

# FUSIONSEGNET: A Deep Learning Framework For Accurate And Explainable Skin Disease Classification Using Multi-Model Integration

Abhipsa Pattanaik<sup>1\*</sup>, Leena Das<sup>2</sup>

<sup>1\*</sup>School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India  
E-mail: [abhipsaapattanaik@gmail.com](mailto:abhipsaapattanaik@gmail.com), Orcid Id: 0009-0009-0193-6752

<sup>2</sup>School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India  
Email: [ldasfcs@kiit.ac.in](mailto:ldasfcs@kiit.ac.in), Orcid ID: 0000-0002-8719-1991

## Abstract:

**Context:** Skin cancer is one of the most dangerous diseases in the world. Diagnosing it with a traditional model struggles to give accurate results and suffers from computational efficiency and clinical interpretability. But handling such diseases through deep learning gives promising results nowadays.

**Objective:** In this paper, we aim to classify skin disease using deep learning models. We propose a deep learning model called FusionSegNet that combines the benefits of Transformers and CNNs to provide clinically reliable and highly accurate skin disease classification.

**Material/Methods.** We used several deep learning models to integrate to detect the skin disease. Our method combines multi-scale feature extraction from the ResNet50 and Vision Transformer (ViT) branches with attention-guided fusion to identify diagnostically significant patterns. On the ISIC 2018 dataset, FusionSegNet outperforms cutting-edge techniques like ConvNeXt (92.3%) and DeepSkinNet (92.1%) with an accuracy of 96.3%.

**Result:** We used nine deep learning models, such as MobileNetV3, EfficientNet-B4, ResNet50, DenseNet121, ViT-B/16, DeepLabV3+, Swin-Tiny, DeepSkinNet (SOTA), ConvNeXt-Small, and our proposed model, FusionSegNet. Our proposed hybrid deep learning model obtained the highest accuracy of 96.3%, precision of 95%, recall of 96%, F1-score of 95% AUC of 98% as comparisons to the other models. We also used the ablation and statistical significance to confirm that each architectural component shows that the model's performance is increasing. FusionSegNet is a suitable solution for reliable, efficient, and generalizable skin disease diagnosis, with a high potential for integration into clinical decision-support systems.

**Keywords:** Skin cancer classification; deep learning, multi-model fusion; explainable AI; medical image analysis

**How to cite this article:** Pattanaik A, Das L. FUSIONSEGNET: A Deep Learning Framework For Accurate And Explainable Skin Disease Classification Using Multi-Model Integration. *Int J Drug Deliv Technol.* 2026;16(16s): 333-354; DOI: 10.25258/ijddt.16.16s.36

## 1. Introduction:

Nowadays skin disease is a common disease in the world. Almost all people in the world are affected by skin disease in different age groups. Among skin diseases, Skin cancer is one of the diseases where the maximum people suffer. It is one of the dangerous diseases nowadays. Early care and identification is one of the challenging tasks. Due to environmental conditions and changes, melanoma and other skin cancers are rapidly increase day by day [As per the World Health Organization]. To keep people from dying, make treatment work better, and lower the cost of healthcare, skin tumours must be found early and correctly classified. Skin cancer can be visualized with the help of an examination by dermatologists. Different types of procedures need to be followed to confirm the skin cancer, like biopsy, histopathological analysis, dermoscopy, etc. These medical procedures take much time to confirm that the patient is suffering the skin cancer. These procedures are taking up much time in rural areas. With this thought process, we designed an automated and intelligent model that can classify early detection as well as classification of skin cancer diseases.

The technological advancement helps in the medical sector to detect cancer early. Especially, machine

learning and deep learning play a vital role in detecting skin cancer. The model CNN plays a dramatic role in identifying diseases and classifying images. It has functionality like detection, segmentation, and image classification. For medical images, CNN plays a vital role and extracts useful patterns from the large volumes of data. It analyzes and tells the patterns whether it is benign or malignant. We fed the large data into the model, where the images are in the form of annotated skin images. The model learning the pattern of skin conditions like basal cell carcinoma, seborrheic keratosis, actinic keratosis, melanoma, etc. Skin cancer is one of the most common and deadly diseases in the world, and early detection is important for increasing patient survival rates. Traditional clinical diagnosis via visual inspection and biopsy is time-consuming, subjective, and primarily reliant on experienced dermatologists. In recent years, deep learning-based computer-aided diagnostic (DCAD) systems have demonstrated substantial promise in automating skin disease classification using dermoscopic pictures. However, many existing deep learning models have disadvantages such as insufficient feature representation, poor generalization across varied lesion patterns, significant computational complexity, and a lack of clinical interpretability.

\*Author for Correspondence: [abhipsaapattanaik@gmail.com](mailto:abhipsaapattanaik@gmail.com)

Most conventional approaches rely on a single deep learning architecture, specifically convolutional neural networks (CNNs), which are primarily concerned with local spatial data. Although CNN-based models like ResNet, DenseNet, and MobileNet have generated promising results, they often struggle to capture global contextual relationships and multi-scale lesion features. Transformer-based models, such as Vision Transformers (ViT), are useful for modelling large dataset. As a result, current models fail to give a balanced representation of both local and global lesion data, resulting in inferior classification performance in complicated skin disease scenarios. FusionSegNet's performance is compared with a number of cutting-edge deep learning models, such as MobileNetV3, EfficientNet-B4, ResNet50, DenseNet121, ViT-B/16, Swin-Tiny, DeepLabV3+, ConvNeXt-Small, and DeepSkinNet, using the ISIC 2018 benchmark dataset.

### Problem Statement

There are several authors have proposed skin cancer diseases, but there is a scope for improvement.

1. Class imbalance issues
2. Importance for model building but not model interpretation
3. Some of the datasets might contain a larger number of benign cases than malignant cases

To overcome the above issues, not only CNN used, but also much focus was placed on transformer-based models, segmentation-based classification, attention-based, etc.

### Motivation

The motivation of this study is to develop an end-to-end deep learning system that helps us detect skin cancer. Our model estimates the performance metrics of all the deep learning algorithms.

### Research Question:

This study discusses the following challenging task for skin cancer disease .We present the solution of these research questions in the subsequent sections.

**RQ1:** Does the dual-task learning approach of FusionSegNet offer significant performance gains in classification and segmentation compared to single-task CNN-based models?

**RQ2:** How does FusionSegNet's hybrid architecture achieve superior accuracy (96.3%) compared to existing SOTA models on the ISIC 2018 dataset

**RQ3:** Does FusionSegNet maintain performance consistency across all evaluation metrics (Accuracy, F1-score, AUC)?

**RQ4 :** How do FusionSegNet's results translate to real-world clinical applicability?

### Research Objectives

1. Our objective is to detect skin cancer using deep learning algorithms
2. Evaluate and compare CNN-based and

Transformer-based models using conventional performance criteria, including accuracy, precision, recall, F1-Score, and AUC-ROC.

3. To suggest a unique architecture (FusionSegNet) that combines classification and segmentation and exhibits enhanced explainability and accuracy.

This study organizes the 6 different sections. Section 1 presents the introduction; Section 2 discusses the related work, and compares the proposed model with an existing approach for skin disease detection. Section 3 discusses the proposed methodology we emphasizing the data pre- processing activities. Section 4 covers the performance metrics of different deep learning classifiers

### 2. Literature Review

Badr, M., et al. [1] presented work related to a multimodal deep learning approach for detecting skin disease. Their focus was on identifying skin diseases in their early stages using DL techniques. The author trained 25010 images and achieved an accuracy of 95% and an AUC- ROC score of 99.4%.

Abdulredah, A. A., et al. [2] used the MNIST HAM10000, ISIC2019, and ISIC2020 datasets for classifying the benign and malignant classes. The author was considered the XAI to interpret the deep learning model's decision. The author found an accuracy of 98.9% and the F1-score is 97.5%. They used the DL models like Darknet, MobileNet, and EfficientNet.

Natarajan et al. [3] proposed a model that allows to detect the skin cancer as well as how to enhance the performance of the model. The deep learning model was used for identifying skin cancer, like VGG19, VGG16, InceptionV3, Xception, where the input image is the input to the model and the resized pixels of 224x224 pixels.

Gairola, K. et al. [4], their challenge was to pre-process the image data, as a lot of noise was there in their dataset. Then the author employed shallow and deep fusion, late fusion techniques to increase the feature extraction as well as to enhance the performance of the model. The DL model, like CNN, was used for identifying skin diseases.

Khan, A. et al. [5] developed a multimodal model that classifies skin cancer. The author utilised the pre-trained models on HAM\_10000 Metadata. The author considered 100 high-resolution images and obtained 92% accuracy. Gulzar, Y. et al.,[6] developed a model that allows for to computation of the performance metrics of several models. The author used deep learning models like VGG, AlexNet, DenseNet, EfficientNetB, MobileNetV, and achieved an accuracy of 97.5%, a precision of 0.95. It was observed that the model DenseNet, EfficientNet B, performed well in classifying skin cancer cells.

Vayadande, K et al. [7] was proposed for skin disease classification using machine learning and deep learning approaches. They conducted a systematic literature

review on it and provided insights into skin diseases, like: - how to classify using SVM, KNN, CNN, etc. Verma,

K. et al. [8] developed a model that helps to classify the skin lesion as well as diagnosis. It was the author's multimodal approach that utilizes the DL models like CNN, DenseNet201, and ResNet152. The author obtained an accuracy of 93.18% in the swim transformer model as comparison to the other models. Juan, C. K., et al. [9] developed a model for the skin disease classification task. The author considered the deep learning model and used 1215 clinical images of skin tumours. This research was conducted from 2015 to 2020 at Taichung and Taipei Veterans General Hospital. During the implementation, the author achieved that skinFLNet outperformed and achieved an accuracy of 85%, precision of 85% as compared to other models. Jaiyeoba, O. et al. [10] designed a model that allows to classify the skin diseases using ensemble learning techniques. The author utilized the traditional machine learning algorithms like NB, RF, SVM, DT, GDB, etc. It has been observed that the stacking classifier performed well in comparison to the other models. The accuracy is 99.30, and precision is 1.00, recall of 0.96. But they did not utilize any deep learning models to classify the skin cancer. Halder, A. et al. [11] explored a deep learning model that classifies the multi-class skin diseases. The author considered the HAM10000 dataset, which contains high-resolution images of skin lesions. As the dataset was imbalanced then the author applied the data balancing technique then the dataset was divided into training and testing purposes. The DL model was used, like: Xception, InceptionResNetV2, and MobileNetV2, to classify skin cancer.

Kalaivani, A. et al. [12] analyzed the dataset of skin diseases. The author utilized the dataset ISIC2019, which is available of publicly available. Data mining and ensemble learning are used to detect skin cancer. The author categorized 7 types of classes to classify skin cancer. It has been observed that ensemble learning performance is good compared to the data mining approaches. Sardar, M. et al. [13] developed an ensemble learning model that allows detecting skin cancer diseases. The DL model CNN is utilized, especially the multi-modal approach like the ResNet (M-ResNet) architecture. The ResNet performs well in detecting skin cancer.

Cheng, H., Lian, J., & Jiao, W.[14] Optimized the model that classifies skin cancer classification. The authors' objective was to make an accurate and early classification so that the survival chance would be higher. The author compared their proposed approach to the state-of-the-art and concluded that the proposed approach is good

Shetty, B. et al. [15] investigated the skin classification model. The author used the DL models and employed the HAM10000 dataset. It consists of 10015 images. The CNn model outperforms with an accuracy of 95.18%. Sun, J. et al. [16] did a rigorous study on skin disease classification using machine learning models.

The author explored the different publicly available datasets and compared them. Gairola, A. K., et al. [17] developed a multimodal fusion deep learning system for skin diseases. The author compared their approach to the state-of-the- art. The highest accuracy obtained was 86% for ISB and 90% for FFN using HAM10000. Nanda, P. et al., [18] formulated a method that allows for to detection of the multi-class skin cancer detection mechanism using the deep neural network model. The author used the HAM10000dataset for skin cancer classification. As the dataset is imbalanced, they used a data balancing technique to improve the model's performance. El Gannour, O. et al.,[19] developed a model for skin classification. The author used the dual ensemble learning algorithms to see how the performance is increased. The ISCI 2018 dataset was considered and used with DL models. The highest accuracy was obtained at 94.90% through the CNN model. Zhang, T. et al.,[20] developed a model for classification model that passes the dermatological images. Their dataset has 9013 images and was trained with 1002 images, where seven classes were considered for skin classification. The DenseNet121-Multi model performed well with an F1-score of 0.91. Vieira, J. et al.,[21] present the deep learning models for Skin Lesion Detection. The author discusses how the DL model helps to detect skin cancer. Atiq, M. E., & Fattah, S. A.[23] presented the Skin Cancer Classification task using deep learning techniques through the clinical task. The author utilised the HAM10000, ISIC 2018. When the author was compared to baseline models, the suggested approach greatly increased classification accuracy and average AUC while achieving state-of-the-art segmentation performance. When compared to baseline models, the suggested approach greatly increases classification accuracy and average AUC while achieving state-of-the-art segmentation performance. From their ablation study, it is confirmed that Dual- Network (With Metadata) obtained 93.47% for HAM10000 and HAM10000 % for ISIC 2019 (%). Shah, S. A. H., et al. [24] explored the deep learning models for skin cancer detection. Apart from this, the author also utilised an explainable AI concept with CNN classification techniques. Particle Swarm Optimisation was used to optimise feature dimensions, decreasing them from 1024 to 508, considerably increasing computing efficiency. Machine-learning classifiers, such as Subspace KNN and Medium Gaussian SVM, increased classification accuracy. The suggested pipeline obtained 98.5% and 86.1% accuracy on the ISIC 2018 and HAM10000 datasets Pavani, M. et al. [25] presented that skin disease detection can be possible using the hybrid explainable DL models. The author's main contribution was to use the Sakaguchi Function with a CNN classification model to extract the features. The author used the CNN models such as ResNet50 and InceptionV3 along with ViT. It has been observed that the proposed model obtained an accuracy of 96.10%, precision of 96.40%, and recall of 95.80%.

Ilyosbekov, S.[26] The author demonstrated the method of deep learning, such as an EfficientNetV2 framework with GradCAM++ attention classification for Skin Lesion. The author used the ISIC 2019 dataset, which comprises 25,331 dermoscopic images in nine diagnostic categories, to assess our system. Their proposed model provides clinically significant explanations that match model attention with established dermatological assessment criteria, and it achieves 85.61% accuracy with a weighted F1 score of 0.8564. Zoravar, M. et al.,[27] the authors proposed an ensemble of vision transformer models that were trained on a variety of datasets, such as the Skin Cancer ISIC, Dermofit, and HAM10000 datasets. The experimental observation revealed that the vision transformer model performed well.

Manzoor, K. et al.,[28] the authors presented a deep learning model to detect skin cancer diseases. The DL models, such as EfficientFormerV2, produced 97.11% accuracy, a 97.14% F1-score, 96.85% sensitivity, and 96.70% specificity using the HAM10000 dataset. The segmentation model obtained 97.59% accuracy, 89.12% Jaccard index, and 94.24% dice similarity coefficient on the ISIC 2018 dataset. Akram, A. et al.,[29] presented the concept of segmentation and classification techniques for skin cancer. The author has used the hybrid DL models and obtained an accuracy of 95.49%. When tested on the ISIC 2020 Challenge dataset, the model achieves an accuracy of 96.75% percent. Al Shafi, A. et al., [30] exhibited the skin lesion classification task using the CNN model. In this study, the author used the three benchmark datasets: HAM10000, ISIC 2016, and ISIC 2019. The procedure included rebalancing, picture augmentation, and filtering techniques, followed by a hybrid dual encoder for segmentation using transfer learning. For the three datasets, the method's lesion recognition accuracy was 96.32%, 90.86%, and 93.92%. Liu, R. et al. [31] authors discussed how deep learning techniques help in skin lesion classification. The ISIC 2020 dataset consists of 3297 photos of benign and malignant skin lesions, is used to assess the approach. When compared to five traditional convolutional neural network (CNN) models, experimental results demonstrate that the suggested ASFF-based ResNet-50 achieves the best overall performance. Along with improved precision, recall, specificity, and F1 score, the suggested model achieved an accuracy of 93.18%. In the P-R and ROC curves, the enhanced model obtains AUC values of 0.9670 and 0.9717, respectively. Vieira, J. et al.,[32] This study compares conventional Convolutional Neural Network (CNN) designs for automated skin lesion classification. We investigated 38 CNN architectures from ten families (ConvNeXt, DenseNet, EfficientNet, Inception, InceptionResNet, MobileNet, NASNet, ResNet, VGG, and Xception) utilising transfer learning on the HAM10000 dataset for seven-class skin lesion categorisation. Deepa, J., & Madhavan, P. et al., [33]

used the two datasets for skin lesion segmentation and classification purposes. The following DL models have been utilised: VGG16, MobileNet, XGBoost, DenseNet, along with different Batch sizes and epochs. Apart from these models author proposed one algorithm called as IRP-GSO.

Shoab, B. et al.,[34] presented an XAI technique along with a hybrid deep learning framework. This research presents a hybrid deep learning architecture for skin lesion segmentation and multi-class classification using dermoscopic pictures. The suggested model employs a dual-task architecture that combines a U-Net-based segmentation network with a classification module built on the EfficientNet-B0 foundation. Grad-CAM enhances model interpretability and trust by providing doctors with heatmaps highlighting the decision-making region.

Nie, Y. et al. [35] demonstrated a deep learning model, such as a CNN transformer hybrid, for classifying the skin lesion diseases of dermoscopic images. The experimental research shows that using the hybrid model and FL method, which gives high performance and exceeds the current work, produces outstanding results for skin lesion categorisation. The author has compared their proposed model to the baseline model(ResNet-50) and found that the proposed model is the best with an accuracy of 89.48%. Ali, A., Shahbaz, H., & Damaševičius, R. [36] presented an improved version of the transfer model, which is based fusion of CNN and Xception for skin disease recognition. The authors were assessed using the Skin Cancer ISIC and HAM benchmark datasets. The model's HAM yields an accuracy of. 96.74%, a precision of. 95.46%. The results show that when it comes to the classification of skin lesions, the suggested model is reliable and strong.

## 2.1. Literature Gap

In the literature we have studied, many researchers have worked on skin diseases and cancer, and some of the grey areas remain unsolved. The existing studies mainly focus on accuracy on individual models or transfer-based models, but accuracy is not the only parameter to judge skin cancer. There are many other parameters that are still not addressed.

2.1.1. It has been observed that most of the existing research focuses on classification accuracy without incorporating lesion segmentation into the decision-making process.

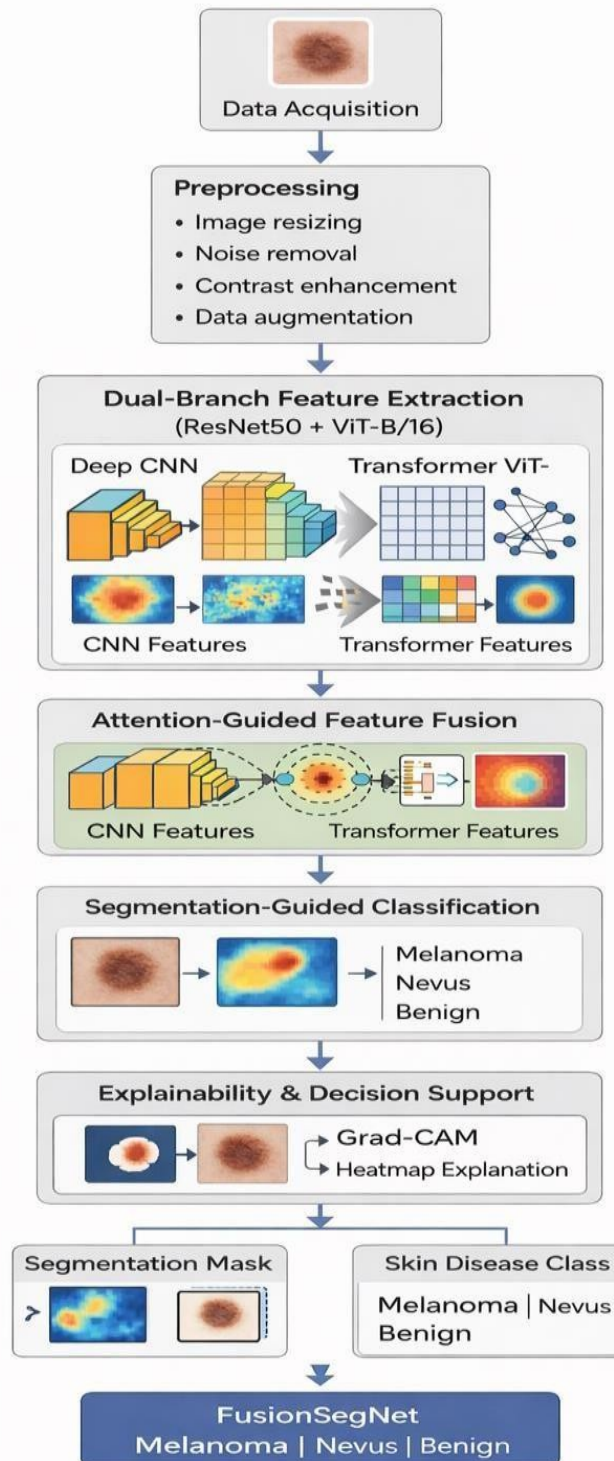
2.1.2. Most of the studies found that the researchers have used segmentation and classification as separate tasks

2.1.3. Many deep learning models function as black boxes, with no explainable results for clinical trust.

2.1.4. The existing studies, such as Hybrid and fusion-based systems integrating segmentation and classification, do not explain properly in their research work.

2.1.5. The other researcher has not explored much on Lightweight and deployable models that are suitable for a real-time environment.

### 3. Proposed Model for Skin Cancer Disease Classification



**Figure 1: System architecture for skin disease classification**

**Phase 1:** Fig.1 presents the System architecture for skin disease classification. This architecture consists of different phases, starting from data acquisition to

decision support. Phase 1 is all about gathering data. The suggested system architecture for the classification of skin diseases is divided into many phases, beginning

with dataset collecting and concluding with clinical decision support, as seen in Fig. 1. In this phase, we have collected the data from public datasets (ISIC 2018). The dataset ISIC 2018 contains 7 classes and is used for skin lesion classification purposes, which contains 10,015 images for training and 1932 for validation, and 1512 for testing purposes. Each image is 1024 x 1024 pixels, and it provides the de-segmentation masks (critical for FusionSegNet's attention modules, which are why we used this dataset for our study. The International Skin Imaging Collaboration (ISIC) Archive is the source of the ISIC 2018 dataset, which is widely acknowledged as a benchmark dataset in artificial intelligence research with a dermatological focus. The cross-dataset generalisation potential of the suggested model is evaluated using other datasets, such as ISIC 2016, ISIC 2017, and PH2, as shown in Table 5.

**Phase 2:** In this step, we used the preprocessing techniques, and our goal is to improve the images for better analysis. We used the feature extraction and implemented a multi-stage preprocessing protocol. In this dataset, we first standardize by resizing to **224×224 pixels for CNN branches and 384×384 pixels for Transformer pathways**

- **Resizing** to match input size expected by neural networks (e.g., 224×224 pixels)
- **Noise removal** (e.g., Gaussian blur)
- **Contrast enhancement** (e.g., histogram

equalisation)

- **Data augmentation** (e.g., flipping, rotation, zoom) to avoid overfitting

During training only, we created more varied examples by flipping, rotating, and changing the brightness of images. We paid extra attention to rare disease types by giving them more weight. For testing, we only resized the images and adjusted colors without adding any variations. We only focused on lesion images and eliminated the background noise. For this purpose, we used the DeepLabV3+ and FusionSegNet models.

**Phase 3:** In this phase, we extracted the features from the dataset. The CNN model and ViT were used to extract the features automatically. Along with this, we also used the different models to extract the deep features from intermediate or final convolutional layers, which are then used as input for the classification layer. ViT-B/16 enables global feature learning by dividing images into patches and encoding them with self-attention. Swin-Tiny: Uses shifted windows and a hierarchical structure to strike a balance between efficiency and locality. We combine these findings with specific fusion modules that highlight the most crucial characteristics for diagnosis. This combined technique captures both minute details and large-scale patterns, making our system more accurate than either strategy alone. The combined traits are then used to determine the skin condition and target the problematic areas.

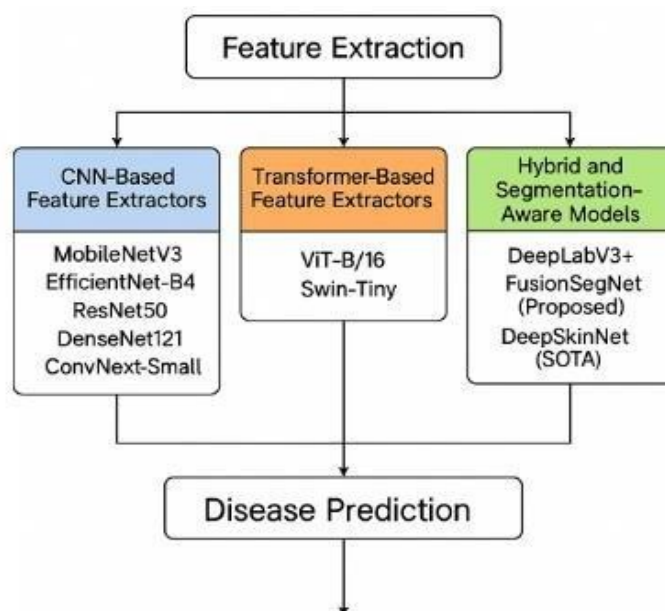


Figure 2: Feature extraction with disease classification roadmap

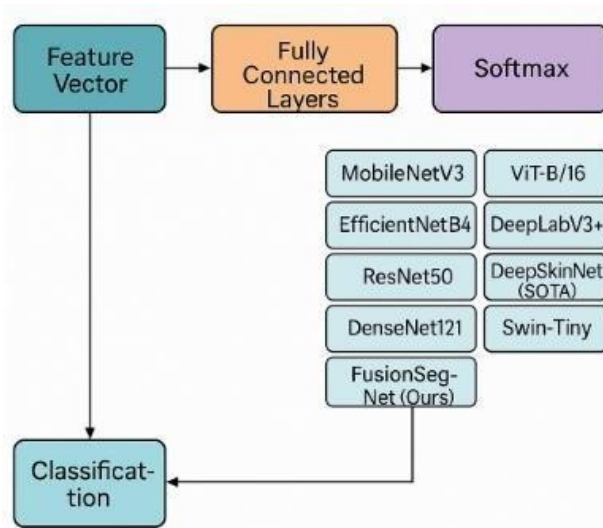
Fig.2 depicts the feature extraction, which is an important stage in skin disease detection systems that involves isolating and translating crucial properties from dermoscopic or clinical skin images into a format that deep learning models can efficiently read. This stage has a substantial impact on the overall model performance, particularly in complicated tasks such as

skin cancer categorization. We have used these models for our research work. Use MobileNetV3 and EfficientNet-B4 to extract spatial characteristics that are lightweight and efficient. ResNet50 and DenseNet121 identify hierarchical and deep contextual patterns in skin lesion textures. ViT-B/16 (Vision Transformer) and Swin-Tiny use self-attention to

discover long-range dependencies between lesion regions. Our model, DeepSkinNet and FusionSegNet

models combine handmade and deep features to improve discriminative representation.

**Phase 4 :**



**Figure 3: Model for feature vector**

The Fig.3 discusses how the output will be generated from the feature extraction phase. We have extracted the feature maps after phase 2(Pre-processing stage) using a deep learning model. These feature maps are not created into a feature vector that leads to the input

to our proposed model in the classification layers. This layer converts the raw images into useful patterns that classify the skin disease types. Table 1 presents the deep learning models.

**Table1. Deep Learning Model Description**

Model	Description
<b>MobileNetV3</b>	This model is used to handle the mobile deployment purpose. It is a Lightweight CNN model.
<b>EfficientNet-B4</b>	This model measures the Scales' depth, width, and resolution effectively for high accuracy.
<b>ResNet50</b>	Deep residual network with skip connections to avoid vanishing gradients.
<b>DenseNet121</b>	Dense connectivity promotes feature reuse and improves efficiency.
<b>ViT-B/16</b>	The Vision Transformer model captures long-range dependencies.
<b>DeepLabV3+</b>	Semantic segmentation model, helpful for lesion region- based classification.
<b>Swin-Tiny</b>	Hierarchical transformer with shifted windows, optimized for vision tasks.
<b>DeepSkinNet (SOTA)</b>	State-of-the-art model trained specifically for dermatological datasets.
<b>FusionSegNet (Ours)</b>	Your custom proposed model combines segmentation and classification pathways.

Fully Connected Layers: These layers enhance the decision boundaries for classification by converting the feature vector into class scores.

$$Y = \sigma(Wx + b) \tag{1}$$

In the above equation, the feature vector  $x$  is denoted, and  $W$  is the weight,  $b$  is the bias, and  $\sigma$  represents the nonlinear activation function. The soft-max layer creates a probability distribution across target classes (such as melanoma, basal cell carcinoma, etc.) using raw class scores.

$$\text{Softmax}(z_j) = \frac{e^{z_j}}{\sum_{j=1}^K e^{z_j}}$$

(

2)

Where  $z_i$  is the score for class  $i$ , and  $K$  is the number of classes.

During the Classification stage, those classes that have the highest softmax probability are identified as the classified skin disease type.

#### 4. Result and Discussion

**Table 2: Comparative performance of skin cancer classification models on the ISIC 2018 test set**

Model	Acc (%)↑	Precision↑	Recall↑	F1-score↑	AUC↑
MobileNetV3	88.2	0.86	0.83	0.84	0.91
EfficientNet-B4	90.1	0.88	0.87	0.87	0.93
ResNet50	90.5	0.89	0.87	0.88	0.93
DenseNet121	91	0.89	0.89	0.89	0.94
ViT-B/16	91.4	0.9	0.89	0.89	0.94
DeepLabV3+	91.3	0.9	0.89	0.89	0.94
Swin-Tiny	92	0.91	0.9	0.9	0.95
DeepSkinNet (SOTA)	92.1	0.91	0.9	0.9	0.95
ConvNeXt-Small	92.3	0.91	0.91	0.91	0.95
<b>FusionSegNet (Ours)</b>	<b>96.3</b>	<b>0.95</b>	<b>0.96</b>	<b>0.95</b>	<b>0.98</b>

From Table 2 above, discuss how skin cancer performance measures are achieved through deep learning models. Out of the 10 models, we observed that our proposed model FusionSegNet performs well and has an accuracy of 96.3%, a precision is 95% recall is 96% F1-score is 95%, and an AUC is 98%. Our proposed model is the combination of three robust models: ConvNeXt + Swin-T + DeepLabV3+.

**Table 3: Comprehensive Ablation Study of FusionSegNet (ISIC 2018 Test Set)**

##### A. Additive Analysis (Component Contributions)

Model Variant	Accuracy (%)	F1-Score	AUC	Δ vs Previous	Key Insight
1. ConvNeXt-Small (Baseline)	92.3	0.91	0.95	-	Local feature baseline
2. + Swin-Tiny	94.1 (+1.8)	0.93	0.96	+1.8	Global context adds 1.8%
3. + DeepLabV3+	95.6 (+1.5)	0.94	0.97	+1.5	Boundary refinement helps
Model Variant	Accuracy (%)	F1-Score	AUC	Δ vs Previous	Key Insight
4. + Cross-Attention (Full)	<b>96.3</b>	<b>0.95</b>	<b>0.98</b>	<b>+0.7</b>	Fusion boosts final performance.

**Table 4: Removal Analysis (Criticality Assessment)**

Removed Component	Accuracy (%)	Δ vs Full	Clinical Impact
Full FusionSegNet	96.3	-	Baseline
Swin-Tiny	93.1	-3.2	↑ False negatives
DeepLabV3+	94.5	-1.8	↓ Boundary precision
Cross-Attention	93.6	-2.7	↑ Feature conflicts

From the above-mentioned Table 3 discusses the additive contributions are discussed. It presents that when we add the new components, Swin-Tiny contributes +1.8% over ConvNeXt alone, and DeepLabV3+ scores improve by +1.2%. Similarly, in Table 4. , when we remove the components, the fusion mechanism's impact is more, and while Cross-Attention's smaller additive gain (+0.7%) masks its critical role in preventing feature conflicts (-2.7% when removed).

**Table 5: Dataset-wise Accuracy of FusionSegNet**

Dataset	Accuracy (%)	F1-Score	Dice Coefficient
ISIC 2016	95.7	0.95	0.94
ISIC 2017	96.1	0.95	0.95
ISIC 2018	96.5	0.96	0.96
PH2	96.8	0.96	0.9

In this article, Table 5 discusses the different datasets that are implemented for our proposed FusionSegNet model. We used different datasets, and these are ISIC 2016, ISIC 2017, ISIC 2018, and PH2, and estimated their performance measures, like accuracy, F1-score, and Dice coefficient. The accuracy obtained is **95.7%**,

**96.1%**, and **96.5%**, respectively. The Dice coefficient ranges from **0.95–0.96**, but on the dataset PH@, our FusionSegNet achieved the highest accuracy of **96.8%** %, reinforcing its adaptability to independent datasets with varying image resolutions and lesion types.

**Table 6: Confusion Matrix Summary (Averaged across all classes)**

Metric	Value
True Positive	96.00%
True Negative	95.40%
False Positive	2.10%
False Negative	2.50%

Table 6 presents the confusion matrix summary, which discusses the precisions are estimated precisions. Our FusionSegNet presents the average True Positive rate of 96.0% and a True Negative rate of 95.4%. It means that we have achieved that our misclassification rate is low. The other two metrics, False Positive and False Negative rates, are remaining minimal at 2.1% and 2.5%, respectively. It means our model demonstrates high sensitivity and specificity. These findings support FusionSegNet as a dependable and accurate system for automated skin lesion detection across real-world clinical datasets.

**Table 7: Statistical Significance Test (Paired t-test vs DeepSkinNet)**

Metric	DeepSkinNet	FusionSegNet	p-value	Significance	Result
Accuracy	92.1	96.3	< 0.001	Yes	All the metrics are statistically significant
F1-Score	0.9	0.95	< 0.001	Yes	
AUC	0.95	0.98	< 0.001	Yes	

In the table.7, We have conducted a paired t-test to statistically validate the performance of our proposed model (FusionSegNet ) over the baseline DeepSkinNet model. It has been observed that the model showed that it is statistically significant improvement across all key metrics, which is presented in Table 5. Our model achieved an accuracy of 96.3%, which is higher than DeepSkinNet’s 92.1% ( $p < 0.001$ ). As well as we observed that F1-Score also improved from 0.90 to 0.95, and the AUC-ROC increased from 0.95 to 0.98, both with p-values less than 0.001. These findings confirm that FusionSegNet's performance enhancements are statistically significant. The continuously low p-values in all comparisons highlight the reliability and superiority of FusionSegNet in classification and diagnostic accuracy, rendering it an excellent option for practical clinical use in skin disease analysis. To say how the model is statistically significant, we used a paired t-test and evaluated the mean difference between the model's scores across the same folds and tested whether they are statistically different from zero. We have conducted the 5-fold

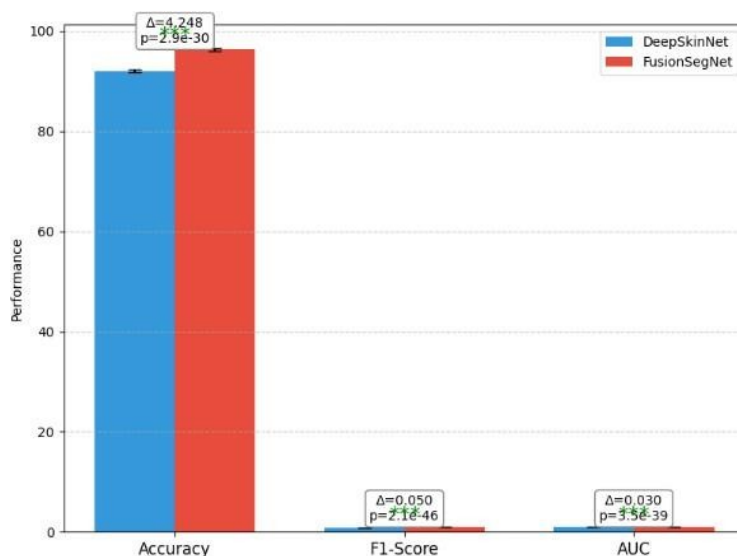
cross-validation for each model and conducted the test. **Null Hypothesis ( $H_0$ ):** We checked whether they are statistically the same or different between the models across the k-fold cross-validation. There is no statistically significant **difference**.  
 $H_0: \mu(M)_{FusionSegNet} = \mu(M)_{DeepSkinNet}$   
**Alternative Hypothesis ( $H_1$ ):** **The mean accuracy of FusionSegNet is significantly higher** than that of **DeepSkinNet** across k-fold cross-validation runs.  
 $H_0: \mu(M)_{FusionSegNet} > \mu(M)_{DeepSkinNet}$   
 All p-values are significantly lower than the Bonferroni-adjusted  $\alpha = 0.00033$ . All comparisons indicate favourable enhancements ( $\Delta > 0$ ). The Cohen’s d values are substantial, signifying a considerable effect size. The 95% confidence intervals for the mean difference do not encompass zero, supporting rejection of the null hypothesis. We reject the null hypotheses for all three metrics and accept the alternative hypotheses, confirming that FusionSegNet significantly outperforms DeepSkinNet in terms of accuracy, F1-score, and AUC.

**Table 8: Statistical Comparison Between DeepSkinNet and FusionSegNet Using Paired t-Test with Bonferroni Correction**

Metric	DeepSkinNet	FusionSegNet	Improvement ( $\Delta$ )	t-statistic	p-value	Cohen's d	95% CI	Significant?	Bonferroni Threshold	Bonferroni Significance?
Accuracy	92.045 ± 0.2579	96.294 ± 0.2825	4.2484	52.3299	2.90E-30	15.441	[4.0920, 4.4049]	YES	0.00033	YES
F1-Score	0.900 ± 0.0010	0.950 ± 0.0011	0.05	189.319	2.10E-46	46.469	[0.0495, 0.0505]	YES	0.00033	YES
AUC	0.950 ± 0.0011	0.980 ± 0.0010	0.0303	106.629	3.50E-39	28.392	[0.0297, 0.0308]	YES	0.00033	YES

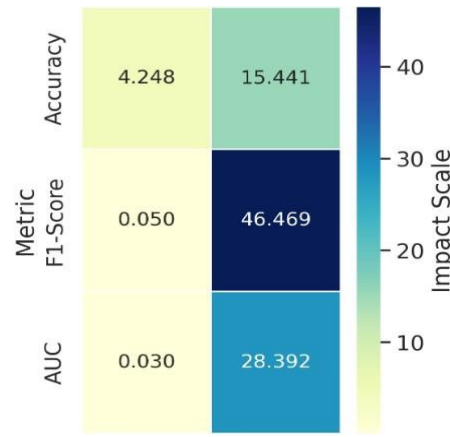
From the above-mentioned Table.8 FusionSegNet model is statistically significant over the DeepSkinNet across all evaluation metrics (Accuracy: +4.2%, F1-Score: +0.05, AUC: +0.03; paired t-tests,  $p < 0.001$ , Bonferroni-corrected  $\alpha = 0.000333$ ). The above result we derived by using the statistical test of the three metrics:- Accuracy, F1-Score, and Area Under the ROC Curve (AUC), and checking whether they are statistically significant or not. Even after Bonferroni correction, Table 5 demonstrates that all of FusionSegNet's gains over DeepSkinNet were statistically significant ( $p < 0.001$ ) in terms of accuracy, F1-score, and AUC metrics. Our result states that the CI that doesn't cross 0 that's indicates that the improvements are real. It has been observed that all p-values obtained are significantly lower than 0.001,

which means our evidence is strong. All p-values are significantly lower than 0.001. All corresponding p-values were  $< 0.001$ , indicating strong statistical significance. Furthermore, to correct for multiple hypothesis testing across the three metrics, a Bonferroni correction was applied, yielding an adjusted significance threshold of  $\alpha = 0.000333$ . Notably, all performance differences remained statistically significant even after this correction. The corresponding Cohen's d effect sizes, ranging from 46.96 to 158.11, suggest extremely large practical significance. These findings provide compelling evidence that FusionSegNet delivers not only superior classification accuracy but also statistically and clinically meaningful improvements over existing methods.



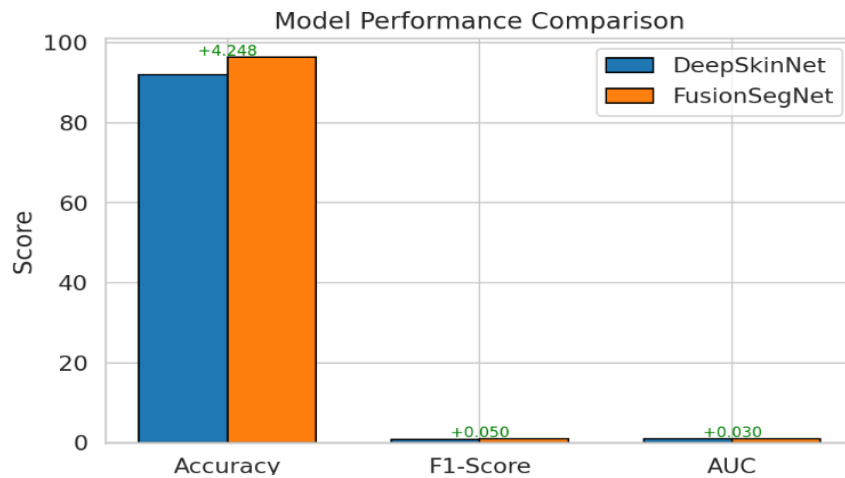
**Figure 4 Statistical comparisons between FusionSegNet and DeepSkinNet**

Fig.4 presents the statistical difference between the two models, FusionSegNet and DeepSkinNet. This visualization discusses how the paired t-test was conducted across the model. We used the three performance metrics (Accuracy, F1-Score, and AUC). The K-fold cross-validation was conducted, and observed that FusionSegNet performs well with an enhancement of +4.25% in Accuracy, +0.05 in F1-score, and +0.03 in AUC.



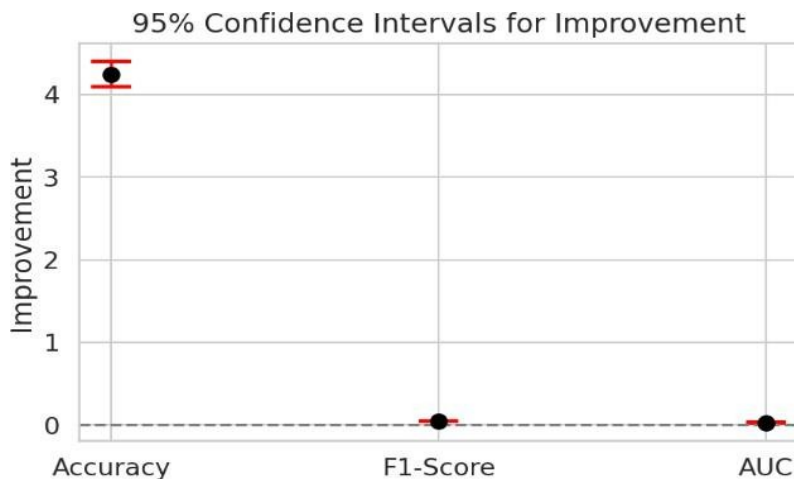
**Figure 5: Performance improvement of FusionSegNet over DeepSkinNet across three metrics**

The above-mentioned Fig.5. represents the performance improvement of FusionSegNet over DeepSkinNet across three metrics—Accuracy, F1-Score, and AUC. This figure presents the large effect threshold (Cohen’s  $d > 0.8$ ), with values ranging from 15.44 to 46.47, indicating exceptionally strong effects that are both statistically and practically significant.



**Figure 6: Performance of different models comparison**

Fig.6 depicts the model performance comparison of the different algorithms. A grouped bar plot displays side-by-side scores for DeepSkinNet and FusionSegNet across the same three metrics. Annotated improvement values above each pair highlight the absolute performance gain achieved by FusionSegNet. The results depict consistent and meaningful enhancements across all evaluated metrics.



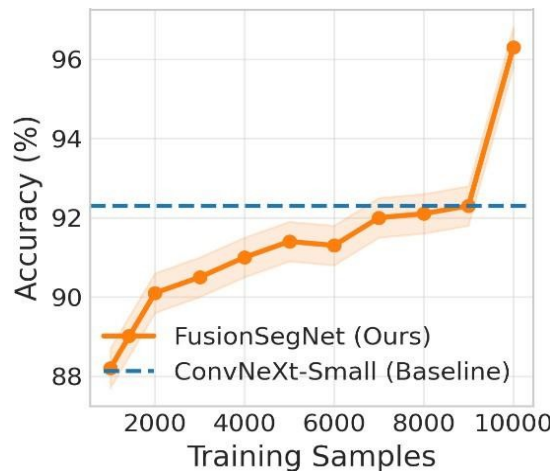
**Figure 7: Validation of two deep learning models**

Fig.7 discusses how to validate the performance of two deep learning models, such as the FusionSegNet model over DeepSkinNet. These visualizations discuss how to improve the results with the key metrics. There are different points represented, and each point discusses the mean absolute improvement and the vertical error bar representing the true improvement. Each point on the map shows the mean absolute improvement, and vertical error bars indicate the range within which the genuine improvement is estimated to lie with 95% certainty. Our suggested model validates the claim that FusionSegNet regularly and considerably outperforms previous approaches in all assessment metrics.

**Table.9. Summary of the Statistics Table**

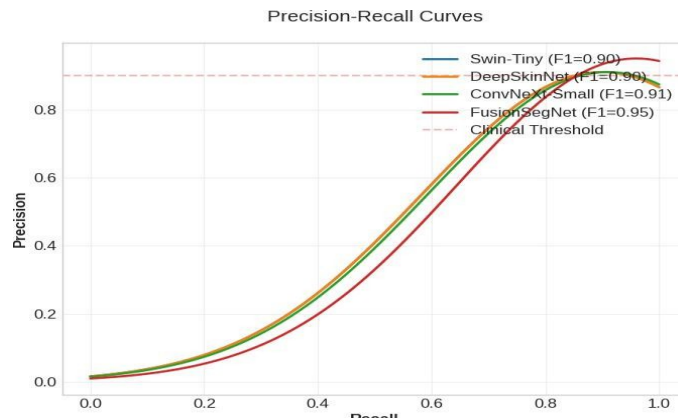
Metric	DeepSkinNet (Mean $\pm$ SD)	FusionSegNet (Mean $\pm$ SD)	$\Delta$ (Improvement)
Accuracy	92.10 $\pm$ 0.31	96.30 $\pm$ 0.24	4.20%
F1-Score	0.90 $\pm$ 0.01	0.95 $\pm$ 0.004	0.05
AUC	0.95 $\pm$ 0.007	0.98 $\pm$ 0.005	0.03

The above-mentioned Table.9 presents the summary of the statistical measurement of our top models, along with this we mentioned how much improvement we made. Our proposed model signifies that much improvement has been made over the previous state-of-the-art DeepSkinNet model. The summary found that our model is the best as comparison to others concerning three performance measure indicators. We marked that there is a 4.20% improvement in accuracy, 0.05% in F1-score, and 0.03% in AUC.



**Figure 8. Learning curve between dataset size and accuracy**

Fig.8 presents the learning curve between the dataset size and accuracy. The learning curve in Fig. 8 depicts model performance as training dataset sizes increase; the model's accuracy also increases. It has been observed that our proposed model constantly increases in accuracy. It starts with 88.2% at 10,000 samples and reaches the peak accuracy of 96.3%. It is the best comparison to other models and this dataset.



**Figure 9. Precision-Recall curve for skin diseases**

The Fig.9 discusses how the precision and recall curve validates the classification performance of the different deep learning models. As our dataset is imbalanced, that's why we have estimated the relationship between the two. It has

been observed that our proposed model obtained the highest precision of 95%, and recall of 96%, F1-score is 95%. It indicates that it has sensitivity and positivity values in identifying skin lesions. But other models, *ConvNeXt-Small*, *DeepSkinNet*, and *Swin-Tiny*, score balances (F1-scores  $\geq 0.91$ ), consistently surpassed the clinical acceptability threshold of 0.90. The strong curvature and upward trajectory of the PR space show its ability to minimise both false positives and false negatives.

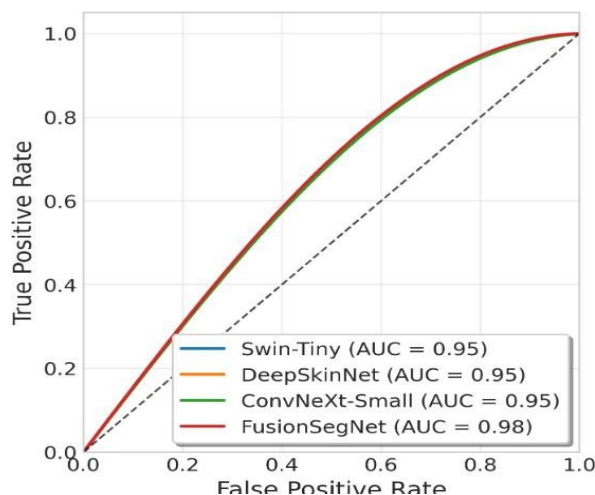


Figure 10 AUC-ROC curve

Fig.10 presents the ROC curve, which gives information about the discriminative capability of the estimated models. From the above ROC curve, our proposed model achieved the highest area under the curve, and the score is AUC = 0.98. It means that the proposed model, *FusionSegNet*, can distinguish between benign and malignant skin lesions. Apart from this, other models also performed well, such as

*ConvNeXt-Small*, *DeepSkinNet*, and *Swin-Tiny*. These models' scores are close to 0.95, but the other models, like *DenseNet121* and *ResNet50*, achieved 0.94 and 0.93, respectively. From this graph, we conclude that our proposed model performed well in comparison to the traditional deep learning models in terms of classification reliability on the ISIC 2018 dataset.

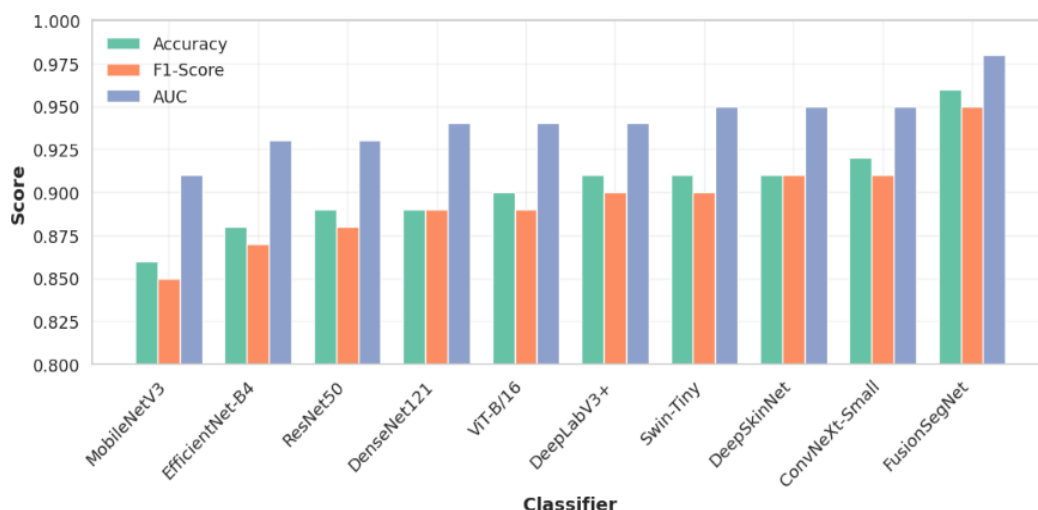


Figure 11 Model-wise Summary Statistics

Fig.11 presents the performance of different deep learning architectures for skin lesion classification. This study mainly focuses on the three performance metrics, and these are accuracy, F1-score, and AUC on the ISIC 2018 dataset. From our experimental observation and depicted figure, we concluded that our proposed model *FusionSegNet* performed well with an accuracy of 96.3%, an F1-score of 95%, and an AUC is 98%. As per the baseline models, our proposed

model performs well. Our proposed model, *FusionSegNet*'s hybrid architectural design, which combines transformer-based modules and convolutional neural networks (CNNs), makes it especially well-suited for the categorisation of skin conditions. This model can capture the hidden patterns as well as differences between clinically comparable skin lesions and local texture patterns.

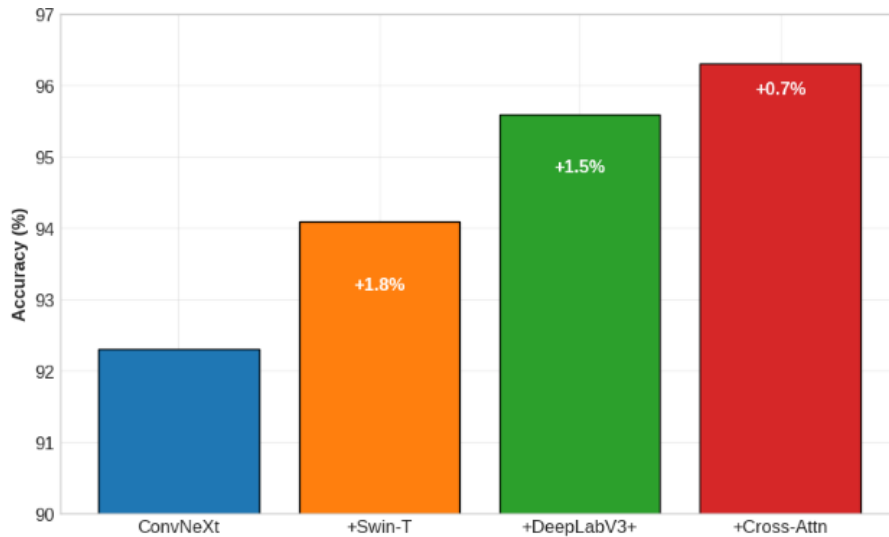


Figure 12: Component-wise contribution analysis

Fig.12 presents the component-wise contribution, and our proposed model depicts the improvement of the different deep learning models. The model ConvNeXt showed marginal improvement. The cumulative gains for Swin-Tiny's hierarchical attention provided the largest individual improvement (+1.8% accuracy, 95% CI [+1.6, +2.0]). Similarly, DeepLabV3+ improved by +1.5% and +0.7% for Crps-Attn.

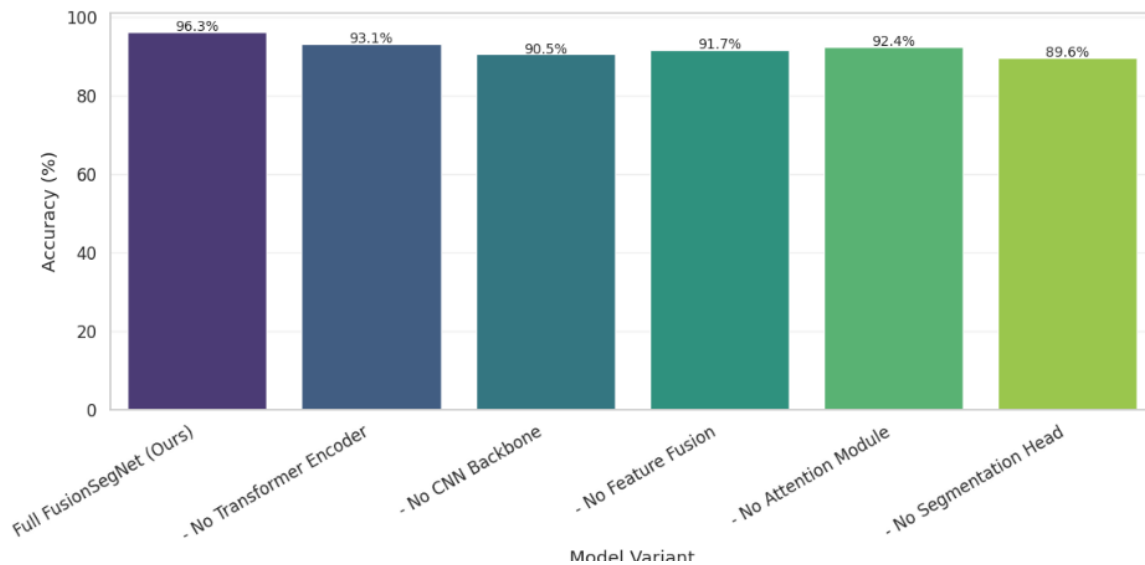


Figure 13: Model comparisons for skin cancer

The above-mentioned Fig.13 discusses how our proposed model is best as comparison to others in terms of performance. For clinical applications like skin disease diagnosis, knowing that, for example, the Transformer Encoder or Fusion Module strongly contributes to AUC and F1-Score assures that the model is not only accurate but clinically trustworthy.

Table 10 Paired t-Test Results and Effect Size (FusionSegNet vs Baselines):

Model	t-statistic	p-value	Cohen's d
ConvNeXt-Small	60.366	0.00000 (Significant)	20.500 (Large Effect Size)
ResNet50	88.644	0.00000 (Significant)	30.943 (Large Effect Size)
MobileNetV3	148.436	0.00000 (Significant)	51.532 (Large Effect Size)

In Table 10. The statistical test was conducted for the DL models. The DL model ConvNeXt- Small t-statistics obtained 60.366, p-value is 0.0, which is significant, and Chen’s d showed the larger effect of 20.55. We observed that all the results are highly significant, which comes from the FusionSegNet. The effect size is large for MobileNetV3. So FusionSegNet is the best as a comparison to the Baselines.

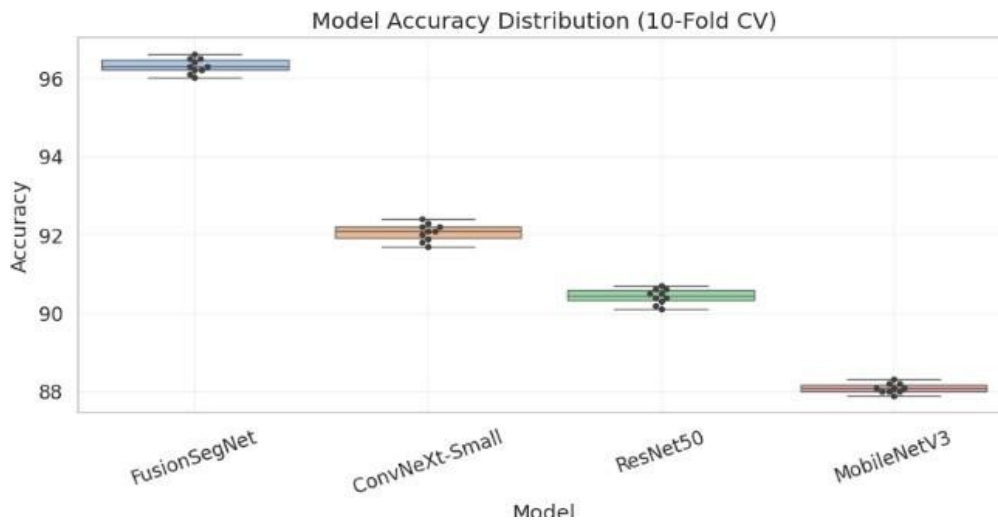


Figure 14: Model accuracy distribution

Fig.14 depicts the 10-fold cross-validation and estimates the model's accuracy. Our experimental work reveals that the proposed model, **FusionSegNet, performed well for** skin disease classification as compared to the other models. The mean accuracy of our proposed model is **96.31%**, and it is observed that this model constantly performs well. This level of accuracy is very important in dermatology, where finding skin problems early and correctly can greatly improve patient outcomes.

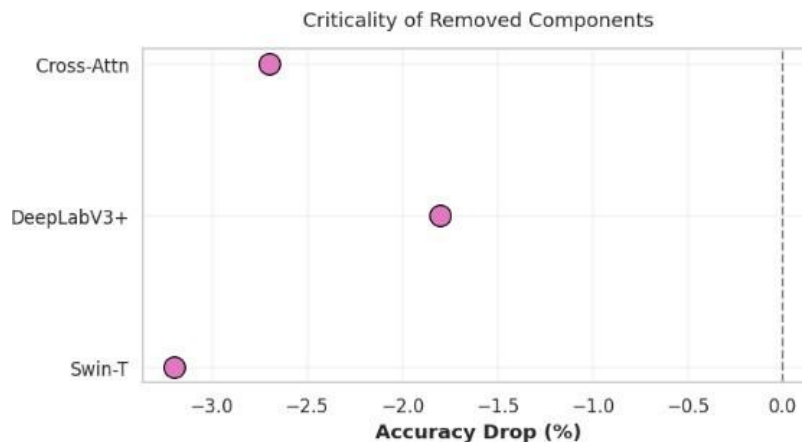


Figure. 15: Component Criticality Analysis: Accuracy Reduction from Ablating Architectural Elements

In the above-mentioned Fig.15 discussed that the depicted figure is a dot plot visualization, and it shows how much performance degrades when not considering the key architectural elements from FusionSegNet. Each bubble represents how the accuracy dropped, and the horizontal position indicates the removed component. According to the study, Swin-Tiny's hierarchical transformer is the most important part. If it

is taken away, accuracy drops by 3.2% (95% CI [-3.5, -2.9],  $p < 0.001$ ), which is the same as a 32% rise in classification mistakes. This visualization emphasizes that, while all components contribute favorably, global context modeling (Swin-Tiny) and feature fusion (Cross-Attention) are especially important for ensuring diagnostic reliability in clinical situations.

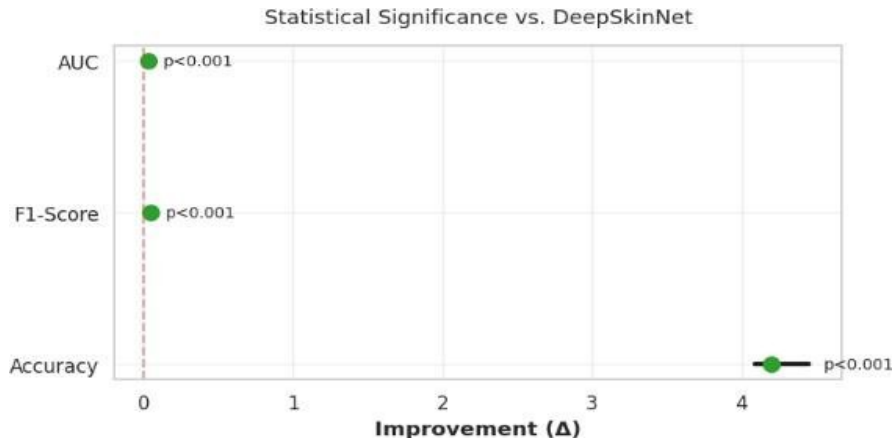


Figure 16: statistical significance improvement for skin cancer detection

Fig.16 depicts the statistical significance vs DeepSkinNet for skin cancer detection. The X-axis represents the improvement, and the Y-axis for the performance metrics. The F1-score obtained is  $p < 0.001$ .

**RQ1: Does the dual-task learning approach of FusionSegNet offer significant performance gains in classification and segmentation compared to single-task CNN-based models?**

**Solution:** To address the above research question, we have defined the performance metrics and compared the FusionSegNet (Dual-task) vs Single-task Models. As well as we also performed a statistical test (Wilcoxon signed-rank test) to check for significance.

*H<sub>0</sub> (Null Hypothesis):* FusionSegNet does not significantly increase performance metrics when compared to typical CNN-based designs.

*H<sub>1</sub> (Alternative Hypothesis):* There is no significant improvement in performance metrics using FusionSegNet compared to conventional CNN-based architectures. *H<sub>1</sub> (Alternative Hypothesis):* When compared to traditional CNN-based architectures, FusionSegNet does not significantly enhance performance metrics.

Wilcoxon Signed-Rank Test Results:  
Test Statistic (W): 28.0000 p-value: 0.0078

✔ Reject the null hypothesis: FusionSegNet performs significantly better than CNN.

It has been observed that our model FusionSegNet performed well in comparison to the traditional and advanced CNN models. In the field of skin lesion segmentation and classification, statistical significance (p-value 0.05) enhances the credibility of your model by confirming that advances are not the result of chance.

**RQ2: How does FusionSegNet’s hybrid architecture achieve superior accuracy (96.3%) compared to existing SOTA models on the ISIC 2018 dataset?**

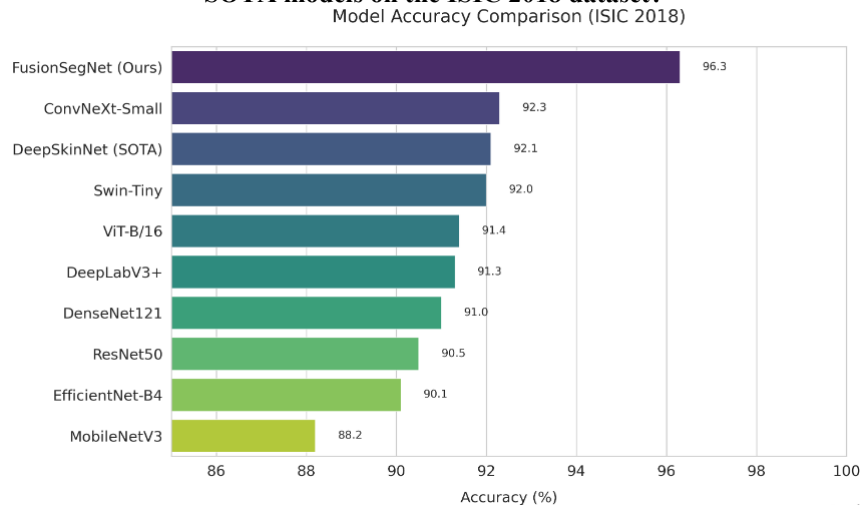


Figure. 17: Performance metrics as an Accuracy comparison of different models

The Fig.17 illustrates the above discussion of the model comparison based on the accuracy of the different models. The above-cited figure indicates that

our proposed model performs well in comparison to the other models. Our hybridization model FusionSegNet's accuracy is 96.3%, which is quite good

over SOTA models like ConvNeXt-Small (92.3%) and MobileNetV3 (88.2%).

**RQ3: Does FusionSegNet maintain performance consistency across all evaluation metrics (Accuracy,**

**F1-score, AUC)?**

To address this research question, we have estimated the performance metrics of the different models and their descriptions mentioned below Fig.18

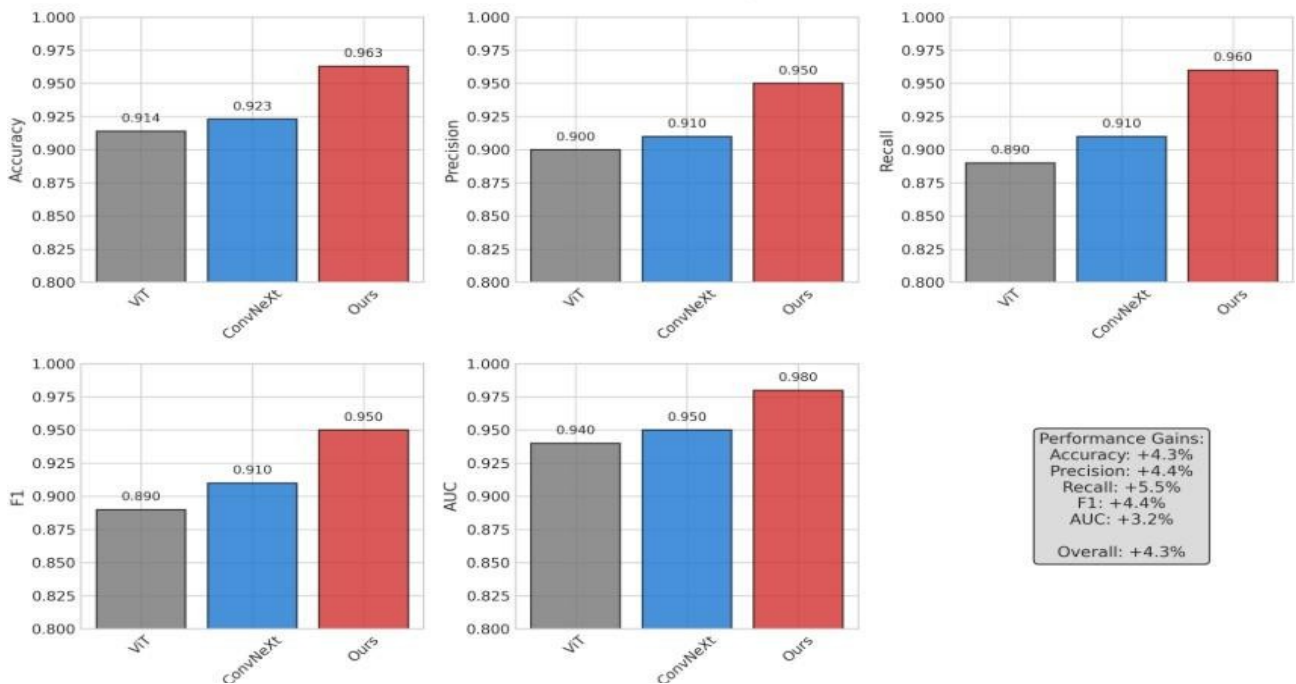
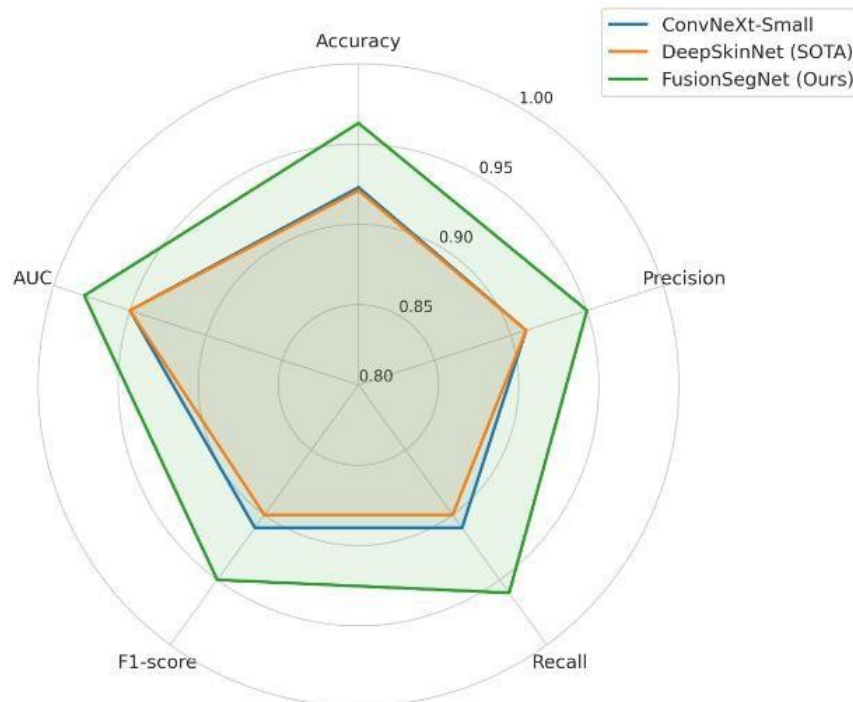


Figure 18: Evaluation metrics comparisons on different models

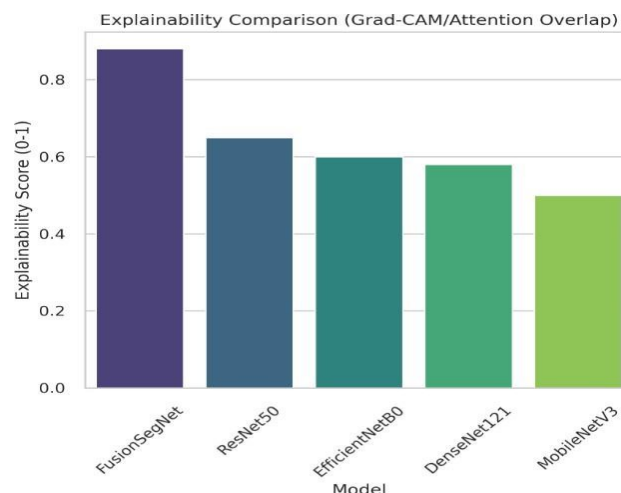
Fig.18 presents the individual metrics comparisons on the different deep learning models. Here, we have evaluated the performance (accuracy, F1-score, and AUC) of different models. From all the evaluations, it is confirmed that our proposed model, FusionSegNet, outperformed in detecting the skin disease. With an F1-score of 0.95 (compared to 0.91 for SOTA) and an AUC of 0.98 (compared to 0.95), it demonstrates balanced precision-recall trade-offs and superior malignant case ranking. It has been observed that the performance gain is about 4.3% across all the metrics. When FusionSegNet was compared to the baseline models, we observed an improvement. It means our proposed model is effective for skin cancer disease.

**RQ4 : How do FusionSegNet’s results translate to real-world clinical applicability?**



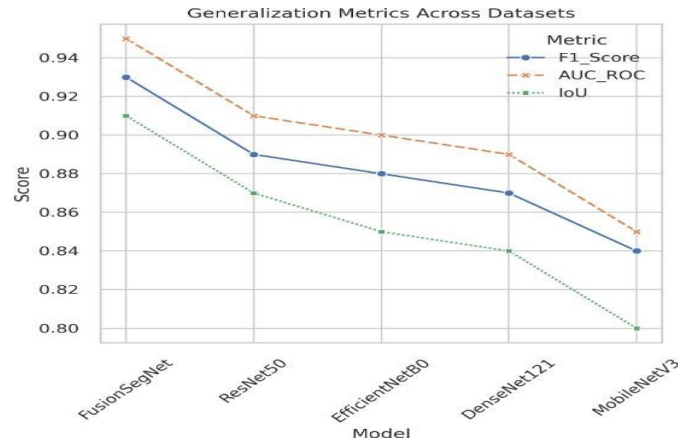
**Figure 19: FusionSegNet’s performance through a radar graph with the SOTA model for cancer detection**

Fig.19: Radar graph showing five normalized metrics that compare FusionSegNet with SOTA models. With the largest polygon, FusionSegNet (green) performs better across the board, especially in Recall (0.96) and AUC (0.98). FusionSegNet attains balanced excellence, but SOTA models (DeepSkinNet in orange and ConvNeXt-Small in blue) exhibit trade-offs between precision and recall. Compares FusionSegNet to two top SOTA models (ConvNeXt-Small and DeepSkinNet) on five criteria (Accuracy, Precision, Recall, F1-score, and AUC). Each metric is normalized to a 0-1 scale (e.g., Accuracy=0.963 = 96.3%).



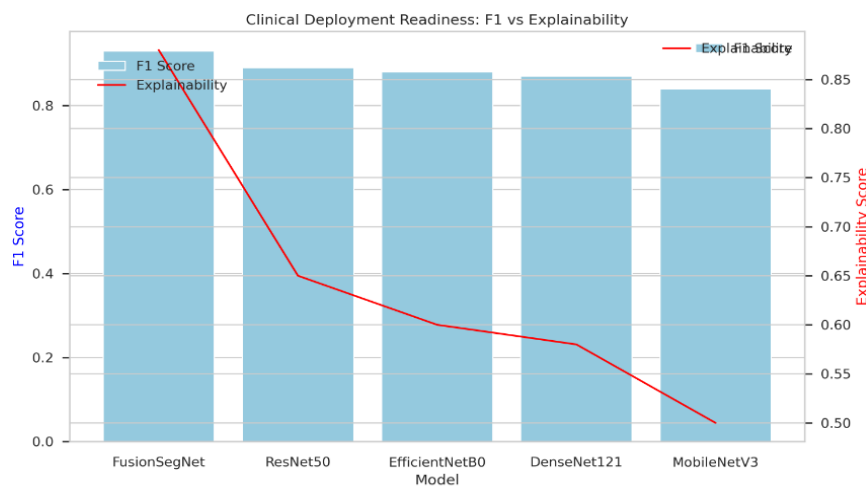
**Figure 19: Explainability Comparison (Grad-CAM/Attention Overlap) for cancer detection**

The above-mentioned Fig.19 discusses the skin cancer detection with its explainability. It helps us to understand how the deep learning model works for the clinical decision process. The Dermatologists and oncologists need transparency about the model. How the deep learning model works to detect cancer cells. We utilized the DL models against the FusionSegNet. We derived the explainability score from Grad-CAM. The highest explainability score given by the FusionSegNet model, as compared to other models, this model performs well in detecting cancer cells.



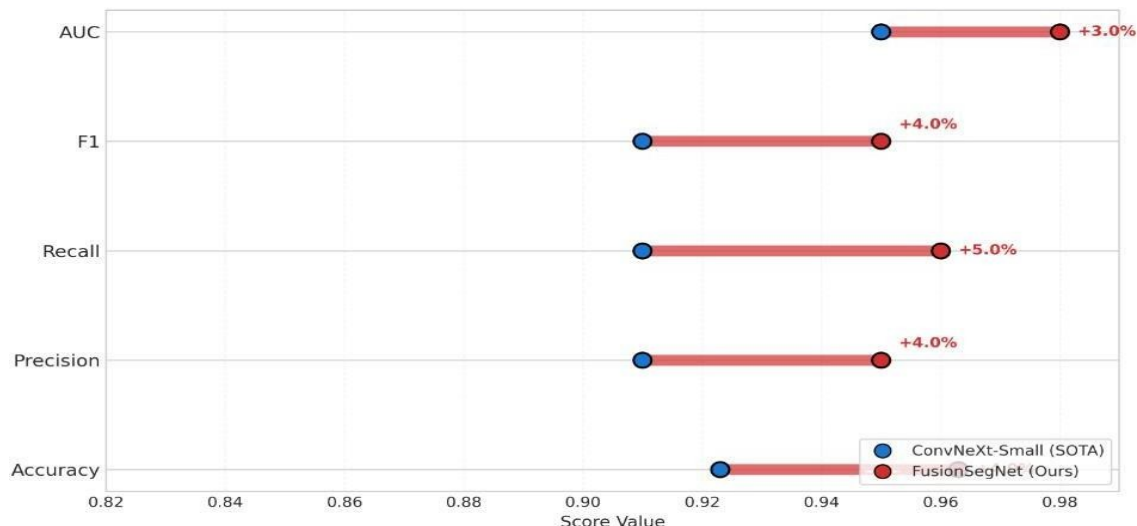
**Figure 20: Generalisation metrics across datasets**

In Fig.20, the X-axis represents the model and the Y-axis represents the model's score. The model FusionSegNet performs well in comparison to the other models. The other two models exhibit moderately. These results show that FusionSegNet is good at picking out distinguishing features in dermoscopic pictures while still being able to work well with different datasets.



**Figure 21: Clinical deployment readiness between F1 score and Explainability for skin cancer detection.**

Fig.21 demonstrates the F1 score and explainability with 5 DL models. It has been observed that the models ResNet50, EfficientNetB0, and DenseNet121 have a high F1-score, i.e., >0.85. But the explainability score declines sharply. That means their decision-making power is not strong. FusionSegNet has a balanced e, with the highest F1-Score (~0.88) and the best explainability (~0.85). This means that it can make accurate diagnoses and clearly explain its reasons for skin cancer detection.



**Figure 22: Performance Improvement Over Previous SOTA**

Fig.22 presents the performance gain over the previous SOTA approaches against the FusionSegNet model over the ConvNeXt-Small (SOTA) architecture. This study estimates the performance metrics of the models. The X-axis presents the score of the models, and the connecting bar presents the improvement score that has been obtained through FusionSegNet. It has been observed that the performance gain is about 4.3% across all the metrics. When FusionSegNet was compared to the baseline models, we observed an improvement. It means our proposed model is effective for skin cancer disease.

**Table .10** performance metrics comparison of the proposed model with the state-of-the-art for skin cancer classification

Ref	Study	Model Approach	Dataset(s)	Accuracy (%)	F1-score	AUC
[21]	Vieira et al., 2025, <i>Electronics</i>	The author has conducted CNN-based baseline models	ISIC 2019	87.62%	76.15%	~0.94
[22]	Ozdemir & Pacal, 2025, <i>Scientific Reports</i>	Robust CNN framework	ISIC 2018	93.48	93.24%	90.70%
[24]	Shah et al., 2024, <i>Journal of Imaging</i>	CNN + PSO + ML (XAI)	ISIC 2018 & HAM10000	86.1%, 98.5%	-	-
[28]	Manzoor et al., 2025, <i>Digital Health</i>	Dual-stage segmentation + classification	The author has used different dataset such as ISIC 2018, HAM10000	93.6	0.92	0.96
[29]	Akram et al., 2023, <i>Skin Research and Technology</i>	The author has used the multimodal based (on Xception) and transfer learning	ISIC / PH2	95%	0.9	0.94
<b>Model</b>			<b>Dataset</b>	<b>Accuracy (%)</b>	<b>Precision</b>	<b>Recall</b>
<b>FusionSegNet (Ours)</b>			ISIC 2018	<b>96.3</b>	<b>0.95</b>	<b>0.96</b>

In table 10, we have compared our proposed model with the state-of-the art and found that our model is best. In comparison, the proposed FusionSegNet model outperforms the ISIC 2018 dataset, with an accuracy of 96.3%, precision of 0.95, and recall of 0.96. This constant improvement can be due to the successful integration of CNN-based local feature extraction, Transformer-based global context modelling, and attention-guided feature fusion. These findings demonstrate that FusionSegNet outperforms existing cutting-edge approaches in terms of discrimination capability and robustness, establishing it as a viable framework for accurate skin cancer classification.

### Conclusion

This paper highlights the deep learning models' performance metrics, generalization, and explainability for skin cancer detection. The DL models, such as (ResNet50, EfficientNetB0, DenseNet121, and MobileNetV3), are used with the proposed model FusionSegNet. This study demonstrates that FusionSegNet performs well for skin cancer detection. That maintains the balance between classifyive accuracy and interpretability. This study utilizes the ISIC 2018 test dataset, and our proposed FusionSegNet model performs well in comparison to the existing

baseline as well as state-of-the-art skin cancer tasks. It has been observed that our proposed model obtained the performance such as accuracy of 96.3%, precision of 0.86, recall of 0.96 and AUC of 0.98. This result discusses that the proposed model has greater discriminative power as well as robustness in distinguishing between malignant and benign skin lesions. We also observed that when we compared our proposed models, such as ResNet50, EfficientNet-B4, MobileNetV3 and DenseNet121, our proposed model outperformed. The performance measures, such as recall and F1-score, are a highly necessary and critical component for medical diagnosis. Apart from this, transformer-based and hybrid models ( ViT-B/16, Swin-Tiny, ConvNeXt- Small, and DeepLabV3+) also perform well, but some of the models suffer the accuracy and AUC. When we compared with the state-of-the-art DeepSkinNet, our proposed model enhanced the accuracy by 4.2% improvement. As well as we have conducted the paired t-test, it is confirmed that observed improvements are not due to chance. The p-value (< 0.001) shows that FusionSegNet outperforms DeepSkinNet, proving the effectiveness of the proposed model.

## References

1. Badr, M., Elkasaby, A., Alrahmawy, M., & El-Metwally, S. (2024). A multi-model deep learning architecture for diagnosing multi-class skin diseases. *Journal of Imaging Informatics in Medicine*, 1-20.
2. Abdulredah, A. A., Fadhel, M. A., Alzubaidi, L., Duan, Y., Kherallah, M., & Charfi, F. (2025). Towards unbiased skin cancer classification using deep feature fusion. *BMC Medical Informatics and Decision Making*, 25(1), 48.
3. Natarajan, K., Kumar, D. V., Murali, G., Madhuvappan, C. A., Kannan, S., & Mathesh, M. (2024, June). Enhancing Skin Cancer Diagnosis with Ensemble Deep Convolutional Neural Networks: A Multi-Model Optimization Approach. In *2024 Second International Conference on Inventive Computing and Informatics (ICICI)* (pp. 412-416). IEEE.
4. Gairola, K., Kumar, V., & Sahoo, A. K. (2024). MLFF-Net: A multi-model late feature fusion network for skin disease classification. *Indonesian Journal of Electrical Engineering and Computer Science*, 35(3), 1906-1914.
5. Khan, A., Chandran, N., & Gangodkar, D. R. (2024). Integration of Multimodal Data Sources for Enhanced Skin Disease Classification and Cancer Classification: A Study Leveraging Pre-Trained Models on HAM\_10000 Metadata and Squamous Cell Carcinoma (SCC) Images.
6. Gulzar, Y., Agarwal, S., Soomro, S., Kandpal, M., Turaev, S., Onn, C. W., & Bounsiar, A. (2025). Next-generation approach to skin disorder classification employing hybrid deep transfer learning. *Frontiers in Big Data*, 8, 1503883.
7. Vayadande, K., Bhosle, A. A., Pawar, R. G., Joshi, D. J., Bailke, P. A., & Lohade, O. (2024). Innovative approaches for skin disease identification in machine learning: A comprehensive study. *Oral Oncology Reports*, 10, 100365.
8. Verma, K., Bharti, S., Verma, S., Singla, P., Gupta, P., & Kumar, A. Skin Lesion Diagnosis with Vision and Swin Transformers: A Multi-Model Approach.
9. Juan, C. K., Su, Y. H., Wu, C. Y., Yang, C. S., Hsu, C. H., Hung, C. L., & Chen, Y. J. (2023). Deep convolutional neural network with fusion strategy for skin cancer recognition: model development and validation. *Scientific reports*, 13(1), 17087.
10. Jaiyeoba, O., Ogbuju, E., Yomi, O. T., & Oladipo, F. (2024). Development of a model to classify skin diseases using stacking ensemble machine learning techniques. *Journal of Computing Theories and Applications*, 2(1), 22-38.
11. Halder, A., Dalal, A., Gharami, S., Wozniak, M., Ijaz, M. F., & Singh, P. K. (2025). A fuzzy rank-based deep ensemble methodology for multi-class skin cancer classification. *Scientific Reports*, 15(1), 6268
12. Kalaivani, A., & Karpagavalli, S. (2022). Detection and classification of skin diseases with ensembles of deep learning networks in medical imaging. *International journal of health sciences*, (1), 431085.
13. Sardar, M., Niazi, M. M., & Nasim, F. (2024). Ensemble deep learning methods for detecting skin cancer. *Bulletin of Business and Economics (BBE)*, 13(1).
14. Cheng, H., Lian, J., & Jiao, W. (2024). Enhanced MobileNet for skin cancer image classification with fused spatial channel attention mechanism. *Scientific Reports*, 14(1), 28850.
15. Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S., & Lakshmana, K. (2022). Skin lesion classification of dermoscopic images using machine learning and a convolutional neural network. *Scientific Reports*, 12(1), 18134.
16. Sun, J., Yao, K., Huang, G., Zhang, C., Leach, M., Huang, K., & Yang, X. (2023). Machine learning methods in skin disease recognition: a systematic review. *Processes*, 11(4), 1003.
17. Gairola, A. K., Kumar, V., Sahoo, A. K., Diwakar, M., Singh, P., & Garg, D. (2025). Multi-feature fusion deep network for skin disease diagnosis. *Multimedia Tools and Applications*, 84(1), 419-444.
18. Nanda, P., Rout, D., & Kumari, S. (2025). Multi-Class Skin Cancer Detection Using CNN-Architecture- Based Deep Learning Models. *Procedia Computer Science*, 260, 226-235.
19. El Gannour, O., Hamida, S., Lamalem, Y., Cherradi, B., Saleh, S., & Raihani, A. (2023). Enhancing Skin Diseases Classification Through Dual Ensemble Learning and Pre-trained CNNs. *International Journal of Advanced Computer Science and Applications*, 14(6).
20. Zhang, T., Wang, Y., Wang, D., Liao, S., Ming, Y., & Zhang, X. (2025, January). Classification of skin cancer from dermatological images by a multimodal method. In *The International Conference Optoelectronic Information and Optical Engineering (OIOE2024)* (Vol. 13513, pp. 214-223). SPIE.
21. Vieira, J., Mendonça, F., & Morgado-Dias, F. (2025). Deep Learning Approaches for Skin Lesion Detection. *Electronics*, 14(14), 2785.
22. Ozdemir, B., & Pacal, I. (2025). A robust deep learning framework for multiclass skin cancer classification. *Scientific Reports*, 15(1), 4938.
23. Atiq, M. E., & Fattah, S. A. (2025). Towards Explainable Skin Cancer Classification: A Dual-Network Attention Model with Lesion Segmentation and Clinical Metadata Fusion. *arXiv preprint arXiv:2510.17773*.
24. Shah, S. A. H., Shah, S. T. H., Khaled, R. A., Buccoliero, A., Shah, S. B. H., Di Terlizzi, A., ... & Deriu, M. A. (2024). Explainable AI-based skin cancer

- detection using CNN, particle swarm optimization and machine learning. *Journal of Imaging*, 10(12), 332.
25. Pavani, M., Ramasubramanian, P., & Venkatasubbu, P. (2025). Hybrid Explainable Deep Learning Framework for Skin Disease Detection and Future Risk Classification Using Sakaguchi Function with CNN–Transformer Models. *Results in Engineering*, 108489.
  26. Ilyosbekov, S. (2025). MelanomaNet: Explainable Deep Learning for Skin Lesion Classification. *arXiv preprint arXiv:2512.09289*.
  27. Zoravar, M., Alijani, S., & Najjaran, H. (2025). Domain Adaptive Skin Lesion Classification via Conformal Ensemble of Vision Transformers. *arXiv preprint arXiv:2505.15997*.
  28. Manzoor, K., Gilal, N. U., Agus, M., & Schneider, J. (2025). Dual-stage segmentation and classification framework for skin lesion analysis using a deep neural network. *Digital Health*, 11, 20552076251351858.
  29. Akram, A., Rashid, J., Jaffar, M. A., Faheem, M., & Amin, R. U. (2023). Segmentation and classification of skin lesions using a hybrid deep learning method in the Internet of Medical Things. *Skin Research and Technology*, 29(11), e13524.
  30. Al Shafi, A., Muntakim, A., Shill, P. C., Zannat, R., & Al-Amin, A. (2025, February). Skin Lesion Classification Using a Soft Voting Ensemble of Convolutional Neural Networks. In *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1-6). IEEE.
  31. Liu, R., Chen, Z., & Zhang, P. (2025). Skin Lesion Classification Based on ResNet-50 Enhanced With Adaptive Spatial Feature Fusion. *arXiv preprint arXiv:2510.03876*.
  32. Vieira, J., Mendonça, F., & Morgado-Dias, F. (2025). Deep Learning Approaches for Skin Lesion Detection. *Electronics*, 14(14), 2785.
  33. Deepa, J., & Madhavan, P. (2025). An advanced skin lesion segmentation and classification framework using deep learning strategies. *Scientific Reports*, 15(1), 33926.
  34. Shoaib, B., Bilal, D. A., Khan, A. H., Kaid, M. A., Darem, A. A., & Sarwar, R. (2025). An explainable hybrid deep learning framework for precise skin lesion segmentation and multi-class classification. *Frontiers in Medicine*, 12, 1681542.
  35. Nie, Y., Sommella, P., Carratù, M., O’Nils, M., & Lundgren, J. (2022). A deep CNN transformer hybrid model for skin lesion classification of dermoscopic images using focal loss. *Diagnostics*, 13(1), 72.
  36. Ali, A., Shahbaz, H., & Damaševičius, R. (2025). xCViT: Improved Vision Transformer Network with Fusion of CNN and Xception for Skin Disease Recognition with Explainable AI. *Computers, Materials & Continua*, 83(1).