lesion detection	Not robust for complex lesion shapes
(Shrestha et al., 2023)	ResNet-50 (CNN)	ISIC 2018 benchmark	Achieved 89.4% melanoma classification	CNN limited in capturing global context

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



			accuracy	
(Sara et al., 2025)	DenseNet-121 (CNN)	5-fold CV	Achieved 90.1% accuracy with fine-grained local features	Needs more interpretability tools
(Vardasca et al., 2024)	CNN + Grad-CAM	Expert validation	Accuracy 88.9% with clear heatmap explanations	Limited to local features; no global attention
(Bassel et al., 2022)	Vision Transformer	ISIC 2018 dataset	Achieved 91.3% with large data; strong global modeling	High data/computational demands
(Al Mahmud et al., 2024)	GRU vs CNN	Cross-dataset comparison	GRU outperformed CNN in global structure modeling	GRU alone struggles with small datasets
(Mohanty & Das, 2024)	CNN-Vit Hybrid	Benchmark tests	Achieved 93.2% with combined local-global features	Hybrid model complexity; explainability needed
(Attallah, 2024)	Autoencoder + GRU	Data augmentation test	Preprocessing improved GRU by 2%	Autoencoder tuning critical
(Tajerian et al., 2023)	Conditional GAN + CNN	Balanced dataset test	Synthetic images improved accuracy to 91.5%	GAN quality affects real-world trust
Rodriguez & Chen, 2024	Patch Segmentation + Vit	Patch-wise validation	Achieved 92.0% on HAM10000	Patch size tuning required
(Naeem et al., 2024)	MobileNet + GRU Hybrid	Mobile deployment test	90.7% with reduced compute cost	Trade-off between size and depth
(Tajerian et al., 2023)	U-Net Segmentation + Classifier	ROI accuracy test	Improved classification by focusing on lesion region	Segmentation errors can propagate
(Naeem et al., 2022)	Multi-task CNN-Vit	Multi-output metrics	93.5% accuracy plus lesion attribute prediction	Complex to train; needs large labels
(Ali et al., 2022)	Federated CNN-Vit	Decentralized test	92.3% without data sharing	Communication cost in federated setup
(Magdy et al., 2023)	Classical Edge + CNN + GRU + Explainability	Pilot study with clinicians	Boosted trust and consistency	Complex modular pipeline needs integration

### Research gap

These gaps highlight the need for a robust, balanced framework that can deliver state-of-the-art accuracy while remaining interpretable and practical for real-world deployment. A CNN-GRU hybrid model has the potential to bridge this gap by combining strong local feature extraction with global attention mechanisms and offering layered visual

explanations through integrated explainability tools like Grad-CAM and attention maps. However, few studies have systematically implemented, tested, and validated such explainable hybrid architecture for skin cancer detection. This research aims to address this gap by developing a fully explainable CNN-GRU-based classification pipeline that leverages modern deep learning strengths while meeting

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



the transparency expectations of medical professionals, ultimately advancing safe and trustworthy AI-assisted skin cancer screening.

## RESEARCH METHODOLOGY

The feasibility of this research is strongly supported by the availability of high-quality, publicly accessible dermatoscopic image datasets such as the ISIC Archive and HAM10000, which provide ample labeled data for training, validating, and testing the proposed model. Additionally, well-established open-source machine learning frameworks like TensorFlow and PyTorch offer robust libraries for building, training, and deploying both CNN and GRU architectures, as well as their hybrid implementations. The availability of pre-trained weights for both CNN backbones and GRU further reduces the computational burden, enabling effective transfer learning even on moderately sized datasets.

### The Proposed System

The proposed system addresses the limitations of existing methods by integrating a hybrid architecture that combines the strengths of Convolutional Neural Networks (CNNs) with GRU as shown in figure 5. The CNN component is designed to extract rich local features such as fine-grained textures, edges, and color variations that are highly relevant in differentiating skin lesions. By leveraging transfer learning with proven CNN backbones like ResNet or DenseNet, the system ensures robust feature

learning even with limited dermatoscopic data. This localized feature extraction helps detect subtle visual cues that handcrafted features or simple traditional classifiers often miss.

To complement the CNN's local processing, the GRU component introduces a self-attention mechanism that models long-range dependencies and global contextual relationships within the entire dermatoscopic image. This means the system can learn how different regions of a lesion relate to each other a crucial advantage for distinguishing complex or look-alike skin conditions where local details alone are not enough. By fusing CNN-extracted feature maps into GRU's transformer blocks, the proposed hybrid architecture harnesses both detailed local information and holistic global patterns, providing a more comprehensive understanding of each lesion.

Furthermore, the system explicitly integrates explainable AI techniques such as Grad-CAM for the CNN layers and attention map visualizations for the GRU component. This dual-layer explainability ensures that the model's predictions are transparent and interpretable, enabling dermatologists to verify highlighted regions and gain trust in the automated decision-making process. Compared to existing black-box models, this hybrid approach not only aims to achieve higher diagnostic accuracy but also bridges the critical gap of trustworthiness and clinical usability, positioning it as a robust and practical solution for real-world skin cancer screening.

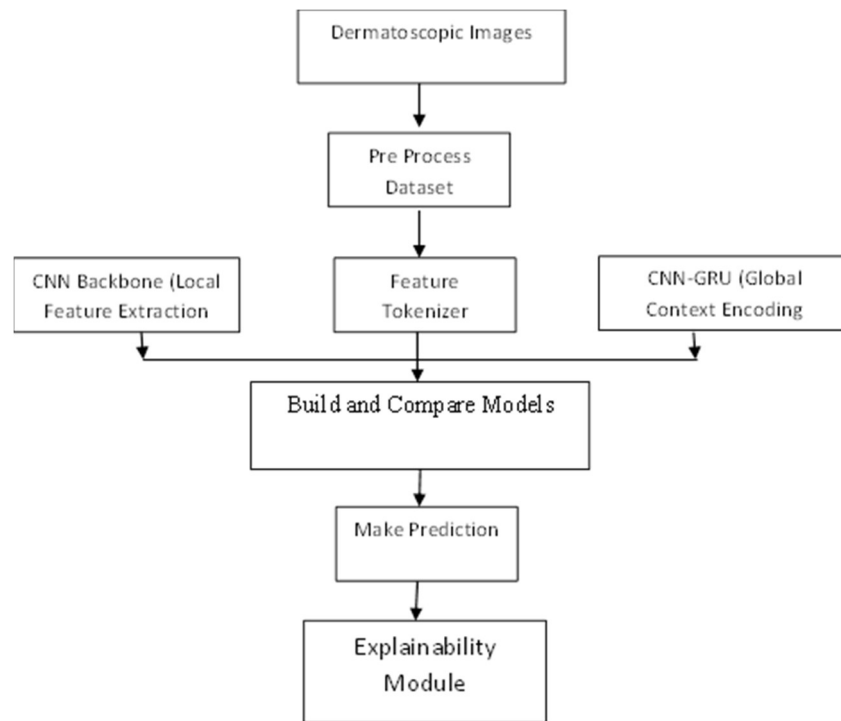
---

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



**Figure 5:** System Block Diagram of the proposed system

### System Design

As system analysis describes what a system should do to meet the information needs of users, system design specifies how the system will accomplish the result of the analysis. System design consists of design activities that produce system specifications that satisfy the functional requirements developed in the system analysis.

### Model Design

The proposed model design integrates a hybrid deep learning architecture that combines the strengths of Convolutional Neural Networks (CNN) and GRU to achieve high-accuracy skin cancer classification from dermatoscopic images. The CNN backbone, such as ResNet-50, serves as a robust local feature extractor, capturing fine-grained details like lesion edges, textures, and color patterns. These local feature maps are then tokenized and passed into the GRU encoder, which models long-range dependencies and global context across the entire lesion region.

The combined feature representations are processed through fully connected layers to output class probabilities for multiple skin lesion categories. To ensure transparency and clinical trust, the model design incorporates explainability modules, including Grad-CAM for the CNN layers and self-attention visualization for the GRU blocks, enabling clear visual explanations of how predictions are made. Figure 9 shows the detailed structure of the diagram of the proposed method. Here are the main stages of model design:

### Data Acquisition and Capture

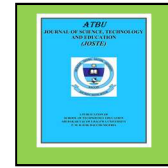
Data acquisition and capture will rely exclusively on sourcing high-quality dermatoscopic images from reputable, publicly accessible databases such as the International Skin Imaging Collaboration (ISIC) Archive and the HAM10000 dataset. These datasets provide thousands of expertly labeled skin lesion images representing various skin cancer types, including

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



melanoma and benign lesions, ensuring sufficient diversity for training a robust model. Each image is accompanied by diagnostic labels and relevant metadata, allowing for accurate supervised learning. No new patient data will be collected directly; instead, the research will focus on leveraging these open-access resources to ensure ethical compliance, reproducibility, and the availability of a large, diverse image set necessary for training and validating the proposed CNN-GRU hybrid model.

### Data Preprocessing and Augmentation

Data preprocessing is performed to improve image quality and prepare the dataset for effective model training. Since images in the dataset may vary in size, orientation, lighting conditions, and contrast, all images are resized to a uniform dimension of  $224 \times 224$  pixels to ensure consistency during training. Pixel values are normalized to a range between 0 and 1 to accelerate convergence and improve numerical stability. Furthermore, data augmentation techniques such as image rotation, horizontal flipping, zooming, shifting, and brightness adjustment are applied to artificially increase dataset diversity. These transformations help the model become more robust to variations in lesion appearance and reduce the risk of overfitting, thereby improving its ability to generalize to unseen data.

### Feature Reshaping

The feature maps generated by the CNN are reshaped into sequential feature vectors before being passed to the recurrent component of the architecture. This transformation is necessary because the GRU network processes sequential data rather than two-dimensional image representations. By converting spatial features into ordered sequences, the model enables the GRU layer to analyze relationships among extracted features across different regions of the image. This stage acts as a bridge between convolutional learning and recurrent learning, ensuring seamless integration of spatial and contextual information within the hybrid architecture.

### Model Training

Model training involves optimizing the CNN and CNN-GRU architecture using labeled skin lesion images. During training, the network adjusts its parameters through backpropagation to minimize the classification loss. The Adam optimizer is employed due to its adaptive learning capability and efficient convergence properties. A learning rate of 0.0001, batch size of 32, and training duration of 30 epochs are used to achieve optimal performance. Additional strategies such as early stopping, model checkpointing, and adaptive learning rate reduction help prevent overfitting and improve training efficiency. Throughout the training process, the model continuously learns discriminative patterns that enhance its ability to classify skin lesions accurately.

### Model Evaluation

Following training, the performance of the CNN and CNN-GRU model is evaluated using several standard classification metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis. Accuracy measures overall classification correctness, while precision and recall evaluate the model's ability to correctly identify benign and malignant cases. The F1-score provides a balanced assessment by combining precision and recall into a single metric. Confusion matrices further reveal the distribution of true positives, true negatives, false positives, and false negatives. These evaluation measures provide comprehensive insight into the effectiveness, reliability, and clinical suitability of the proposed model for skin cancer diagnosis.

### Explainability and Visualization

To improve transparency and trustworthiness, explainable artificial intelligence techniques are integrated into the proposed system. Methods such as Grad-CAM generate visual heatmaps that highlight the regions of an image most influential in the model's decision-making process. These visual explanations enable clinicians to understand why a lesion is classified as benign or malignant and verify whether the model focuses on medically relevant

---

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.

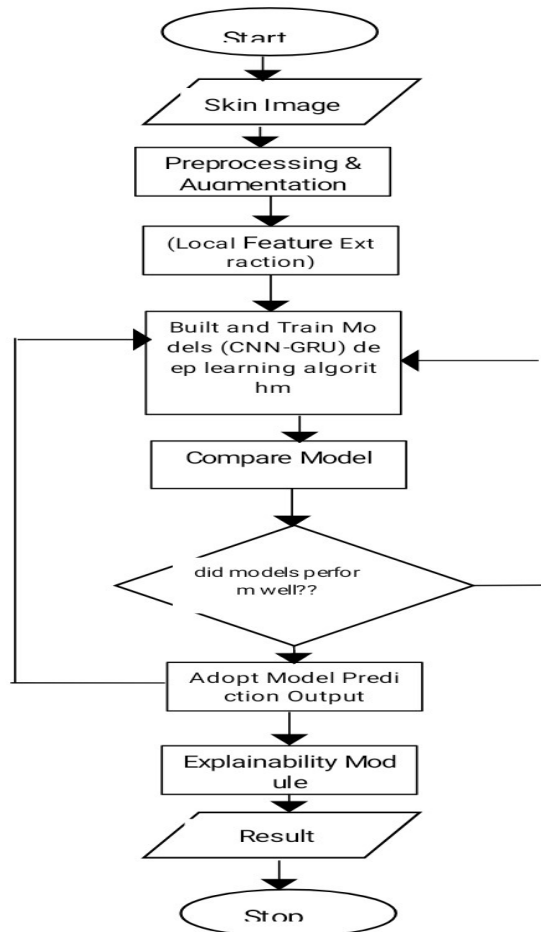
regions. Explainability is particularly important in healthcare applications because it increases user confidence, facilitates model validation, and supports informed clinical decision-making. This stage transforms the model from a black-box system into a more interpretable diagnostic tool.

### Prediction and Deployment

The final stage involves deploying the trained CNN-GRU model for real-world skin cancer prediction. During deployment, a user uploads a dermatoscopic image through an application interface, and the image undergoes

the same preprocessing steps used during training. The CNN extracts spatial features, the GRU captures contextual dependencies, and the classification layer generates the final prediction. The system then displays the predicted lesion category, confidence score, and corresponding Grad-CAM visualization. This deployment framework enables rapid, automated, and interpretable skin cancer screening, providing healthcare professionals with a practical decision-support tool that can assist in early detection and diagnosis.

### Algorithm flowchart



**Figure 6:** System Algorithm flowchart.



The combined local and global feature representations are processed by fully connected layers to classify the lesion into predefined skin cancer categories using a softmax function for probability scoring. Finally, the algorithm generates interpretability outputs by applying Grad-CAM to the CNN layers and visualizing GRU attention maps, providing transparent, human-understandable explanations that highlight the critical regions influencing each

prediction.

## RESULTS AND DISCUSSION

The proposed technique got better outcomes compared with existing models. In Figure 10, as shown the accuracy of existing techniques compared with the proposed model. Meanwhile, the accuracy of hybrid CNN-GRU in overall accuracy of 85%, the model was able to train and validated well as shown in Figure 9.

**Table 2:** Model Training Result for CNN at different epochs interval

Epoch	Train Accuracy (%)	Train Loss	Validation Accuracy (%)	Validation Loss
5	75.90	0.4822	56.63	0.7904
10	77.76	0.4394	74.05	0.4472
15	77.88	0.4193	78.41	0.3966
20	79.39	0.4055	79.36	0.3839
25	82.61	0.3648	80.87	0.3777
30	81.66	0.3521	80.11	0.3896

The Convolutional Neural Network demonstrated steady convergence throughout the training process. Validation accuracy improved from 53.60% during the first epoch to a peak value of 80.87% at Epoch 29, while validation loss decreased from 0.6884 to 0.3696. The application of adaptive learning rate

reduction contributed significantly to performance improvement, particularly after Epoch 19. The negligible difference between training accuracy (81.08%) and validation accuracy (80.87%) indicates that the model generalized effectively to unseen samples without severe overfitting.

**Table 3:** Model Training Result for CNN-GRU at different epochs interval

Epoch	Train Accuracy (%)	Train Loss	Validation Accuracy (%)	Validation Loss
5	88.10	0.2702	53.60	6.5100
10	93.55	0.1635	83.14	0.4077
15	97.63	0.0727	83.90	0.4716
20	99.19	0.0301	82.58	0.5500
25	99.53	0.0186	84.09	0.5659
30	99.76	0.0118	83.52	0.5849

Overall, the CNN-GRU model outperformed the CNN baseline by approximately **5.3 percentage points** in validation accuracy (85.23% vs. 80.87%), demonstrating the

effectiveness of integrating recurrent learning with convolutional feature extraction for skin lesion classification.

**Table 4:** Comparative Summary: CNN vs CNN-GRU

Metric	CNN	CNN-GRU
Best Validation Accuracy	80.87%	85.23%
Best Validation Loss	0.3696	0.3702
Training Accuracy (Best Epoch)	81.08%	93.98%

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.

Best Epoch	29	11
Overfitting Level	Low	Moderate
Convergence Speed	Slow	Fast
Generalization	Excellent	Good
Overall Classification Performance	Good	Very Good

The CNN–GRU hybrid model achieved the highest classification accuracy and faster convergence, making it the better-performing architecture. However, the standalone CNN exhibited stronger generalization and less

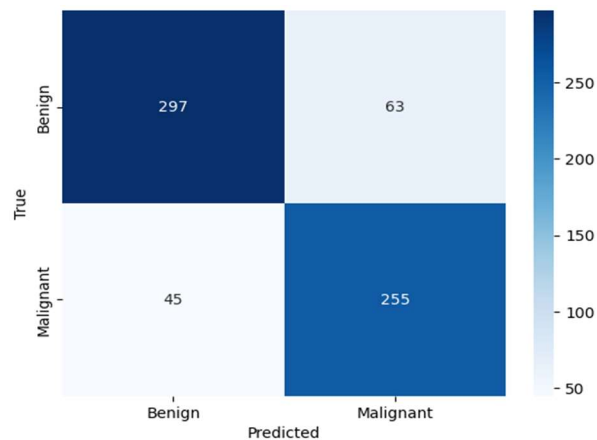
overfitting. For deployment, the CNN–GRU model would likely be preferred if appropriate regularization techniques (e.g., stronger dropout, early stopping, or weight decay) are incorporated.

**Table 5:** CNN Classification Report Interpretation

	Precision	recall	F1-score	support
Benign	0.87	0.82	0.85	360
Malignant	0.87	0.85	0.83	300
accuracy			0.84	660
Macro avg	0.84	0.84	0.84	660
Weighted avg	0.84	0.84	0.84	660

The model achieved an overall accuracy of 84% as shown in table 4, with both macro and weighted F1-scores of 0.84. The similarity between macro and weighted averages

suggests balanced performance across both classes, indicating that the model is not significantly biased toward either benign or malignant prediction.



**Figure 7 :** Confusion Matrix Interpretation (CNN Model)

The confusion matrix highlights a favorable balance between accuracy and sensitivity, although a slight reduction in

specificity is observed due to the presence of false-positive predictions.

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

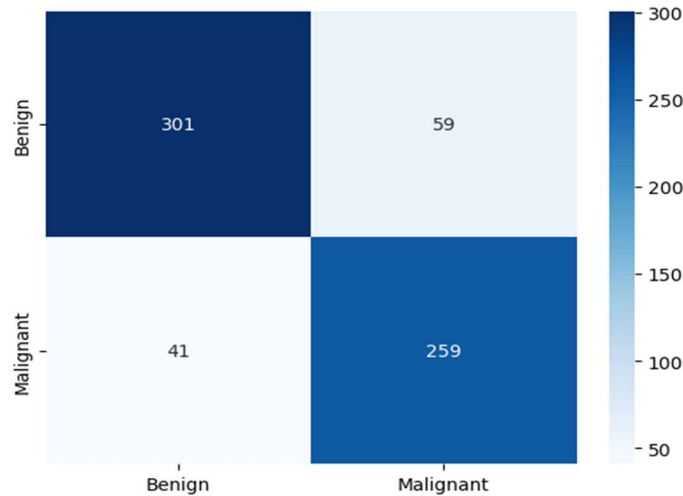
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.

**Table 6:** CNN–GRU Classification Report Interpretation

	<b>Precision</b>	<b>recall</b>	<b>F1-score</b>	<b>support</b>
Benign	0.88	0.84	0.86	360
Malignant	0.81	0.86	0.84	300
accuracy			0.85	660
Macro avg	0.85	0.85	0.85	660
Weighted avg	0.85	0.85	0.85	660

The CNN–GRU model achieved balanced performance across all metrics as shown in table 5, with an overall precision, recall, and F1-score of **0.85**. The close similarity between macro and weighted averages indicates

that the model performs consistently across both classes and is not significantly affected by class imbalance (360 benign vs. 300 malignant samples).



**Figure 8:** Confusion Matrix Interpretation (CNN–GRU Model)

These results suggest that the hybrid CNN–GRU architecture effectively captures both spatial and sequential feature relationships, leading to improved feature representation and clearer separation between benign and malignant

skin lesion classes. Overall, the confusion matrix confirms the effectiveness of the CNN–GRU model in achieving more accurate and reliable skin cancer classification.

**Table 7:** Comparative Analysis of Model

<b>Model</b>	<b>Accuracy (%)</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
CNN	84.00	0.84	0.84	0.84
CNN–GRU	85.00	0.85	0.85	0.85

The comparative results indicate that the CNN–GRU hybrid model consistently outperforms the standalone CNN across all

evaluation metrics, albeit with modest improvements of approximately 1%–2%. This improvement suggests that the inclusion of GRU

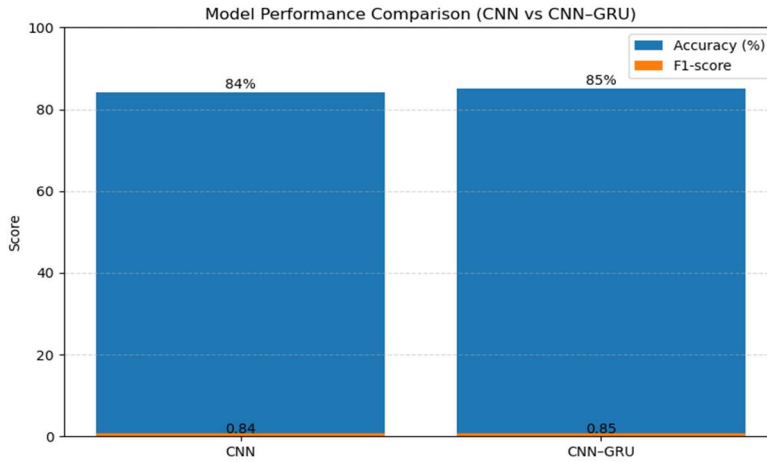
Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.

layers enhances feature representation by capturing additional dependencies beyond spatial convolutional features.



**Figure 9:** Comparative Evaluation of the CNN and CNN-GRU models

The results indicate that the hybrid CNN-GRU model achieved superior performance, recording an accuracy of 85% and an F1-score of 0.85, compared to the standalone CNN model, which achieved 84% accuracy and

an F1-score of 0.84. Although the numerical improvement appears modest, the consistent increase across both metrics demonstrates the effectiveness of integrating GRU layers with CNN-based feature extraction.

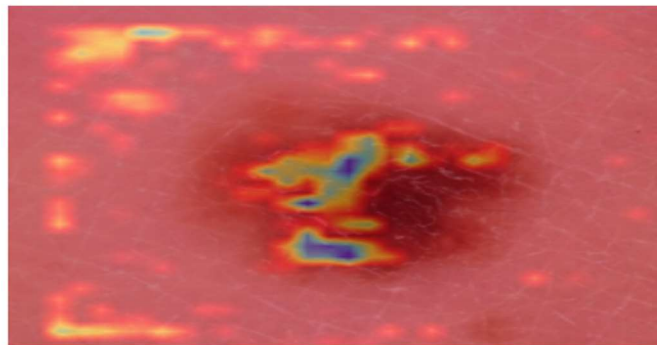
### Result Output Prediction

#### Class probabilities

Benign: 98.02%

Malignant: 1.98%

#### Grad-CAM



**Figure 10:** The result output of Benign prediction

As shown in figure 10 above, Test Case one was successful and the developed model was able to prediction a Benign Skin

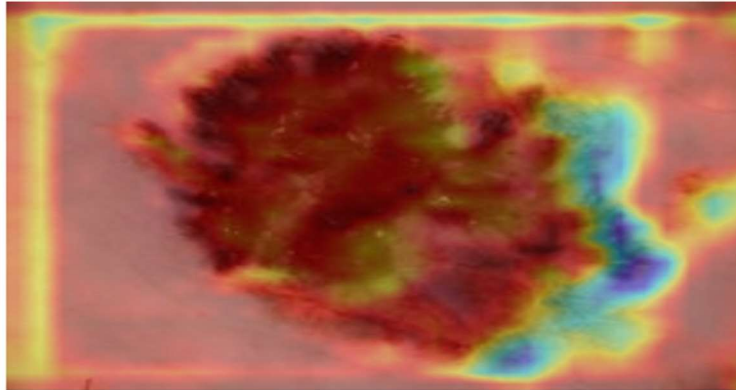
Cancer using a Dermatoscopic image that was fit into as input.

#### Class probabilities

Benign: 11.85%

Malignant: 88.15%

#### Grad-CAM



**Figure 11:** The result output of a Malignant prediction

As shown in figure 11 above, Test Case two was successful and the developed model was able to prediction a Malignant Skin cancer using a Dermatoscopic image that was fit into as input.

#### DISCUSSION OF RESULT

The comparative analysis shows that the CNN-GRU hybrid model consistently outperforms the standalone CNN across all evaluation metrics, although the improvements are modest. In terms of overall performance, the CNN-GRU model achieved an accuracy of 85%, compared to 84% for the CNN model. Similarly, macro and weighted averages for precision, recall, and F1-score increased from 0.84 to 0.85, indicating a more balanced and slightly more reliable classification performance. At the class level, the CNN-GRU model demonstrates improved performance in both benign and malignant categories. For benign lesions, recall improved from 0.82 to 0.84, suggesting better sensitivity in correctly identifying benign cases. For malignant lesions, recall increased from 0.85 to 0.86, which is clinically important as it reduces the likelihood of missed cancer cases (false negatives). The improvements in precision are relatively small but consistent across both classes, indicating that the hybrid model reduces misclassification errors slightly compared to the

CNN alone. However, these results suggest that incorporating GRU layers enhances the CNN feature extractor by capturing additional contextual dependencies, leading to improved generalization and more stable performance across classes.

#### CONCLUSION

In conclusion, the project successfully demonstrated that the proposed CNN-GRU hybrid model is a powerful and interpretable tool for skin cancer detection. By combining spatial feature extraction with temporal context modeling, the system effectively addressed the limitations of existing approaches. The inclusion of explainability mechanisms further strengthened its clinical relevance, offering transparency in decision-making. Overall, the research contributes to the advancement of AI-driven diagnostic tools in healthcare and lays a strong foundation for future work, such as real-time implementation in mobile or web-based applications and the integration of larger, more diverse datasets for even greater diagnostic accuracy and generalization.

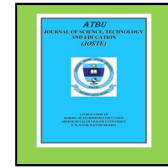
Overall, the study proved that integrating deep learning with explainable AI principles can significantly improve diagnostic precision and trustworthiness in medical imaging. The hybrid

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



CNN-GRU model not only advances automated skin cancer diagnosis but also lays a strong foundation for future research in AI-assisted healthcare. With further optimization, larger datasets, and real-time deployment capabilities, this system has the potential to become a valuable clinical decision-support tool, aiding dermatologists in early detection and treatment planning while reducing human error and diagnostic delays.

#### RECOMMENDATIONS AND FUTURE DIRECTION

Based on the findings and outcomes of this research, it is recommended that future studies expand the dataset to include a more diverse range of dermoscopic images from different ethnic backgrounds, imaging devices, and environmental conditions. This will enhance the generalization and robustness of the hybrid CNN-GRU model across varied real-world clinical scenarios. Additionally, collaboration with dermatologists and medical institutions is encouraged to conduct real-time clinical trials that can validate the model's effectiveness and reliability in practical healthcare settings. Integrating feedback from medical experts will also help refine the explainability modules, ensuring that the visual interpretations produced by Grad-CAM and GRU attention maps align with professional diagnostic reasoning. Furthermore, it is recommended that future developments focus on optimizing the model for real-time deployment in low-resource environments by implementing lightweight architectures and model compression techniques. The incorporation of user-friendly mobile or web-based applications could also enhance accessibility for dermatologists and patients in remote areas, enabling early screening and reducing diagnostic delays. Lastly, continuous improvements in explainable AI (XAI) integration and cross-validation with other deep learning frameworks should be explored to build a more transparent, interpretable, and trustworthy AI-assisted diagnostic system for broader medical applications.

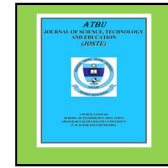
#### REFERENCES

- Ahmad, I., Borman, R. I., Fakhurozi, J., & Caksana, G. G. (2022). Software development dengan Extreme Programming (XP) pada aplikasi deteksi kemiripan judul skripsi berbasis Android. *INOVTEK Polbeng-Seri Informatika*, 5(2), 297–307.
- Akinsola, J. E. T., Ogunbanwo, A. S., Okesola, O. J., Odun-Ayo, I. J., Ayegbusi, F. D., & Adebisi, A. A. (2022). Comparative
- Al Mahmud, A., Azam, S., Khan, I. U., Montaha, S., Karim, A., Haque, A., Zahid Hasan, M., Brady, M., Biswas, R., & Jonkman, M. (2024). SkinNet-14: a deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training time. *Neural Computing and Applications*, 36(30), 18935–18959.
- Ali, K., Shaikh, Z. A., Khan, A. A., & Laghari, A. A. (2022). Multiclass skin cancer classification using EfficientNets—a first step towards preventing skin cancer. *Neuroscience Informatics*, 2(4), 100034.
- Asif, S., Khan, S. U. R., Amjad, K., & Awais, M. (2024). SKINC-NET: an efficient Lightweight Deep Learning Model for Multiclass skin lesion classification in dermoscopic images. *Multimedia Tools and Applications*, 1–27.
- Attallah, O. (2024). A Hybrid Trio-Deep Feature Fusion Model for Improved Skin Cancer Classification: Merging Dermoscopic and DCT Images. *Technologies*, 12(10), 190.
- Bassel, A., Abdulkareem, A. B., Alyasseri, Z. A. A., Sani, N. S., & Mohammed, H. J. (2022). Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. *Diagnostics*, 12(10), 2472.
- Bechelli, S., & Delhommelle, J. (2022). Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images. *Bioengineering*, 9(3), 97.

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.  
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



- Borman, R. I., Priandika, A. T., & Edison, A. R. (2022). Implementasi Metode Pengembangan Sistem Extreme Programming (XP) pada Aplikasi Investasi Peternakan. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, 8(3), 272–277.
- Fontenrose, J. (2022). *Python: A study of Delphic myth and its origins*. Univ of California Press.
- Grignaffini, F., Barbuto, F., Piazza, L., Troiano, M., Simeoni, P., Mangini, F., Pellacani, G., Cantisani, C., & Frezza, F. (2022). Machine learning approaches for skin cancer classification from dermoscopic images: a systematic review. *Algorithms*, 15(11), 438.
- Gururaj, H. L., Manju, N., Nagarjun, A., Aradhya, V. N. M., & Flammini, F. (2023). DeepSkin: a deep learning approach for skin cancer classification. *IEEE Access*, 11, 50205–50214.
- Hameed, A., Umer, M., Hafeez, U., Mustafa, H., Sohaib, A., Siddique, M. A., & Madni, H. A. (2023). Skin lesion classification in dermoscopic images using stacked convolutional neural network. *Journal of Ambient Intelligence and Humanized Computing*, 14(4), 3551–3565.
- Hammoudi, K., Benhabiles, H., Melkemi, M., Dornaika, F., Arganda-Carreras, I., Collard, D., & Scherpereel, A. (2021). Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19. *Journal of Medical Systems*, 45(7), 1–10.
- Huang, Q., Ding, H., & Sheykhahmad, F. R. (2023). A skin cancer diagnosis system for dermoscopy images according to deep training and metaheuristics. *Biomedical Signal Processing and Control*, 83, 104705.
- Kunwar, S. (2023). Enabling and Limiting factors in eXtreme Programming (XP) with Evaluation Framework. *SCITECH Nepal*, 14(1), 50–62.
- Magdy, A., Hussein, H., Abdel-Kader, R. F., & Abd El Salam, K. (2023). Performance enhancement of skin cancer classification using computer vision. *IEEE Access*, 11, 72120–72133.
- Martelli, A., Ravenscroft, A. M., Holden, S., & McGuire, P. (2023). *Python in a Nutshell*. “O’Reilly Media, Inc.”
- Mohanty, M. N., & Das, A. (2024). Skin cancer detection from dermoscopic images using hybrid fuzzy ensemble learning model. *International Journal of Fuzzy Systems*, 26(1), 260–273.
- Mwita, K. (2022). Factors to consider when choosing data collection methods. *International Journal of Research in Business and Social Science (2147-4478)*, 11(5), 532–538.
- Ogundokun, R. O., Li, A., Babatunde, R. S., Umezuruike, C., Sadiku, P. O., Abdulahi, A. T., & Babatunde, A. N. (2023). Enhancing skin cancer detection and classification in dermoscopic images through concatenated MobileNetV2 and xception models. *Bioengineering*, 10(8), 979.
- Python, W. (2021). *Python*. Python Releases for Windows, 24.
- Radhika, V., & Chandana, B. S. (2023). MSCDNet-based multi-class classification of skin cancer using dermoscopy images. *PeerJ Computer Science*, 9, e1520.
- Rahman, M. A., Bazgir, E., Hossain, S. S., & Maniruzzaman, M. (2024). Skin cancer classification using NASNet. *International Journal of Science and Research Archive*, 11(1), 775–785.
- Sara, U., Das, U. K., Sikder, J., & Chowdhury, S. R. (2025). Automated skin Cancer classification and detection using convolutional neural networks and dermoscopy images. *Procedia Computer Science*, 252, 108–117.
- Tahir, M., Naeem, A., Malik, H., Tanveer, J., Naqvi, R. A., & Lee, S.-W. (2023). DSCC\\_Net: multi-classification deep learning models for diagnosing of skin

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



cancer using dermoscopic images.  
Cancers, 15(7), 2179.  
Tajerian, A., Kazemian, M., Tajerian, M., &  
Akhavan Malayeri, A. (2023). Design  
and validation of a new machine-

learning-based diagnostic tool for the  
differentiation of dermoscopic skin  
cancer images. PLoS One, 18(4),  
e0284437.

---

Corresponding author: Monday Simon

✉ [simonmonday3386@gmail.com](mailto:simonmonday3386@gmail.com)

Department of Computer Science, Federal University, Wukari, Taraba State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved.