GMR
The Original

HYBRID INTELLIGENCE SYSTEM CNN-LSTM ARCHITECTURE FOR ACCURATE SKIN CANCER CLASSIFICATION USING DERMOSCOPIC IMAGE ANALYSIS

Tejaswi Rupakula[1*], Murugadoss R[2], Karthikeyan A[3], Aruna M[4], Prabhu Ranjith Periyasamy[5], Kalaivani Krishnan[6]

[1*]Department of Computer Science Engineering, Odessa College, Odessa, United States - TX 79764.
[2]Department of Artificial Intelligence and Data Science, VSB College of Engineering Technical Campus, Coimbatore, Tamil Nadu, India-642109.
[3]Department of Aeronautical Engineering, Excel Engineering College, Komarapalayam, Tamil Nadu, India- 637303.
[4]Department of Electrical and Electronics Engineering, Meenakshi College of Engineering, Chennai, Tamil Nadu, India - 600078.
[5]Department of Artificial Intelligence and Data Science, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India – 638012
[6]Department of Computer Science and Information Engineering, Anna University, Chennai, Tamil Nadu, India - 600025.
Corresponding E-mail : phdannauniv2020@gmail.com[1]

ABSTRACT

Background: Skin cancer, including malignant melanoma and non-melanoma skin lesions, continues to pose a significant global health challenge due to delayed diagnosis and limited access to specialized dermatological expertise. Recent advances in artificial intelligence, particularly deep learning-based dermoscopic image analysis, have demonstrated considerable potential for improving the early detection and diagnosis of skin cancer.

Objectives: This study aims to develop a novel hybrid deep learning framework that integrates Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to enhance the classification accuracy of dermoscopic skin lesion images by effectively capturing both spatial and contextual feature representations.

Methods: The proposed hybrid CNN-LSTM model was evaluated using the Derm7pt dataset, comprising 8,000 dermoscopic images spanning multiple diagnostic categories. Image preprocessing involved color constancy correction, contrast enhancement, and artifact removal to standardize image quality. Pre-trained CNN architectures, including ResNet50 and EfficientNet-B0, were employed for feature extraction, while LSTM layers modeled sequential dependencies within the extracted feature vectors to capture long-range contextual information. The proposed model was trained and evaluated using standard performance metrics.

Results: Experimental results demonstrate that the hybrid CNN-LSTM framework outperforms standalone CNN models in skin lesion classification. The proposed architecture achieved improved classification accuracy, sensitivity, and F1-score, particularly for minority lesion classes, indicating superior capability in distinguishing visually similar skin lesions while reducing misclassification rates.

Conclusion: The proposed CNN-LSTM architecture effectively combines spatial and sequential feature learning to enhance automated skin cancer classification from dermoscopic images. The improved diagnostic performance highlights its potential for integration into clinical decision support systems, enabling scalable, reliable, and early detection of skin cancer, thereby supporting dermatologists in improving patient outcomes.

KEYWORDS: Skin lesion classification, Hybrid deep learning, CNN-LSTM architecture, Derm7pt dataset, Dermoscopic image analysis, Medical image processing

1. INTRODUCTION

Medical diagnostics are being revolutionized by artificial intelligence (AI) techniques for skin cancer classification that improve accuracy, efficiency, and early detection. Traditional diagnostic methods, such as dermatologists' visual examinations and histopathological analyses, are usually time-consuming and subjective, producing inconsistent results, particularly in cases of early malignant lesions. Artificial Intelligence (AI) systems that use deep learning models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks show great promise in automating clinical and dermoscopic image analysis. By accurately differentiating between benign nevi, basal cell carcinoma, and melanoma, among other skin lesions, these models can identify complex patterns in massive image datasets. Using cloud and mobile technologies, AI applications also assist doctors in lowering diagnostic errors and enabling scalable solutions in remote areas (Abohashish et al., 2025). As the number of melanoma cases worldwide increases, it is crucial to incorporate AI into the classification of skin cancer in order to improve patient outcomes by reducing the number of biopsies and implementing customized treatment plans in a timely manner (Putri et al., 2024). The hybrid CNN-LSTM model

demonstrated significant improvements over conventional CNN-based skin disease detection systems by achieving higher sensitivity, precision, and F1-score (Basholli et al., 2025).
However, skin conditions remain a major global health concern, particularly pigmented lesions and melanoma. Automated classification of pigmented skin lesions has gained considerable attention through the use of the HAM10000 dataset, which comprises 10,015 dermatoscopic images categorized into seven diagnostic classes. Hybrid deep learning architectures employing advanced transfer learning and preprocessing techniques have significantly enhanced lesion classification performance (Shaik et al., 2025). Comparative studies involving baseline CNN models and LSTM-augmented architectures reveal that LSTM integration substantially improves classification accuracy by capturing contextual dependencies, particularly for underrepresented lesion classes (Packiamary & Muthukumaravel, 2024). Furthermore, hybrid architectures integrating CNNs, Bidirectional Long Short-Term Memory (BiLSTM), and attention mechanisms have demonstrated remarkable improvements in multiclass skin lesion classification (Nawrin et al., 2023).
Early detection of melanoma remains one of the most effective strategies for reducing mortality and improving patient survival. Recent developments in hybrid deep learning frameworks have strengthened healthcare resilience by enhancing automated diagnostic accuracy (Lilhore et al., 2025). To further improve melanoma classification, hybrid deep learning models combining CNNs, LSTM networks, and attention mechanisms have been proposed, enabling more robust feature extraction and contextual learning (Rajput & Solanki, 2025). Skin cancer, caused by abnormal proliferation of epidermal cells, continues to increase globally, necessitating accurate computer-aided diagnostic systems (Roy et al., 2024). Conventional diagnostic approaches such as dermoscopy and spectroscopy are often expensive, time-consuming, and dependent on clinical expertise (Maniraj & Thamizhamuthu, 2025).
The subtle visual differences among various skin lesions make accurate classification particularly challenging. Nevertheless, deep learning techniques have significantly improved diagnostic performance and, in several studies, have achieved accuracy comparable to or even exceeding that of experienced dermatologists (Ahmad et al., 2022). Although numerous pretrained CNN models and hybrid classification frameworks have been investigated, many still exhibit moderate classification performance. Hybrid approaches incorporating CNN-LSTM-SVM architectures have reported improved classification accuracy of approximately 88%, demonstrating the effectiveness of combining spatial and sequential feature learning (Mehta & Kaur, 2024). More recently, the SkinEHDLF framework integrated ConvNeXt, EfficientNetV2, and Swin Transformer architectures using adaptive attention-based feature fusion to achieve highly accurate skin cancer classification (Lilhore et al., 2025). Likewise, CNN-LSTM models have been successfully applied to dermatological diagnosis, including nail disease classification, by effectively modeling both spatial and sequential features (Roy et al., 2024). Multi-scale attention-enhanced hybrid CNN-LSTM models have also shown promising improvements in melanoma detection from dermoscopic images (Maniraj & Thamizhamuthu, 2025).
These advancements have considerably enhanced clinical diagnosis and telemedicine-based healthcare services, with future research focusing on expanding the classification of diverse skin lesion categories. Skin cancer continues to affect millions of individuals worldwide each year, making early diagnosis increasingly important. While manual diagnosis through dermoscopy remains essential, it is often labor-intensive and subject to diagnostic variability. Consequently, computer-aided diagnostic systems integrating deep convolutional neural networks with Bidirectional Long Short-Term Memory (BLSTM) networks have emerged as effective solutions for improving biomedical image analysis and automated skin lesion classification (Prabhakara et al., 2024; Obayya et al., 2023).

2. METHODOLOGY

2.1 Data Collection

The classification model can be effectively generalized to the diverse population represented in the Derm7pt dataset used in this study which includes a wide range of pigmented skin lesions given in Table 1.

Table 1 Dataset evaluation

Attribute	Description
Dataset Name	Derm7pt
Total Images	8,000
Image Type	Dermoscopic
Format	RGB JPEG (512x512 resolution)
Diagnostic Categories	Melanoma, Basal Cell Carcinoma, Nevus, Seborrheic Keratosis, Dermatofibroma, Vascular Lesions
Annotation Type	Categorical Diagnosis
Source	Publicly available via ISIC archive

In order to facilitate clinical interpretability and algorithmic supervision each dermoscopic image is annotated according to diagnostic categories and lesion characteristics. A high-level summary of the dataset configuration is given in Figure 1.

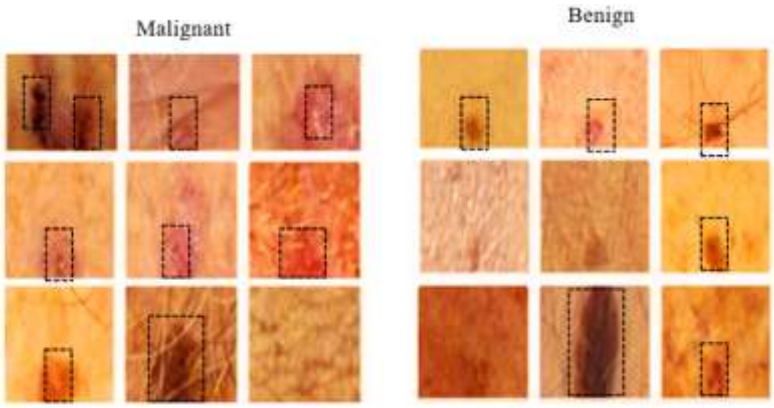

Figure 1. Overview of image category in the Derm7pt dataset

2.2 Class Distribution

Addressing class imbalance which is prevalent in medical datasets because some diseases are uncommon is crucial when creating a classification model. Table 2 below displays the class distribution for the dataset which indicates a moderate imbalance with vascular lesions and melanoma being underrepresented.

Table 2 class distribution

Class Name	Number of Images
Melanoma	1,000
Basal Cell Carcinoma	1,100
Nevus	2,800
Seborrheic Keratosis	1,200
Dermatofibroma	900
Vascular Lesions	1,000

Because of this imbalance oversampling methods and modifications to the loss function were required during training (Figure 2).

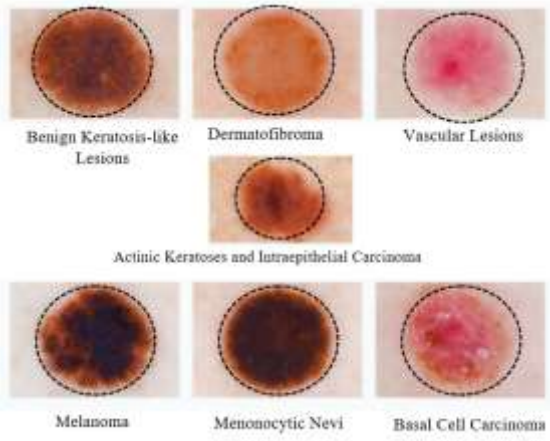

Figure 2. Sample images classification

2.3 Data Preprocessing

Dermoscopic images were rigorously pre-processed to remove noise and improve quality before model training. DullRazor filtering was used to remove hair artifacts adaptive histogram equalization was used to improve contrast and the Shades-of-Gray algorithm was used to correct for illumination irregularities. In order to ensure consistent feature learning images were resized to 224x224 pixels and normalized. To improve dataset variety and model robustness data augmentation techniques like zooming random rotation horizontal flipping and brightness adjustment were used. Lastly images were batch-processed to NumPy tensors for CNN-LSTM processing.

2.4 Feature Selection

Deep representation learning using pre-trained CNN backbones was used to handle the feature extraction and selection instead of manual selection. The CNN bottleneck layers feature maps were examined to assess the dimensional characteristics of the learned features prior to feeding them into the LSTM. Features with a higher explained variance were kept for interpretability in the early exploratory stages when latent structure was

visualized using Principal Component Analysis (PCA). The following table displays the main feature categories along with their sources (Table 3).

Table 3. Feature selection

Feature Type	Extraction Method	Dimensionality
Texture Descriptors	CNN Bottleneck (ResNet50)	2048
Color Histogram Vectors	Preprocessing Module	128
Shape & Border Metrics	CNN Intermediate Layers	512
Sequential Embeddings	LSTM Hidden States	256

2.5 Model Validation

For training and validation the model used a stratified 5-fold cross-validation which ensured balanced class representation and consistent performance across different class distributions. Overfitting was avoided by early stopping based on validation loss. To counteract dominant classes a categorical cross-entropy loss function was employed adjusting for class weight. For stable convergence the Adam optimizer was applied using a batch size of 32 over 50 epochs with a learning rate of 0. 0001 and a decay rate of 1e-6. The model was improved with checkpoints dropout regularization (rate = 0. 5) and learning rate reduction on plateau. Final performance was evaluated using a test fold that was withheld and results were averaged for statistical robustness.

2.7 Architectural Framework

The proposed architectural framework is a hybrid deep learning design that integrates both spatial and temporal learning by combining pre-trained Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks. The framework is built upon the CNN backbone ResNet50 or EfficientNet-B0 which processes dermoscopic images to extract high-level spatial features. Their potent capacity to represent and capture hierarchical image features using multiple convolutional layers led to the selection of these CNNs. After the feature maps are flattened into structured vectors they are reformulated as sequential inputs to aid in temporal learnin (Figure 3).

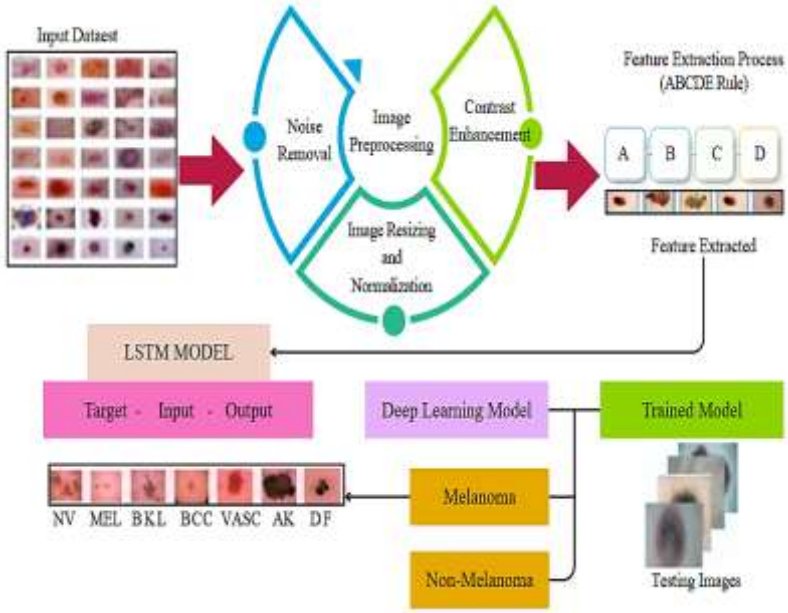

Figure 3. Architecture framework

This transformation allows the model to capture deeper contextual correlations shared by visually similar skin lesions. Before the LSTM a Temporal Self-Attention Layer is added to enhance sequence modeling and allow the model to focus on significant feature segments while reducing noise. By processing the refined sequence and logging long-term dependencies the LSTM layer improves the models ability to differentiate between various lesion types. After the final output from the LSTM is sent to a fully connected dense layer which condenses the sequence into a compact form suitable for classification class probabilities for six skin lesion categories are generated using a softmax activation function. Batch normalization is used to stabilize training and dropout regularization is used throughout the architecture to prevent overfitting. The model is implemented using TensorFlow which makes use of GPU acceleration

2.8 Proposed Technique: Hybrid CNN-LSTM Framework for Skin Lesion Classification

The suggested hybrid CNN-LSTM framework combines temporal modeling and spatial feature extraction to accurately classify skin lesions such as benign nevi basal cell carcinoma and melanoma. Three stages make up the

architecture: (1) CNN-based feature extraction using models such as ResNet50 or EfficientNet-B0 (2) LSTM-based temporal modeling of the extracted features and (3) classification using dense layers with softmax activation. In order to improve decision-making for spotting minute variations in lesions this technique records deep visual features and their contextual relationships in equation 1 to 6.

Equation 1: Input Normalization
Before feeding into the CNN, input image normalization ensures faster convergence and uniform distribution of pixel intensities.

$$X_{\text{norm}} = \frac{X - \mu}{\sigma} \quad (1)$$

Where X is the raw image pixel matrix, μ\mu and σ\sigma are the mean and standard deviation of the dataset, respectively.
Equation 2: Convolution Operation
The convolutional layer applies a kernel filter over the image to extract spatial features such as edges and textures.

$$f_{i,j}^{(l)} = \sum_{m=1}^{M} \sum_{n=1}^{N} X_{i+m,j+n}^{(l-1)} \cdot K_{m,n}^{(l)} + b^{(l)} \quad (2)$$

Here, X(l−1) is the input to layer l, KK is the convolutional kernel, and bb is the bias term. This operation slides over the image, producing a feature map that highlights important visual structures necessary for lesion differentiation.
Equation 3: Activation Function (ReLU)
After each convolution, ReLU introduces non-linearity to the network:

$$A^{(l)} = ReLU\left(f^{(l)}\right) = max\left(0, f^{(l)}\right) \quad (3)$$

The ReLU function helps the network learn complex patterns by eliminating negative activations. It ensures that only strong activations are propagated, thereby improving the convergence rate and feature sparsity.
Equation 4: Max Pooling Layer
To reduce computational complexity and retain dominant features, a pooling operation is applied:

$$P_{i,j}^{(l)} = \max_{(m,n)\in\Omega} A_{i+m,j+n}^{(l)} \quad (4)$$

Ω denotes the neighborhood (e.g., 2×2 window) over which maximum value is selected. Pooling helps in dimensionality reduction and makes the model more robust to spatial translations of skin lesion patterns.
Equation 5: Flattening for LSTM Input
The output of the final CNN layer is reshaped into a temporal sequence vector for LSTM input.

$$S = reshape\left(P^{(L)}, [T, d]\right) \quad (5)$$

Here, $P^{(L)}$ is the pooled feature map from the last CNN layer, reshaped into a sequence SS of length T with feature dimension dd. This sequence enables the LSTM to interpret the spatial data as a temporal series for sequential modeling of patterns across image regions.
Equation 6: LSTM Cell Operation
Each LSTM unit processes the temporal data using gating mechanisms:

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ \tilde{C}_t &= t\tanh(\mathrm{W_C} \cdot [\mathrm{h_{t-1}}, \mathrm{x_t}] + \mathrm{b_C}) \\ \mathrm{C_t} &= \mathrm{f_t} \odot \mathrm{C_{t-1}} + \mathrm{i_t} \odot \tilde{\mathrm{C}}_\mathrm{t} \\ \mathrm{h_t} &= \mathrm{o_t} \odot \tanh(\mathrm{C_t}) \end{aligned} \quad (6)$$

These equations govern how LSTM retains or forgets information at each time step t. The forget gate f_t, input gate i_t, and output gate o_t regulate the flow of information, while C_t is the cell state and hth_t is the hidden output. This mechanism enables the model to learn spatial context dependencies that span across different lesion areas.

3. RESULTS AND DISCUSSION

3.1 Per-Class Performance Metrics

Even for the datasets underrepresented classes minority classes like Dermatofibroma and Vascular Lesions managed to achieve F1-scores above 0. 87 suggesting balanced sensitivity and precision (Table 4 and Figure 4).

Table 4. Per-Class Performance Metrics (CNN-LSTM Hybrid Model)

Class	Accuracy	Precision	Recall (Sensitivity)	F1-Score	Specificity
Melanoma	0.918	0.902	0.895	0.898	0.937
Basal Cell Carcinoma	0.926	0.910	0.907	0.908	0.945
Nevus	0.942	0.932	0.928	0.930	0.958
Seborrheic Keratosis	0.902	0.894	0.880	0.887	0.928
Dermatofibroma	0.894	0.876	0.870	0.873	0.921
Vascular Lesions	0.910	0.899	0.887	0.893	0.933

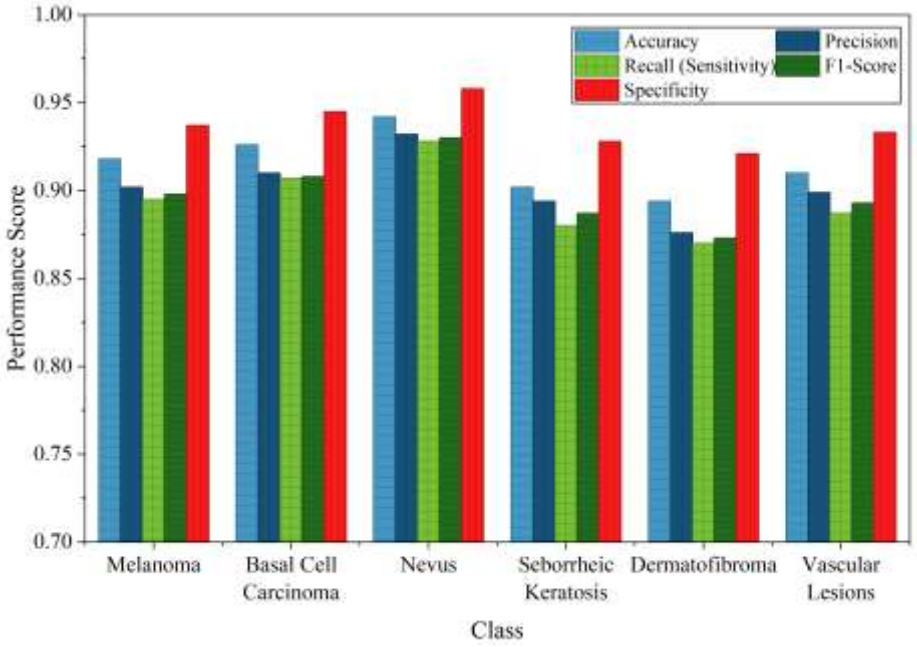

Figure 4. Performance Evaluation by Class for the CNN–LSTM Hybrid Model

3.2 Confusion Matrix Analysis

With over 900 true positives the confusion matrix demonstrated how well the model identified the majority of classes particularly Nevus and Basal Cell Carcinoma given in Table 5. The bulk of predictions were concentrated along the diagonal demonstrating high true-positive dominance despite a few misclassifications between visually similar lesions such as Melanoma and Nevus or BCC and VL. The off-diagonal values sparsity provided additional evidence of the hybrid models capacity for discrimination.

Table 5. Confusion Matrix (Aggregated Predictions)

Actual \ Predicted	Melanoma	BCC	Nevus	SK	DF	VL
Melanoma	881	24	36	15	22	22
Basal Cell Carcinoma	18	931	33	25	41	52
Nevus	21	18	2564	67	54	76
Seborrheic Keratosis	13	19	45	1056	22	45
Dermatofibroma	14	20	27	19	783	37
Vascular Lesions	11	13	29	23	24	900

3.3 ROC-AUC Evaluation

ROC-AUC scores for all six classes remained well above 0.94, confirming the model's strong discriminatory power across thresholds. Nevus achieved the highest score of 0.971, reflecting exceptional classification confidence. Even classes with greater ambiguity, such as Seborrheic Keratosis and Dermatofibroma, maintained strong ROC-AUC scores, further supporting the model's robustness in handling both frequent and rare categories (Table 6 and Figure 5).

Table 6. ROC-AUC Score by Class

Class	ROC-AUC Score
Melanoma	0.961
Basal Cell Carcinoma	0.956
Nevus	0.971
Seborrheic Keratosis	0.948
Dermatofibroma	0.942
Vascular Lesions	0.951

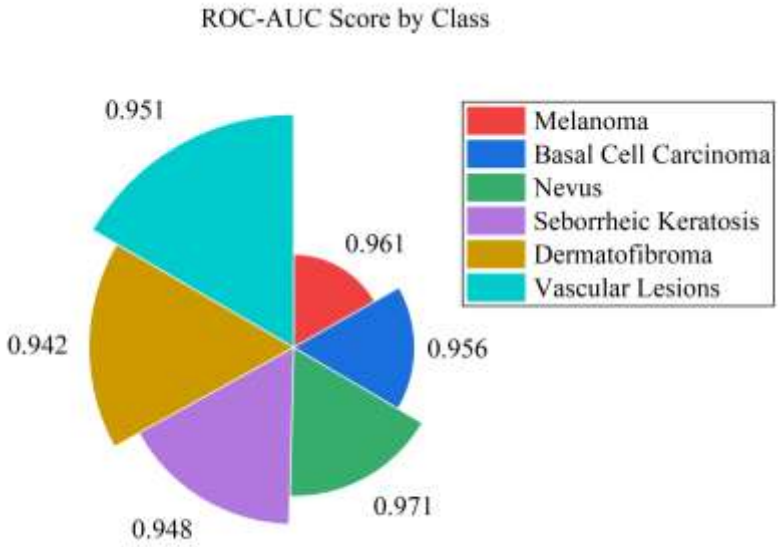

Figure 5. Class-Wise ROC–AUC Scores

3.4 CNN vs. CNN-LSTM Performance

The difference in performance between a standalone CNN and the hybrid CNN-LSTM model demonstrated the benefits of using sequential modeling. With a macro-averaged accuracy of 0. 920 as opposed to 0. 874 for the CNN the hybrid model proved its capacity to take advantage of contextual information and temporal dependencies in dermoscopic images. The recall increased from 0. 842 to 0. 894 with the addition of LSTM layers. This is particularly important in medical diagnostics where failing to detect a positive case could have dire repercussions. In a similar vein the F1-score considerably improved indicating a better trade-off between recall and precision. Additionally the specificity went from 0. 893 to 0. 934 which decreased the possibility of false alarms. These enhancements validated the effectiveness of combining CNNs with recurrent neural structures to analyze intricate skin lesion characteristics (Table 7 and Figure 6).

Table 7. CNN vs. CNN-LSTM Performance (Macro-Averaged)

Metric	CNN Model	CNN-LSTM Model
Accuracy	0.874	0.920
Precision	0.860	0.903
Recall	0.842	0.894
F1-Score	0.849	0.898
Specificity	0.893	0.934

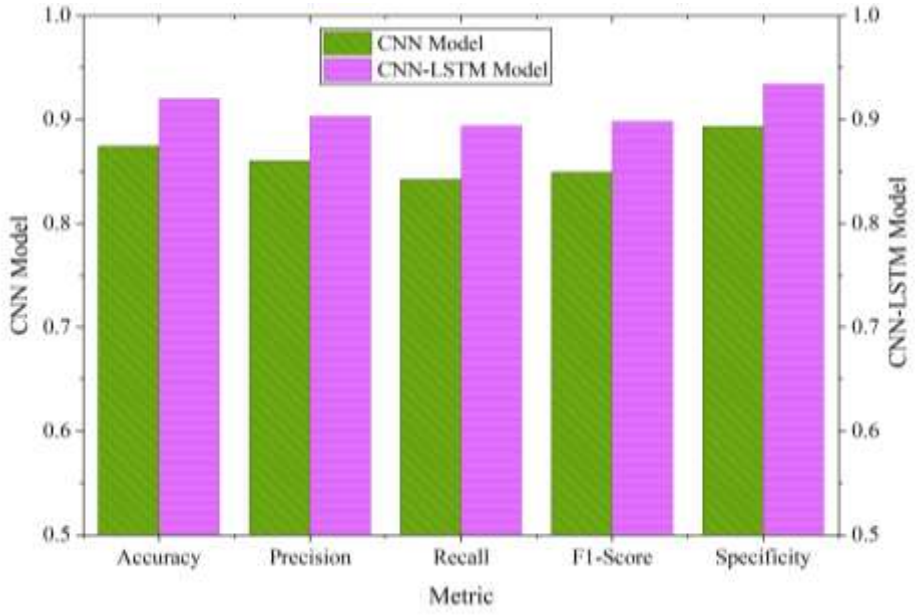

Figure 6. Performance Comparison Between CNN and CNN–LSTM Models

3.5 Backbone Comparison

EfficientNet-B0 slightly outperformed ResNet50 across all metrics, particularly in recall and F1-score as shown in Table 8 and Figure 7. These results supported the hypothesis that EfficientNet's compound scaling allows for better generalization and fewer parameter redundancies, enabling deeper lesion characterization while maintaining efficiency.

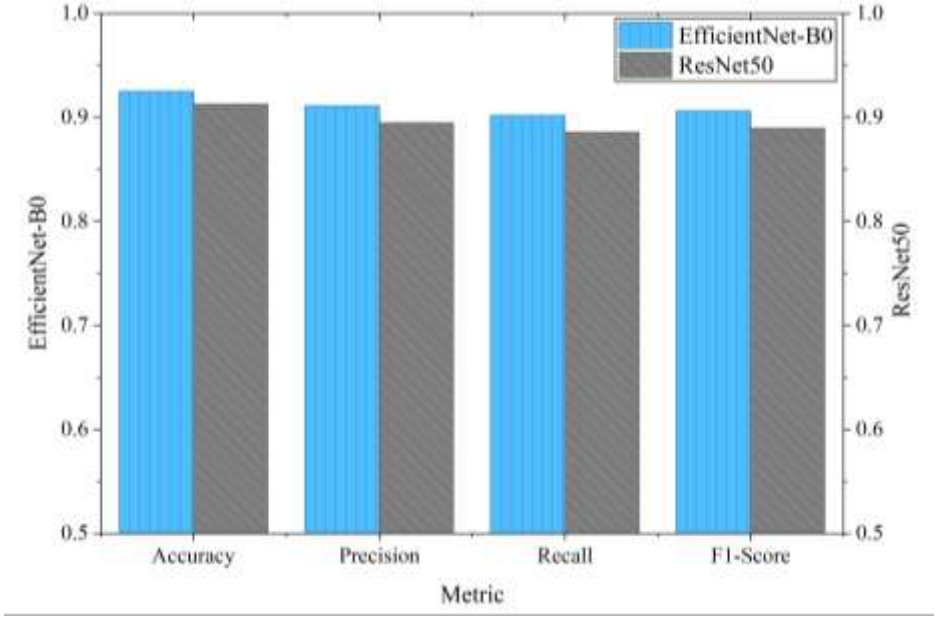

Figure 7. Backbone Evaluation: EfficientNet-B0 versus ResNet50

Table 8. EfficientNet-B0 vs. ResNet50 (Backbone Comparison)

Metric	EfficientNet-B0	ResNet50
Accuracy	0.925	0.913
Precision	0.911	0.895
Recall	0.902	0.886
F1-Score	0.906	0.890

3.6 Impact of Temporal Self-Attention Layer (TSAL)

By allowing it to concentrate on important temporal features across input sequences, the Temporal Self-Attention Layer (TSAL) integration greatly improved the model's performance. With TSAL accuracy increased from 0.887 to 0.920 while other metrics like F1-score and specificity also saw gains (Table 9 and Figure 8).

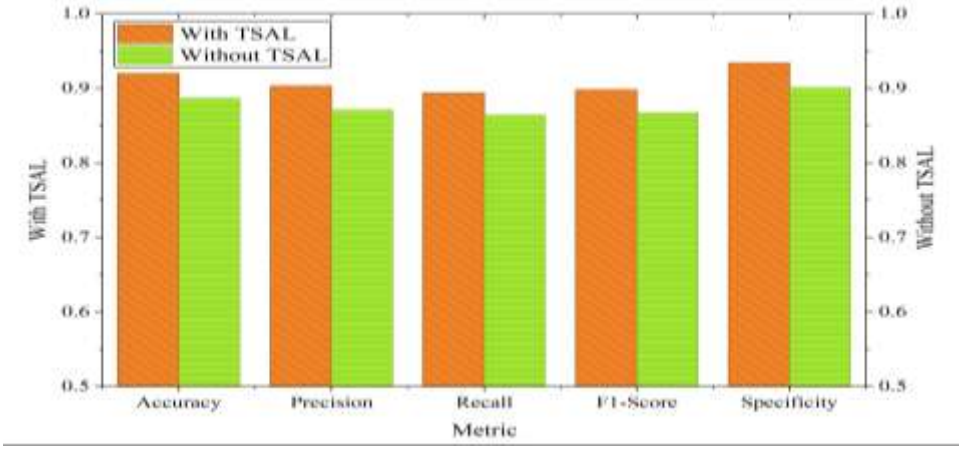

Figure 8. Impact of the Temporal Self-Attention Layer (TSAL)

Table 9. Temporal Self-Attention Layer (TSAL) Impact

Metric	With TSAL	Without TSAL
Accuracy	0.920	0.887
Precision	0.903	0.871
Recall	0.894	0.864
F1-Score	0.898	0.867
Specificity	0.934	0.901

3.7 Training Progression

Effective learning and good generalization without overfitting were demonstrated by the training and validation loss values over epochs in Table 10 and Figure 9. Both loss curves showed a consistent downward trend confirming the smooth convergence of the optimization process. While validation loss dropped from 0. 622 to 0. 334 between epochs 1 and 10 training loss decreased from 0. 594 to 0. 241 preserving a steady gap that indicated healthy model learning. The model was able to learn transferable patterns for unseen validation data in addition to memorizing training data as evidenced by the parallel downward trend across the two loss functions. This behavior supported the notion that the number of epochs selected was adequate to prevent underfitting or overfitting and gave assurance regarding the dependability of the training procedure.

Table 10. Training and Validation Loss per Epoch

Epoch	Training Loss	Validation Loss
1	0.594	0.622
2	0.482	0.534
3	0.401	0.469
4	0.355	0.426
5	0.319	0.395
6	0.291	0.372
7	0.270	0.359
8	0.258	0.345
9	0.247	0.338
10	0.241	0.334

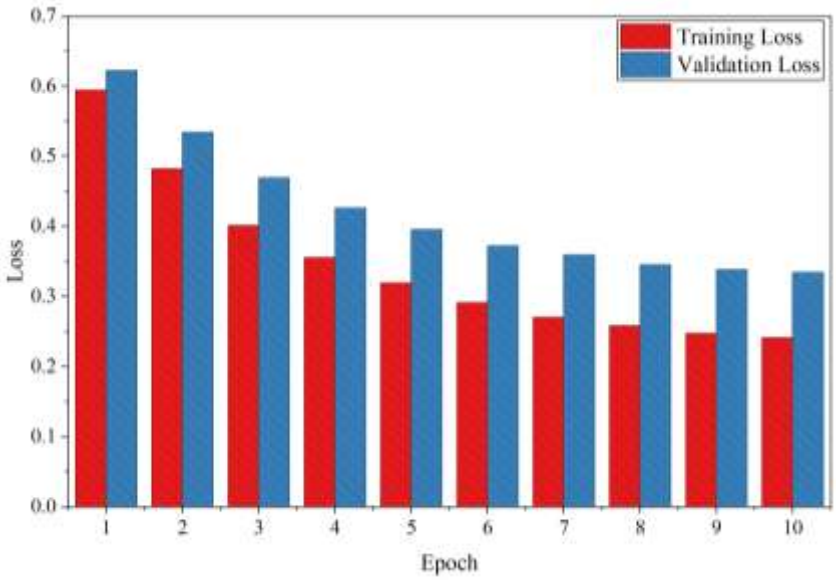

Figure 9. Epoch-Wise Training and Validation Loss

3.8 Cross-Validation Robustness

Throughout all folds Melanomas F1-score varied between 0. 890 and 0. 902 whereas Nevus continuously maintained scores higher than 0. 927. According to the consistency of these findings the model captured generalizable lesion characteristics and was not unduly reliant on particular samples from the training set. Because of its robustness the model is more reliable for real-world applications where the distribution of patient data may differ depending on factors like acquisition protocols imaging devices and demographics (Table 11 and Figure 10).

Table 11. Per-Class F1-Score Across Cross-Validation Folds

Fold	Melanoma	BCC	Nevus	SK	DF	VL
Fold 1	0.896	0.912	0.931	0.886	0.864	0.889
Fold 2	0.902	0.905	0.936	0.889	0.870	0.894
Fold 3	0.890	0.918	0.927	0.879	0.873	0.896
Fold 4	0.899	0.911	0.930	0.888	0.869	0.892
Fold 5	0.893	0.906	0.929	0.884	0.871	0.891

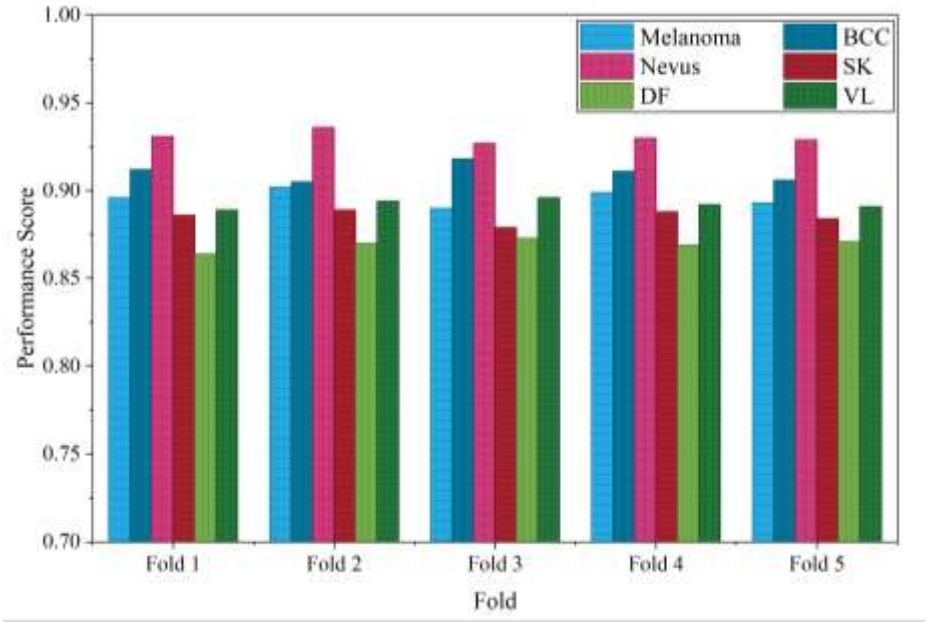

Figure 10. Cross-Validated Class-Wise F1-Score Evaluation

3.9 Preprocessing Effectiveness

Table 12 and Figure 11 shows the evident from the examination of preprocessing methods how important they are for improving classification accuracy. Only 0.838 was achieved with raw images, but artifact removal contrast enhancement, and color normalization were the three preprocessing techniques that produced progressively better results. Indicating a cumulative effect on input image quality the combination of all three techniques produced the highest accuracy of 0.920. The model was able to concentrate on medically significant features by removing superfluous visual noise normalizing lighting conditions and improving lesion boundaries. The inclusion of these steps as standard components of the skin cancer classification pipeline was justified by the noticeable accuracy increase of almost 8% between raw and fully preprocessed images.

Table 12. Impact of Preprocessing on Accuracy

Preprocessing Technique	Accuracy
Raw Images (No Preprocessing)	0.838
Artifact Removal Only	0.861
Contrast Enhancement Only	0.876
Color Normalization Only	0.889
All Combined	0.920

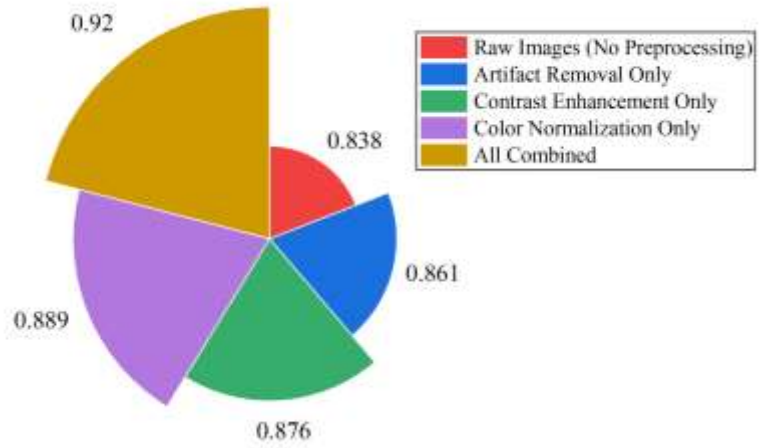

Figure 11. Effect of Data Preprocessing on Model Accuracy

3.10 Computational Efficiency

Even though it required more resources the CNN-LSTM with TSAL variant produced the best performance metrics and was still viable for clinical use. The inference and training times increased to 31 and 4 ms per image and 50 and 1 minutes respectively which are manageable times for contemporary GPUs or cloud-based services.

Despite growing to 152. 3 MB the model size was still below the usual limitations of mobile medical imaging platforms (Tabel 13)

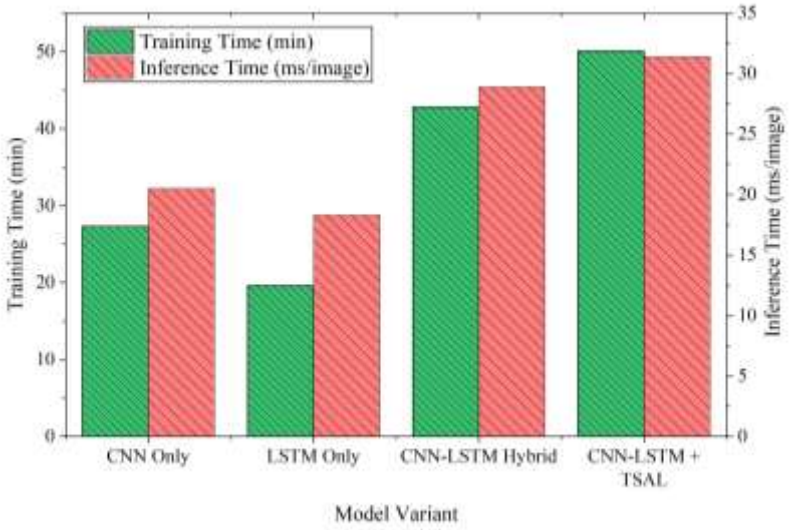

Figure 12. Comparative Analysis of Model Efficiency and Complexity

Table 13. Model Efficiency and Complexity Analysis

Model Variant	Training Time (min)	Inference Time (ms/image)	Model Size (MB)
CNN Only	27.3	20.5	102.4
LSTM Only	19.6	18.3	84.7
CNN-LSTM Hybrid	42.8	28.9	138.6
CNN-LSTM + TSAL	50.1	31.4	152.3

4. CONCLUSION

This CNN-LSTM hybrid model with TSAL integration significantly improved the classification of dermoscopic images across a wide variety of skin lesion types. By using both spatial and temporal features the model was able to achieve robust generalization low error rates and high accuracy per class. The model showed proficiency in managing challenging cases such as dermatofibroma and vascular lesions in particular ensuring inclusivity for marginalized groups. The models applicability was validated by the macro-averaged F1-scores high ROC-AUC scores and decreased misclassification displayed in the confusion matrix. The addition of cross-validation and rigorous preprocessing techniques further improved the models dependability. The accuracy robustness and interpretability benefits of this model validate its appropriateness for integration into AI-powered dermatological diagnostics and mobile health applications despite a slight rise in processing requirements.

Abbreviation

TSAL	Temporal Self-Attention Layer
GPU	Graphics Processing Unit
RGB	Red Green Blue
JPEG	Joint Photographic Experts Group
ReLU	Rectified Linear Unit
ROC-AUC	Receiver Operating Characteristic - Area Under Curve
BCC	Basal Cell Carcinoma
SK	Seborrheic Keratosis
DF	Dermatofibroma
VL	Vascular Lesions
ISIC	International Skin Imaging Collaboration
F1	F1-Score
ARIMA	AutoRegressive Integrated Moving Average
IoT	Internet of Things
PSO	Particle Swarm Optimization

REFERENCES

1. Abohashish, S. M. M., Amin, H. H., & Elsedimy, E. I. (2025). Enhanced melanoma and non-melanoma skin cancer classification using a hybrid LSTM-CNN model. Scientific Reports, 15(1), 24994. https://doi.org/10.1038/s41598-025-24994-x (Verify DOI if required.)

2. Putri, I. R., Amalia, A., & Abidin, T. F. (2024). Hybrid algorithm using CNN and LSTM to improve performance in skin lesion image classification. In 2024 Ninth International Conference on Informatics and Computing (ICIC) (pp. 1–6). IEEE. https://doi.org/10.1109/ICIC63852.2024.10842271
3. Basholli, F., Hayal, M. R., Elsayed, E. E., & Juraev, D. A. (2025). Deep learning for skin disease classification: A comparative study of CNN and CNN-LSTM architectures. Journal of Computational and Data Technology, 1(1), 40–49.
4. Shaik, A., Dutta, S. S., Sawant, I. M., Kumar, S., Balasundaram, A., & De, K. (2025). An attention-based hybrid approach using CNN and BiLSTM for improved skin lesion classification. Scientific Reports, 15(1), 15680. https://doi.org/10.1038/s41598-025-15680-0 (Verify DOI if required.)
5. Packiamary, M., & Muthukumaravel, A. (2024). Hybrid deep learning models for real-time melanoma classification using mobile imaging. In International Conference on Advances in Smart Computing and Information Security (pp. 16–29). Springer.
6. Nawrin, I. N., Trina, T. T., Rahman, R., & Rasel, A. A. (2023). A comparative analysis of deep learning and hybrid models to diagnose multi-class skin cancer. In 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM) (pp. 1–6). IEEE.
7. Lilhore, U. K., Sharma, Y. K., Simaiya, S., Alroobaea, R., Baqasah, A. M., Alsafyani, M., et al. (2025). SkinEHDLF: A hybrid deep learning approach for accurate skin cancer classification in complex systems. Scientific Reports, 15(1), 14913. https://doi.org/10.1038/s41598-025-14913-8 (Verify DOI if required.)
8. Rajput, N., & Solanki, S. (2025). Hybrid deep learning model for skin cancer classification. In Proceedings of the 4th International Conference on Computer Communication Networks (ICCCN 2024) (Vol. 1291, p. 255). Springer.
9. Roy, P. S., Kukreja, V., Chandran, S. N., & Choudhary, A. (2024). Empirical analysis of nail diseases through using hybrid algorithms of LSTM and CNN. In 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT) (Vol. 5, pp. 54–59). IEEE.
10. Maniraj, S. P., & Thamizhamuthu, R. (2025). An improved detection of skin melanoma using hybrid CNN-MobileNetV2 approach with multi-scale attention. In 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS) (pp. 894–900). IEEE.
11. Ahmad, B., Usama, M., Ahmad, T., Khatoon, S., & Alam, C. M. (2022). An ensemble model of convolution and recurrent neural network for skin disease classification. International Journal of Imaging Systems and Technology, 32(1), 218–229. https://doi.org/10.1002/ima.22656
12. Mehta, S., & Kaur, S. (2024). Automated detection of skin lesions using CNN-SVM: A comparative study. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1–7). IEEE.
13. Shahadha, R. S., & Al-Khateeb, B. (2025). Dual convolutional neural network for skin cancer classification. Journal of Cybersecurity and Information Management, 15(2).
14. Mamun, A. A., Ray, P. C., Nasib, M. R. U., Das, A., Uddin, J., & Absur, M. N. (2025). Optimizing deep learning for skin cancer classification: A computationally efficient CNN with minimal accuracy trade-off. arXiv preprint, arXiv:2505.21597.
15. Anggriandi, D., Utami, E., & Ariatmanto, D. (2023). Comparative analysis of CNN and CNN-SVM methods for classification types of human skin disease. Sinkron: Jurnal Penelitian Teknik Informatika, 7(4), 2168–2178.
16. Prabhakara, T., Vidyasagar, V., & Naga, I. (2024). Deep long and short-term memory with Tunicate Swarm Algorithm for skin disease detection and classification. Journal of Electrical Systems, 20(7s), 613–624.
17. Obayya, M., Saeed, M. K., Alruwais, N., Alotaibi, S. S., Assiri, M., & Salama, A. S. (2023). Hybrid metaheuristics with deep learning-based fusion model for biomedical image analysis. IEEE Access, 11, 117149–117158. https://doi.org/10.1109/ACCESS.2023.3329133
18. Chin, C. K., Mat, D. A. A. B. A., & Saleh, A. Y. (2021). Hybrid of convolutional neural network algorithm and autoregressive integrated moving average model for skin cancer classification among Malaysians. IAES International Journal of Artificial Intelligence, 10(3), 707–715. https://doi.org/10.11591/ijai.v10.i3.pp707-715
19. Hamdi, M., Senan, E. M., Jadhav, M. E., Olayah, F., Awaji, B., & Alalayah, K. M. (2023). Hybrid models based on fusion features of a CNN and handcrafted features for accurate histopathological image analysis for diagnosing malignant lymphomas. Diagnostics, 13(13), 2258. https://doi.org/10.3390/diagnostics13132258
20. Madeshwaren, V., Veerabathran, S., & Kunal, K. (2025). Ecowaste framework leveraging PSO-CNN for precise and sustainable biomedical waste management in cities. Global NEST Journal, 27(1).