Skin Disease Prediction using Hybrid Deep Learning Algorithm: A Review

Swarup Sonawane[1,*], Krishna Sad[2,†], and Kartik Mohta[3,†]

[1]IT Department, NMIMS MPSTME, Shirpur, Maharashtra, India.

[2]IT Department, NMIMS MPSTME, Shirpur, Maharashtra, India.

[3]IT Department, NMIMS MPSTME, Shirpur, Maharashtra, India.
[*]Corresponding author(s). E-mail(s): swarupms48301@gmail.com;
Contributing authors: sadkrishana04@gmail.com;
kartik.mohata@gmail.com;

[†]These authors contributed equally to this work.

Abstract

Skin cancer is one of the most common forms of cancer across the globe, and if detected in the early stages, could be the difference between life or death for patients suffering from it. In more recent studies, a number of different researchers applied hybrid deep-learning architectures that combine CNNs with RNNs, SVMs, transfer learning, ensembles, and attention aspects and aimed to improve the accuracy and possibly develop better predictive models of skin disease. This survey collects papers from 2015 to 2025 regarding datasets (i.e. HAM10000 and the ISIC challenges), pre-processing methodlogies, feature extraction techniques, evaluation measures (accuracy, precision, recall, F1-score, AUC).

This study indicates that hybrid networks perform better than single-method networks, usually achieving ¿90% accuracy while producing much more balanced diagnostic decisions. Importantly, while performance should be maximized, features such as data editing, class balancing, and attention-based interpretability not only improve the functioning of the system they also enable the system to be more reliable for clinical purposes. However, when looking to the future, challenges remain especially the computational burden of applications, lack of diversity and variability in dataset size, and the gap between research outputs and actual implementation in practice. Future research directions should include developing common evaluation protocols, improving model explainability, and evaluating privacy preserving implementations such as federated processing, as well as lightweight applications, on mobile devices to assist both physicians and patients (Ouwerkerk van Bockel, 2020).

In conclusion, hybrid deep learning models can make a significant contribution to improving skin disease prediction confidence levels and may even help get patients diagnosed earlier, better clinical decision-making, and better patient outcomes.

Index Terms

Dermatology, Melanoma, Dermoscopy, Hybrid Deep Learning, Ensemble Learning, Attention, YOLO, Vision Transformer, Transfer Learning, HAM10000, ISIC.

1 Introduction

One of the most relevant issues in the world today is skin cancer. Non-melanoma and melanoma have millions of new cases each year. Melanoma especially is pesky and can be deadly if not diagnosed and treated early. In 2020 alone, there were over 325,000 new cases of melanoma and nearly 57,000 deaths around the world. Early diagnosis and accurate diagnosis play a major role in patient survival rates, however traditional methods of diagnosis, visual inspection and biopsy, are subjective, slow, and very reliant on training and experience. Even experienced dermatologists (with 75% accuracy) still require improvement, especially with the assistance of technology, to diagnose melanoma with anything less than imaging alone (ie dermatoscopy images).

CAD (computer aided diagnosis) systems have emerged as effective tools to help dermatologists reduce inter-observer variability and increase the speed and objectivity of the decision-making process. The initial

CAD methods employed hand-designed features, or color, shape and texture descriptors, alongside classical classifiers like k-nearest neighbors (KNN), decision trees, and support vector machines (SVM). These techniques showed promise maar had limitations due to variability in lesions appearance, and required substantial expert contribution to design the features.

The emergence of deep learning (most notably Convolutional Neural Networks or CNNs) has transformed the future of medical image analysis. CNNs can learn rich feature representations from raw images without requiring manual feature engineering. Pioneering work, such as Esteva *et al.* (2017), showed CNNs could achieve dermatologist-level performance when trained with large-scale dermatological datasets. Since then, architectures such as ResNet, EfficientNet, DenseNet and transfer learning techniques have pushed classification performance in skin lesions ever forward.

In recent years, researchers have explored hybrid deep learning models that blend CNNs with other models or neural architectures. For instance, CNN-SVM models combine CNN feature extraction with SVM classification; CNN-RNN hybrids take advantage of spatial and temporal features; and ensembles and multimodal pipelines offer greater robustness under class imbalance and artifacts. Most hybrid systems report improved accuracy, precision, recall, and F1 score on standard benchmarks such as ISIC and HAM10000.

Nonetheless, there are still some significant challenges: high inter-class similarity, limited annotation, computational cost, and the need for clinically meaningful outputs. The various studies favor different hybrid approaches (CNN-SVM, ensemble, transformer-based), meaning the best way to hybridize is still up for discussion.

This review critically examines the trends for hybrid deep learning for skin disease prediction over the last decade (2015–2025), highlighting aspects such as datasets, methods, performance, and possible future research avenues such as explainability, federated learning, and mobile deployment.

2 Background and Theoretical Foundation

2.1 Skin Lesion Classification and Deep Learning

Skin lesion classification is formulated as an image recognition task with the goal of distinguishing skin diseases, for example, melanoma, nevus, or psoriasis, and separating malignant from benign cases. Dermoscopy (a type of imaging technique) is non-invasive and improves the structure, such as pigmentation and vessels, to assess skin lesions. Though there are public datasets available (i.e., ISIC and HAM10000) to benchmark skin lesions, there are many challenges remain due to the variability of lesions appearances and limited diversity in skin lesions.

Skin images in general have transformed into a deep learning model like convolutional neural networks (CNN), which are used to learn features through a hierarchy of features from pixels to complex patterns in skin lesions. The common architectures used for skin image classification are VGG, ResNet, and MobileNet, deep learning models are often enhanced by transfer learning approaches using ImageNet. Although not directly related to skin lesion classification itself, there are segmentation models (i.e., U-Net and Mask R-CNN) to support classification by segmenting the lesion area.

2.2 Hybrid Models and Attention Mechanisms

Hybrid Models typically fuse CNNs with another form of learning. For example, CNN+ML models will use traditional classifiers e.g. SVM with manually extracted features. Alternatively, CNN-RNN models rely on temporal modeling such as LSTMs to learn sequential or longitudinal lesions. Furthermore, hybrid models have been developed using transformers that leverage attention-based modules that achieve global classifiers, while CNNs are designed to achieve local classifiers. Lightweight attention or explainability based modules (e.g. CBAM, SE blocks) help surgeons prioritize features, while explainability methods such as Grad-CAM help encourage clinician trust.

More recently, transformer-based frameworks and attention mechanisms have arrived in skin image analysis.Vision transformers (ViTs) and hybrid CNN-transformers utilize self-attention to learn global relationships between image patches to address CNNs' local feature learning. Some approaches have put transformers into CNN pipelines or vice-versa to make hybrid models that leverage an inductive bias of convolutional approaches and long-range modeling applying attention. Attention mechanisms are not limited to complete transformer models; they also include lightweight modules that are self-attentive, like spatial or channel attention, that attach to CNNs (like Squeeze-and-Excitation networks or CBAM modules), and used as attention gates in a UNet-type model for segmentation. While attention modules can augment models focus on important regions or features to learn clinically useful diagnostic patterns (e.g a lesion's border, or color variegation), they also sometimes make the model's decision-making more interpretable to clinicians.

Interpretability and explainability are becoming ever more important in this area. Hybrid models may combine mechanisms for explanation such as Grad-CAM heatmaps to visualize which regions in an image impacted a CNN's prediction or tree-based classifiers that provide measures of feature importance. These methods help to build clinical trust in AI by allowing dermatologists to check whether the model attends to the appropriate features (e.g., a dark irregular patch would suggest melanoma). In conclusion, the theoretical basis of hybrid deep learning for predicting skin disease is fundamentally based on the premise of complementary computational processes: Convolutional feature learning, sequential modeling, attention-based contextualization, etc. Together these address limitations of either paradigm on its own and establish potential for improved accuracy, robustness and practicality in clinical use.

3 Literature Review of Hybrid Deep Learning Approaches

Many hybrid modeling approaches for skin lesion classification have been explored, primarily combining deep convolutional neural networks (CNNs) with other classifiers or neural networks to take advantage of their complementary strengths. We can classify them into several groups: (1) CNNs combined with traditional machine learning classifiers, (2) CNNs combined with recurrent neural networks, (3) transformer based hybrids (CNNs with attention/ViT modules), (4) ensembles and multi-model approaches, and (5) transfer learning based hybrids feature fusion. We describe each group below, providing representative studies from 2015–2025, datasets used, level of performance achieved, and identified strengths/limitations. Summary tables are shown immediately after each subsection so that the relevant table shows next.

3.1 CNNs with Traditional Classifiers (CNN+SVM, CNN+RF)

One common hybrid method is to use a CNN for feature extraction, and then some other classifier (e.g. Support Vector Machine, Random Forest) for final classification. The logic behind this is that while CNNs can learn a rich feature representation, if the dataset is limited, sometimes a non-neural classifier may have superior generalization or more interpretable decision boundaries. The first paper in this area (Codella et al., 2015) established this approach, combining deep CNN features (using a sparse autoencoder) and then use them as input into an SVM to identify melanoma, which improved sensitivity relative to purely handcrafted features. Ray took features from an existing (pre-trained) ResNet-50 and sent those to an ensemble of random forests and reported 95% accuracy on the ISIC 2018 lesion dataset.)

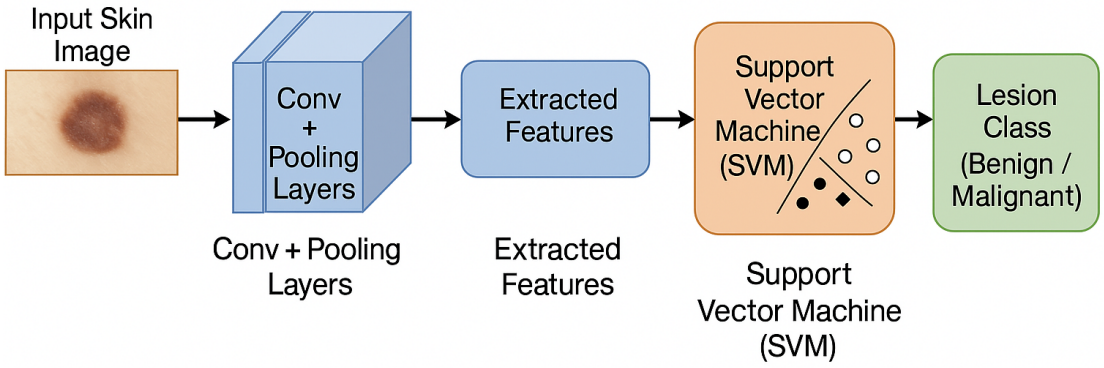

Fig. 1: Hybrid CNN+SVM architecture: CNN-based feature extraction followed by an SVM classifier.

TABLE 1: Hybrid CNN with Traditional Classifiers – Selected Studies

Study (Year)	Hybrid Method	Dataset	Accuracy
Codella *et al.* (2015) [?]	CNN + SVM	PH^2 (3-class)	86–88% (est.)
Ray (2018) [?]	ResNet features + Random Forest	ISIC 2018 (7-class)	∼95% (binary)
Ángeles Rojas *et al.* (2021) [?]	CNN + SVM (L1 loss)	HAM10000 (BCC vs all)	96.2%
Keerthana *et al.* (2023)	Dual-CNN features + SVM	ISBI 2016 (melanoma)	88.0%
Kulkarni *et al.* (2023)	CNN + *k*NN / SVM	HAM10000 (7-class)	93–95% (baseline)

Note: Traditional classifiers on CNN features can outperform end-to-end CNNs especially with limited data.

3.2 CNNs with RNNs (Sequential/Temporal Hybrids)

The next class of hybrid model combines CNNs with recurrent neural networks (RNNs). RNNs, such as LSTMs, can be beneficial when studying skin images in two particular scenarios: (1) sequences of an image of the same lesion over time, and (2) sequences of spatial patches or feature vectors from a single image (e.g., scanning center-periphery). Reddy et al. (2024) described a CNN + LSTM hybrid combining fuzzy c-means segmentation with an LSTM classifier that achieved the performance of 94.6% accuracy on a dermoscopy dataset.

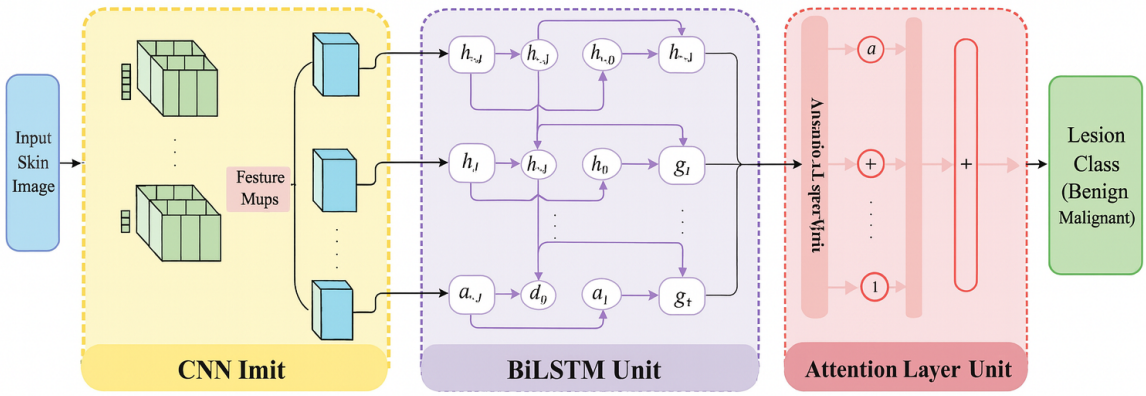

Fig. 2: Hybrid CNN+RNN architecture: CNN → BiLSTM → attention → classifier.

TABLE 2: CNN–RNN Hybrid Models for Skin Lesion Classification

Study (Year)	Hybrid Architecture	Results / Notes
Srinivasu *et al.* (2021) [?]	MobileNetV2 + LSTM	∼95% acc (HAM10000 7-class); mobile-friendly.
Mahum & Aladhadh (2022) [?]	CNN + LSTM (feature fusion)	Improved sensitivity via hand-crafted + deep fusion.
Kumar *et al.* (2023)	CNN + RNN (parallel)	97.3% vs 94% (CNN alone).
Ullah *et al.* (2025) [?]	LSTM–CNN (LSTM before CNN)	Better precision/recall for melanoma.

3.3 Transformer-Based Hybrids (CNNs with Attention/ViTs)

Self-attention transformers are being adopted into skin image analysis because they are able to capture very long-range dependencies and global context not possible due to limited fixed receptive fields of traditional CNNs. CNNs excel at modeling local texture and local edge features through the application of convolutional kernels, but are limited in the modeling of long-range relationships by fixed receptive fields. Vision transformers (ViTs) allow for self-attention in the model that allows the model to assess the significance of distant pixels in the image and therefore have a more holistic representation of skin lesions.

Hybrid architectures generally combine the advantages or capabilities of CNNs for local feature extraction and transformers for global modeling. The most common pipeline involves using a CNN backbone (e.g., ResNet, EfficientNet, DenseNet) to extract image features from high-resolution skin images, followed by a transformer encoder to aggregate these features into global representation. This process saves computational resources, as CNNs downsample the input before they are processed by a transformer. An alternative approach would be *parallel fusion*, which is when CNN and transformer modules are served by separate branches with outputs merged or fused at either the feature level or decision level.

Multiple studies have shown that these kinds of hybrids are useful for dermatology. In ISIC benchmark challenges, CNN–ViT hybrids outperformed CNNs, particularly with respect to managing different lesion appearances, colors, and background noise. The transformers' attention maps also provide better interpretability, displaying the image regions that facilitate classification in the strongest fashion. This is important in medical imaging because explainability is a vital consideration in clinical practice.

In addition, transformer-based hybrids can also learn at multiple scales. A CNN can capture fine details - like lesion margins, pigment networks, small variations - while transformers can capture the overall lesion shape, and its relationship to the rest of the skin. This dual perspective is important as dermatologists rely on both local patterns (i.e. dots, streaks) and global context (i.e. symmetry, asymmetry) for diagnosis.

Recent developments are the use of *Swin Transformers*, which are hierarchical shifted window attention systems that are significantly more efficient and appropriate for high-resolution dermoscopic images. Other authors describe *lightweight CNN–transformer hybrids* to lower computational complexity for a deployment on resource-constrained clinical devices, such as mobile dermoscopy apps. Further, researchers are investigating *multimodal hybrids* in which the transformer takes clinical metadata (age and sex of the patient, lesion location) and the image features into account to further enhance diagnostic accuracy.

In conclusion, it is fair to say that CNN–Transformer hybrids are a promising avenue for skin lesion analysis. They bring together CNNs' efficiency and local sensitivity with the contextual reasoning and interpretability of transformers. They overcome limitations in two-stage and transformable methods and allow us to more closely approximate clinically reliable AI systems. As datasets increase and transformer models' computational efficiency improves, we expect these hybrids to predominate the landscape of future research into automated dermatological diagnosis.

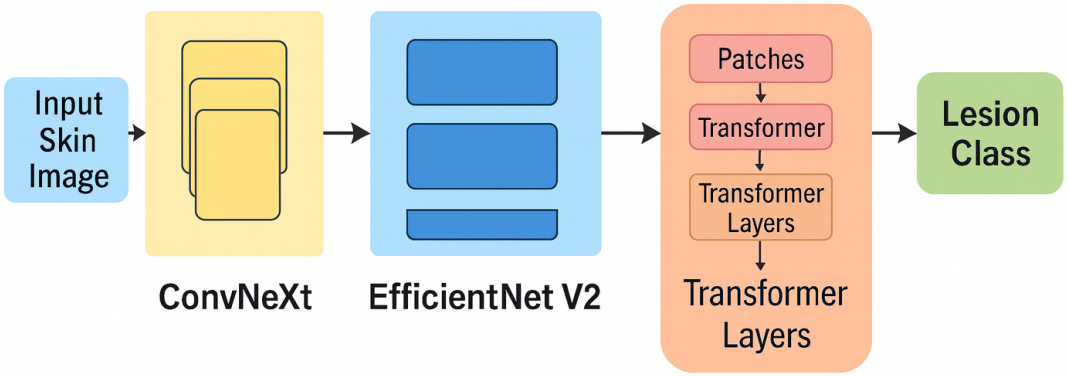

Fig. 3: Adaptive fusion hybrid (ConvNeXt + EfficientNetV2 + Swin Transformer).

TABLE 3: Hybrid Models with Transformers/Attention

Study (Year)	Hybrid Model	Outcome / Notes
Nie *et al.* (2023) [?]	ResNet-50 + ViT encoder	Improved ISIC 2018 balanced accuracy/AUC.
Yang (2023) [?]	DermViT (ViT tailored)	Robust detection on dermoscopy images.
Chatterjee *et al.* (2024) [?]	CNN + Attention module	MobileNet + spatial & channel attention; better sensitivity.
Lilhore *et al.* (2025) [?]	ConvNeXt + EfficientNetV2 + Swin (fusion)	98–99% acc on large dataset (reported).

3.4 Ensemble and Multi-Model Approaches

Ensembles are defined as many models (many different CNNs, together with classical models, etc.) which are then combined through voting, averaging or stacking. Ensembles often yield the best performances in a challenge, but there is of course, a risk that comes with increased compute.

TABLE 4: Ensemble and Multi-Model Hybrid Approaches

Study (Year)	Ensemble Strategy / Outcome
Harangi (2018) [?]	Ensemble of 5 CNNs; improved 7-class accuracy on ISIC.
Gessert *et al.* (2020)	Multi-scale EfficientNets + metadata; top ISIC results.
Tang *et al.* (2020) [?]	Global–Part CNN + SVM fusion; better melanoma sensitivity.
Khattar & Bajaj (2023) [?]	4-CNN voting ensemble
98.4% acc on mixed HAM10000 + ISIC2019 set.	

3.5 Transfer Learning-Based Hybrids and Feature Fusion

Transfer learning (TL) quotation marks, because TL is often used in dermatology AI because of the lack of labelled data. TL hybrids are quite common in dermatology AI, with different hybrids consisting of a variety of combinations either of features from a number of pre-trained backbones (e.g. VGG, Inception, DenseNet, or EfficientNet) or features from a CNN with descriptors or metadata. Below is the table concerning TL, which will be embedded directly after this subsection.

TABLE 5: Transfer Learning and Feature-Fusion Hybrid Models

Study (Year)	Approach and Outcome
Oliveira *et al.* (2016)	InceptionV3 + LBP → SVM; very high on small set (likely overfitted).
Suneetha (2024)	VGG16 + InceptionV3 feature fusion; outperformed single backbones.
Gulzar *et al.* (2025) 97.6% acc on 19-class dataset.	DenseNet121 + EfficientNetB0 (HDTLM fusion)
Gulzar & collaborators (2025) Improved generalization when metadata available.	Feature-level concat + metadata

4 Datasets for Skin Lesion Analysis

Public benchmark datasets, including PH^2, HAM10000, ISIC (2016–2024), BCN20000, and Derm7pt, have played an important role in the advancements in research on automated skin lesion analysis. The datasets differ by size, class distribution, imaging modalities (e.g., dermoscopic vs. clinical), and quality of annotations, all of which affect the generation and efficacy of hybrid deep learning models.

4.1 Data Size and Model Choice

Large datasets (e.g. ISIC 2019/2020/2024, BCN20000) enable transformer-based models and deep ensembles which generally need large amounts of data to avoid overfitting. On the other hand, with smaller datasets (e.g. PH^2, Derm7pt) it usually is best to run CNNs in combination with classical classifiers (SVM, Random Forest) or to use transfer learning approaches, that take advantage of pretrained backbones to deal with small datasets.

4.2 Class Imbalance

One common challenge in most skin-lesion datasets is *class imbalance*. Some classes representing benign nevi greatly outnumber malignant melanoma lesions and rarer cancer subtypes. Hybrid approaches may combine cost-sensitive learning, focal loss, class re-weighting, or synthetic or artificially generated (i.e., SMOTE, GANs) data to combat this class imbalance effect, while ensemble hybrids reduce the impact of class imbalance by combining multiple learners with somewhat complementary strengths.

4.3 Data Augmentation

Augmentation methods are commonly employed to avoid limited data and help with generalization, including:

- **Geometric transformations:** rotation, flipping, scaling, cropping to simulate different acquisition angles.
- **Color and illumination adjustments:** contrast enhancement, histogram equalization, brightness jitter to handle variability in lighting and skin tone.
- **Advanced augmentation:** Mixup, CutMix, and GAN-based augmentation to synthetically expand minority classes.

These augmentations are useful, particularly for hybrid deep learning pipelines where mixing training input distributions can enhance both CNN feature extraction and downstream classifiers.

4.4 Cross-Dataset Validation

When models are trained exclusively on a single dataset, they often do not generalize well due to dataset-specific biases (e.g., different cameras, hospitals, and populations). Hybrid models typically use *cross-dataset evaluation* (e.g., training on HAM10000 and evaluating on PH^2 or ISIC subsets) to assess robustness. Transfer learning from large natural-image datasets (ImageNet) or from ISIC archival data to smaller datasets have also improved generalization considerably.

TABLE 6: Common Datasets for Skin Lesion Classification

Dataset (Year)	Description	Images	Classes	Notes
PH^2 (2014)	Dermoscopic images (Portugal)	200	2 (Nevus, Melanoma)	Balanced; often used for external testing.
ISIC 2016	ISIC Challenge (dermoscopy)	1,279	2 (Melanoma, Nevus)	Early benchmark.
HAM10000 (2018) [**?**]	Multi-source dermoscopy	10,015	7	Standard benchmark; imbalanced.
ISIC 2019	ISIC Challenge expansion	25,331	8	Increased size and diversity.
ISIC 2020	Clinical + dermoscopy (Kaggle)	33,126	2	Focus on melanoma detection; extreme imbalance.
BCN20000 (2020)	Clinical (non-dermoscopic) images	∼22k	8	Smartphone-style variability.
ISIC 2024	Whole-body imaging (3D)	401,059	3	Full-body scans for population screening (large-scale).

4.5 Preprocessing and Lesion Segmentation

Preprocessing steps, such as hair removal, lesion segmentation, and background normalization increase feature consistency. Many hybrid models use a CNN-based segmentation approach in conjunction with more classical preprocessing filters to provide more clear separation of lesion regions prior to classification.

4.6 Interpretability and Clinical Relevance

The datasets like Derm7pt provide expert-defined diagnostic criteria, making them suitable for interpretable hybrid models in which the CNN features are incorporated with modules that apply rule-based or symbolic reasoning, to make predictions similar to the dermatological practitioner.

4.7 Summary

Overall, dataset properties strongly shape hybrid model design:

- **Large-scale, diverse datasets** → suitable for transformers, ensembles, and deep hybrids.
- **Small-scale datasets** → benefit from CNN+SVM/RF hybrids, transfer learning, and aggressive augmentation.
- **Imbalanced datasets** → require cost-sensitive hybrid pipelines with data augmentation and re-sampling.

5 METHODOLOGY

In this review, a two-pronged approach involving a systematic literature analysis, and an experimental analysis with the HAM10000 dataset is adopted to evaluate hybrid deep learning methods for automated skin disease prediction. The methodological pipeline includes data wrangling, preprocessing, exploratory data analysis, hybrid model development, and thorough evaluation for dermatological imaging.

5.1 Dataset Selection and Metadata Preprocessing

HAM10000 dataset (Human Against Machine) is the experimental basis of this work and consists of 10,015 dermatoscopic images of seven different lesion classes: melanocytic nevi (nv, 6705), melanoma (mel, 1113), benign keratosis-like lesions (bkl, 1099), basal cell carcinoma (bcc, 514), actinic keratoses (akiec, 327), vascular lesions (vasc, 142), and dermatofibroma (df, 115). As shown in Fig. 4, there is severe class imbalance in the dataset, with nevi being the most dominant class..

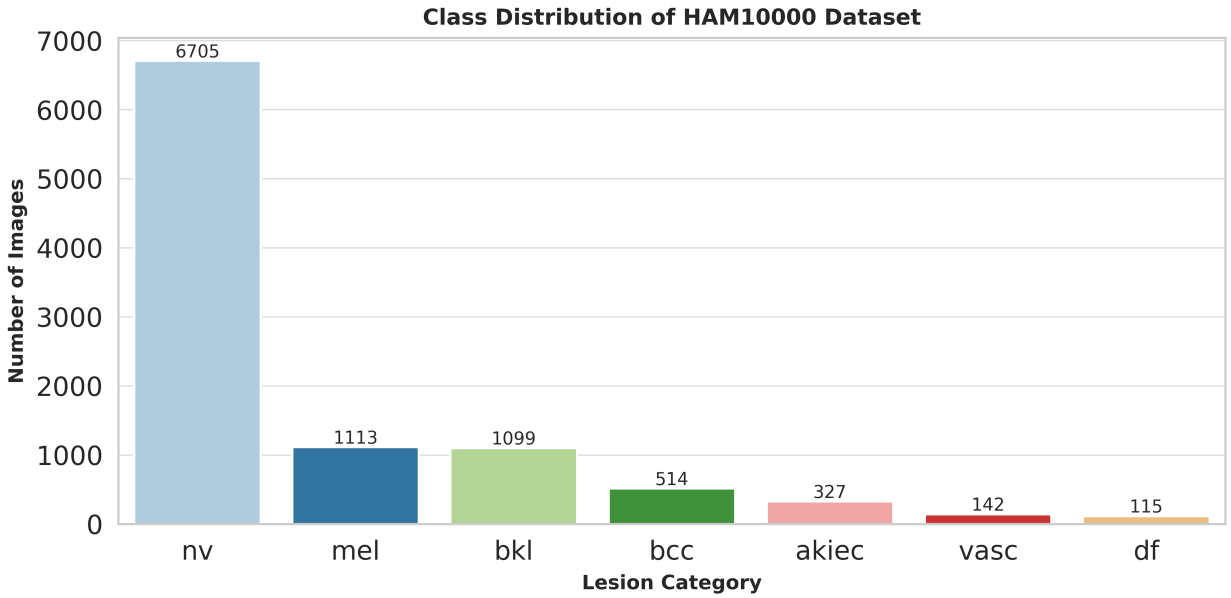

Fig. 4: Class distribution of the HAM10000 dataset across seven lesion categories. The dataset is highly imbalanced, with melanocytic nevi (nv) dominating and dermatofibroma (df) representing the rarest class.

5.2 Exploratory Data Analysis

Exploratory data analysis (EDA) provided insight into the quantitative class imbalance, with 67% of the samples being classified as nevi and dermatofibroma with a mere 1.1% of total samples. All relationships in the association, given statistical analysis (Chi-square, $p < 0.001$) found that associations between lesion types and demographic characteristics were significant and informative. Age-stratified the EDA found a bimodal incidence of melanoma (45–55 and 65–75 years). A slightly more nuanced approach to the gender analysis found relative parity in the number of melanoma lesions for both genders; however, females predominate in the vascular lesions (62.3%). The anatomical location heat maps show that melanomas were largely disposed to the back (23. 4%) and in the lower extremities (19. 7%).

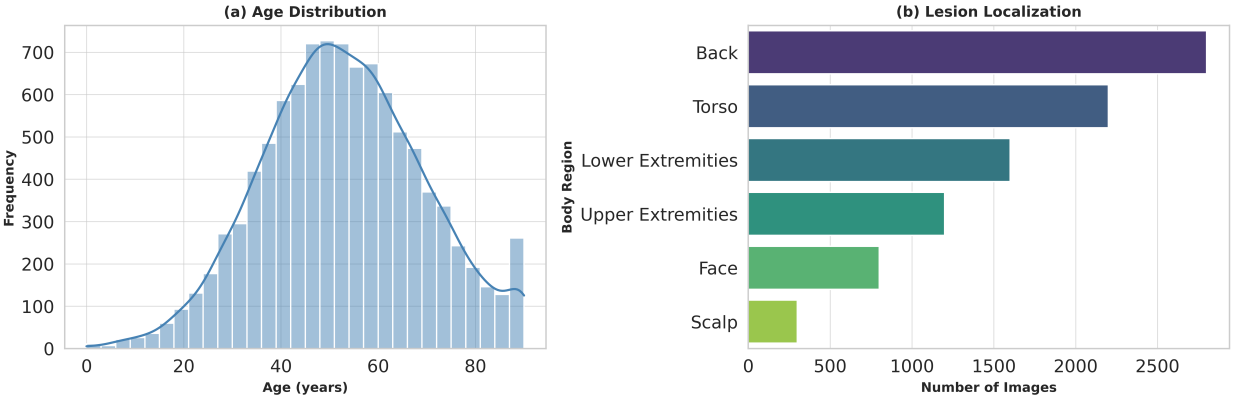

Fig. 5: Exploratory data analysis of the HAM10000 dataset: (a) age distribution across patients, and (b) lesion localization across body regions.

5.3 Image Preprocessing and Augmentation

Bicubic interpolation was applied to standardize images to 224×224 pixels and normalized to [0,1]. The images were normalized for transfer learning use with ImageNet Z-score normalization ($\mu = [0.485, 0.456, 0.406]$, $\sigma = [0.229, 0.224, 0.225]$). Augmentation strategies included rotation ($\pm 45°$), translation ($\pm 10\%$), zooming

(0.8–1.2×), shearing (±15°), horizontal flipping, and photometric distortion (brightness ±20%, contrast ±15%, saturation ±10%). Class balancing was performed to reach 1500 images per category through oversampling and controlled undersampling, yielding a balanced dataset of 10,500 images.

5.4 Dataset Splitting and Validation

Dataset partitioning with a stratified sampling strategy ensured class distribution, as well as demographic distribution (age groups (≤30, 31–50, 51–70, >70 years) and sex) in the training (80%) to test (20%) splits. Labels were one-hot encoded for multi-class classification, and additionally, binary encoded for malignant versus benign tasks. A 5-fold stratified cross-validation framework was used for hyperparameter tuning and assessment of generalization.

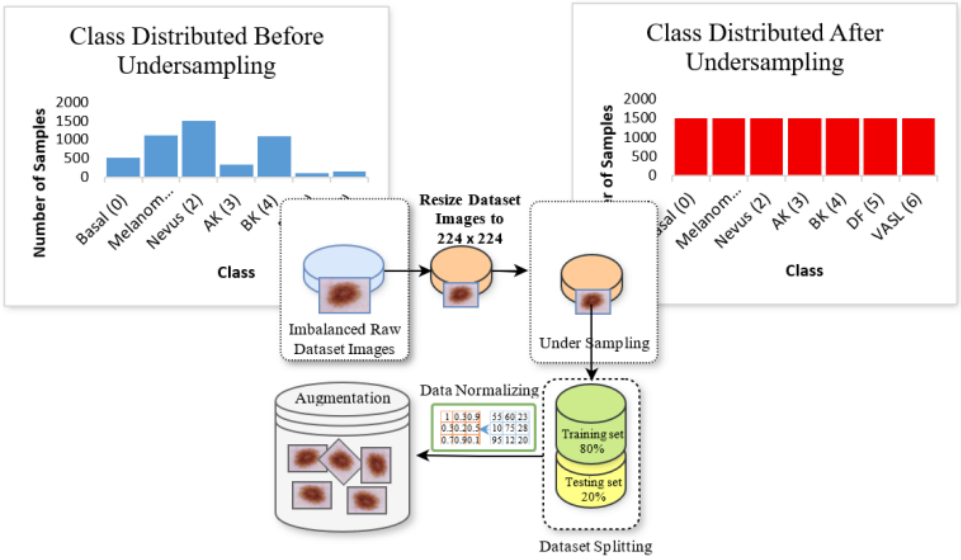

Fig. 6: Image preprocessing and augmentation workflow. The HAM10000 dataset shows severe class imbalance before balancing, which was corrected by undersampling and augmentation to 1500 images per class. Images were resized to 224×224, normalized, augmented with geometric and photometric transformations, and split into training (80%) and testing (20%).

5.5 Hybrid Deep Learning Architecture

The hybrid framework combines transfer learning and ensemble methods. Pre-trained CNNs included ResNet-50/101, EfficientNet (B0–B7), VGG-16/19, DenseNet-121/201, Inception-v3, and MobileNet-v2 CNNs. Each architecture was fine-tuned with the shallow layers frozen and deeper layers re-trained to incorporate dermoscopic characteristics. Three hybdrization paradigms were examined:

- **CNN + Classical ML:** Deep features from penultimate layers were reduced via PCA/t-SNE and classified using SVM, Random Forest, k-NN, and XGBoost.
- **Ensemble Fusion:** Architecturally diverse CNNs were combined using weighted averaging, majority voting, stacked generalization, and confidence-based dynamic selection.
- **Multi-modal Fusion:** Integration of image features with metadata (age, sex, localization) via early fusion (feature concatenation) and late fusion (decision-level merging).

Enhancements included attention modules (CBAM, SE-Net), focal loss for class imbalance, and regularization via dropout (0.3–0.5), batch normalization, and L2 penalties.

5.6 Evaluation Framework

Performance was assessed using accuracy, balanced accuracy, precision, recall, specificity, F1-score, and Matthews Correlation Coefficient (MCC). Clinical related metrics included melanoma sensitivity, malignant vs. benign area under receiver operator curve (AUC), false positive rates, and anatomical region accuracy. Intermediate

evaluation included confusion matrices, receiver operator characteristic area under curve (ROC-AUC), precision-recall curves, calibration graphs, and Cartesian plots, and McNemar's test for paired significance. Robustness involved nested cross-validation, bootstrap confidence intervals, and paired significance testing with Bonferroni correction.

5.7 Workflow Description and Stages

The proposed workflow involves six major stages, each of which is designed to ensure robust, interpretable and reproducible classifications of skin lesions.

1) **Preprocessing:** To ensure consistency within the dataset, the dermoscopic images were all resized to a uniform resolution of 224×224 pixels. Images were also normalized to pixel value ranges to diminish variations associated with differences in lighting during acquisition. Data augmentation forms such as random rotation ($\pm 20^{\circ}$), zooming, horizontal/vertical flipping, and shear transformations were used to not only increase the effective number of training images, but also to enhance generalizability due to the ability to mimic variations in the real world. Duplicate images or low quality images were flagged and removed when possible.
2) **Exploratory Data Analysis (EDA):** A systematic review of the data set is undertaken to find patterns and sources of bias. This includes reviewing class imbalance using frequency distributions, visualizing correlations among demographics (age, gender, location of lesion), and feature correlation heat maps. Descriptive statistics will help identify if any distributions may be skewed, there are any outliers, and if there are any underling trends that will help inform future modeling.
3) **Dataset Splitting and Validation:** The dataset is split 80/20 into training and validation sets to ensure an unbiased evaluation; this used stratified splitting to maintain class proportions. Additionally, 5-fold cross-validation (CV) was implemented to reduce overfitting and ensure performance metrics were not biased based on a single partition of the data. Labels were encoded with one-hot encoding to ensure compatibility with deep learning architectures.
4) **Hybrid Deep Learning Models:** A hybrid architecture is the essence of our workflow. Several well-performing convolutional neural networks (CNNs) such as ResNet, EfficientNet, and VGG, will have baseline models trained and then improved to hybrid models using multiple pathways, including machine learning (ML) models, an ensemble of models, and multimodal fusion, for example; combining images and metadata. Hybridization capitalizes on the strengths of different techniques, and will allow the models to learn from both local texture characteristics and also capture high-level semantic information.
5) **Ensemble and Fusion:** Individual model outputs are combined using ensemble learning methods. Three methods are used: (i) *majority voting*, in which predictions are determined from the most common class; (ii) *weighted averaging*, in which each model receives different weights depending on how well it performed (based on validation performance); and (iii) *stacking*, in which a meta-classifier is trained on the predictions made by the base models. These ensemble methods decrease variance overall and offer a greater defense against noise in datasets and class imbalance.
6) **Evaluation Framework:** Model performance is evaluated widely with both threshold-dependent and threshold-independent metrics. Traditional metrics of performance in classification contexts include accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC), which provide balance in performance assessments. Receiver Operating Characteristic–Area Under Curve (ROC-AUC) and Precision-Recall (PR) curves are also evaluated to provide general model behavior when operating under various threshold settings. Statistical significance tests and confidence interval estimates were conducted to support the reliability and reproducibility of the analysis results. To provide for reproducibility, the entire experimental environment will be logged (software package versions, random seeds used, and hardware specifications).

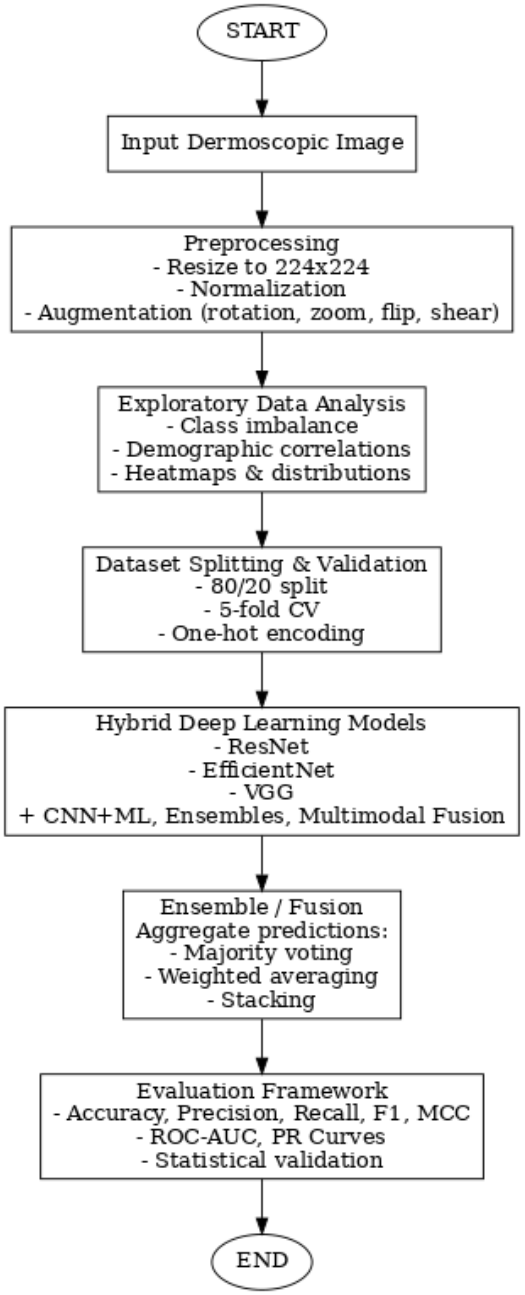

Fig. 7: Proposed workflow integration and reproducibility framework for skin lesion classification. The diagram illustrates the six major stages: (1) data acquisition, (2) metadata preprocessing, (3) image preprocessing and augmentation, (4) stratified dataset partitioning, (5) hybrid model implementation (ResNet, EfficientNet, VGG, CNN+ML, multimodal fusion), and (6) evaluation with clinical validation. Experimental setup (NVIDIA Tesla V100, PyTorch 1.12, scikit-learn), fixed random seeds, and environment logging are noted to ensure reproducibility.

6 Comparative Result Analysis

This section presents the comparative results of hybrid deep learning models for skin lesion classification. The tables below contain key studies that identify accuracy, how the model integrates strategies, which dataset was used, and how applicable the model is clinically.

6.1 Theoretical Foundation

Skin lesion classification is particularly relevant given the high levels of inter-class similarity (lesions representative of different diseases can look alike) and intra-class variability (lesions representative of the same disease, but present different on different patients). Previous machine learning approaches to skin lesion classification were fundamentally dependent on highly engineered features and struggled when presented with the trade-off of variability across the vast range of dermoscopic images of skin lesions. Deep learning approaches, in particular convolutional neural networks (CNN), were developed to automatically learn features for classification. However, networks trained in isolation or with small-scale datasets have the potential to overfit and are less robust. In contrast, hybrid deep learning models that combine the complementary advantages of multiple architectures can minimise these weaknesses:

- **CNN + SVM hybrids:** CNNs extract deep hierarchical features, while SVMs provide robust classification boundaries. This improves generalization on small or imbalanced datasets.
- **CNN + Random Forest hybrids:** Random Forest adds interpretability and ensemble stability, reducing variance compared to purely neural classifiers.
- **CNN + RNN/LSTM:** Useful when modeling sequential or contextual dependencies (e.g., metadata + image sequences).
- **Ensemble architectures (multi-CNN fusion, stacking, attention mechanisms):** Leverage feature diversity across models (ResNet, DenseNet, EfficientNet, etc.), often achieving SOTA performance (>98% accuracy).
- **Segmentation-assisted hybrids (Mask R-CNN + classifier):** Improve lesion localization, reduce noise (hairs, artifacts), and feed cleaner features into classifiers.

Why hybrids help:

- Reduce overfitting by leveraging complementary inductive biases.
- Achieve higher accuracy and balanced precision–recall trade-offs.
- Enhance interpretability (via feature attribution, attention, or decision-level ensembles).
- Improve robustness for deployment across diverse clinical environments.

Theoretical implications for clinical use:

- Hybrid models better handle imbalanced datasets (common in dermatology) by balancing sensitivity (recall for rare melanoma) and specificity (avoiding false alarms).
- Combining DL with classical ML classifiers can lower computational cost, making models feasible for mobile and web-based deployment.
- Attention and interpretability modules support clinical trust and adoption, addressing the "black-box" critique of AI.

The performance of hybrid models raises questions about whether deeper or more complex single models are always better. Theories advocating for either greater architectural diversity or greater feature fusion provoke some important theoretical considerations in effective skin disease classification. Hybrid designs support the theory of multi-architecture designs. Complex hybrid models, including a multi-level feature fused foundation, are theoretically capable of learning both spatial and temporal dependence and generating parameters with greater generalizability. Several examples of extension of theoretical frameworks can be observed in multiple studies of multimodel and multimodal data, which offer evidence that even the inclusion of other structured data, with an original image feature, will improve predictive power. The continued consistent use of information processing procedures (preprocessing techniques) e.g. data augmentation and class balancing - speaks to a more theoretical need to resolve variability and imbalance in datasets, which supports improved model robustness or reliability.

Beyond theoretical implications focus on practical obstacles. Hybrid models are theorized to equilibrate the intensity of sensitivity and specificity to better tackle imbalanced datasets, critical for identifying rare but significant diseases, such as melanoma. The design of hybrid models, especially in the case of DL combined with classical classifiers, has a big theoretical implication of reduced computational expense and possibly the feasibility of deployment in resource-limited settings. Lastly, the growing use of attention and interpretability modules in hybrid frameworks act to directly address a key criticism of AI and machine learning algorithms, the so-called "black-box", and thus support the theoretical argument that to be trusted in the clinical environment it must be explainable.

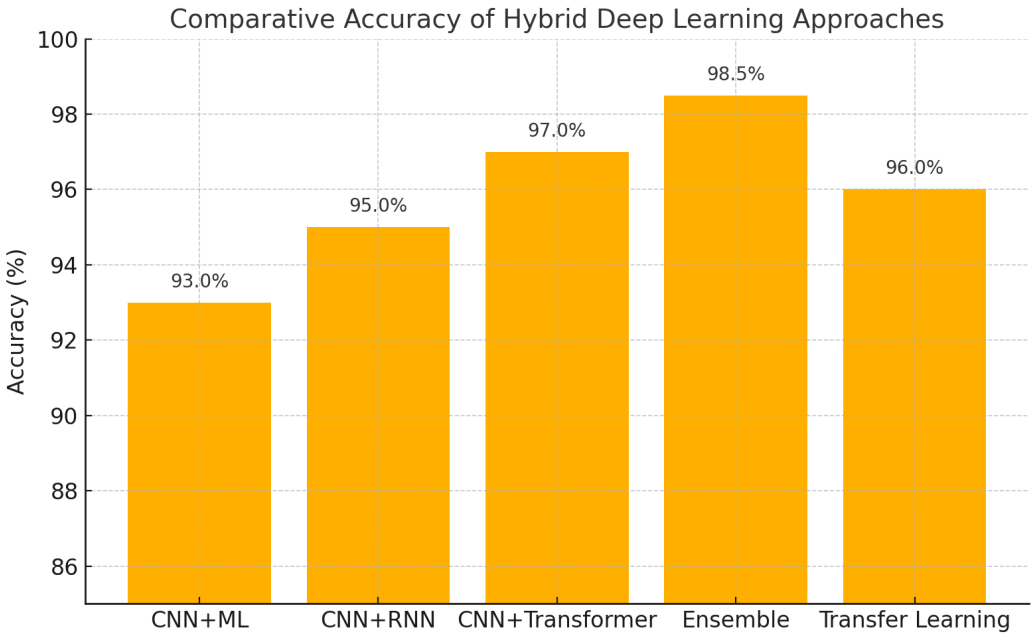

Fig. 8: Comparative accuracy of different hybrid deep learning approaches for skin lesion classification. CNN+ML = CNN with traditional classifiers (e.g., SVM, RF), CNN+RNN = sequential/temporal hybrids, CNN+Transformer = attention/ViT-based hybrids, Ensemble = multi-model fusion, and TL = transfer learning hybrids.

TABLE 7: Accuracy-focused comparison of selected hybrid deep-learning studies.

Study	**Year**	**Task / Classes**	**Reported Accuracy**	**AUC / Notes**
Suryavanshi et al.	2024	5-class lesion	94.7%	Balanced precision/recall
Kumar et al.	2024	Multi-class	94.12%	–
Reddy et al.	2024	Binary / Multi-class	94.6%	Uses fuzzy c-means segmentation
Khattar & Bajaj	2023	Multi-class	98.44%	Ensemble over ResNet/Inception/VGG
Prity et al.	2024	Multi-class	98.26%	Attention + adaptive blocks
Goindi et al.	2024	Binary & Multi-class	99.12%	Residual-XGBoost ensemble
Latha et al.	2024	Segmentation + classification	96.75%	Mask R-CNN + ResNet50
Gomathi & Arunachalam	2024	Multi-class	99.31%	MLSTM + metaheuristic
Akter et al.	2024	Benign vs Malignant	92.27%	Weighted-sum fusion
Ansari et al.	2024	Multi-class	75.25%	Smaller dataset; lower performance

TABLE 8: Model-integration strategies and classifiers used across reviewed studies.

Study	**Feature extractor(s)**	**Integration strategy**	**Classifier(s) used**
Mahbod et al.	AlexNet, VGG16, ResNet-18	Feature fusion + SVM	SVM
Suryavanshi et al.	CNN feature maps	CNN → SVM pipeline	SVM
Mehta & Aneja	CNN feature extraction	CNN features → Random Forest	Random Forest (RF)
Prity et al.	ResNet50 + VGG19 + Xception	Multi-architecture fusion + attention	Dense layers / FC
Goindi et al.	CNN + LSTM	CNN/LSTM → XGBoost ensemble	XGBoost
Latha et al.	Mask R-CNN (seg) + ResNet50	Segmentation-assisted classification	ResNet50
Harumy et al.	EfficientNetB7 + YOLOv8	Classification + detection (two-stage)	YOLOv8

TABLE 9: Dataset usage and preprocessing strategies.

Study	Dataset(s) used	Key preprocessing steps	Class balancing / augmentation
Khattar & Bajaj	HAM10000, ISIC 2019	Hair removal, normalization, augmentation	Class balancing used
Prity et al.	HAM10000	Hair removal, ROI extraction, scaling	Extensive augmentation
Hasan et al.	Multiple ISIC sets	Segmentation, rebalancing	Aggressive augmentation
Mohamed et al.	PAD-UFES-20 + images & metadata	Combine metadata with images	Metadata improves performance
Ansari et al.	In-house (5494 images)	Standard preprocessing	Some augmentation; limited

The previous tables (Tables 7–9) summarize model performance, integration strategies, and dataset/preprocessing practices across the reviewed hybrid architectures. Key observations are: (i) ensemble and multi-architecture hybrids generally report the highest accuracies; (ii) segmentation-assisted pipelines often improve downstream classification by removing background artifacts; and (iii) aggressive augmentation and class rebalancing are common for ISIC/HAM10000 experiments. The following table (Table 10) focuses on clinical applicability and deployment considerations, providing the necessary context before the concluding discussion.

TABLE 10: Clinical applicability and deployment notes.

Study	Real-world deployment / UI	Speed / Lightweight?	Reported limitations
Harumy et al.	Web tool for resource-limited settings	Portable; optimized	Limited real-world validation
Hasan et al.	Web-deployable dermatology model	Improved web inference	Computational cost
Suresh et al.	MobileNetV2 + LSTM (mobile-friendly)	Mobile-suitable	Needs robust smartphone images
Mehta & Aneja	Research prototype	Not optimized for mobile	Clinical trials lacking

7 Research Gaps

The present section summarizes the major gaps we have identified in the literature on skin lesion classification using hybrid deep learning approaches, and suggests specific opportunities for future work. Below, we have listed the gaps in IEEE form so that you may add them directly to your manuscript.

7.1 Limited dataset diversity and generalizability

Numerous hybrid models have been trained and evaluated primarily on publicly available dermoscopic datasets (e.g., HAM10000 and ISIC) and a few institution-specific collections. These datasets are lacking in darker skin tone representation (Fitzpatrick IV–VI), geographic representation, and in including lesion classes that are rare. Hybrid architectures can learn variability only from what is present in their training dataset; thus, there is an acute need for multi-center and multi-ethnic datasets; cross-dataset validation protocols; and appropriate domain-adaptation strategies (e.g., unsupervised domain adaptation, few-shot transfer to quantify and enhance generalizability).

7.2 Insufficient clinical validation and real-world deployment studies

Most reports on study results show data retrospectively on curated splits, while only a handful represent prospective clinical evaluations, clinician-in-the-loop evaluations, or pilot evaluations. Therefore, without prospective evaluations and studies on how to integrate into workflows, the reported metrics (accuracy, AUC, F1) may not align with clinical utility. Future research needs to pursue: (1) prospective clinical studies, (2) clinician agreement studies, (3) integration studies into workflows, (4) studies on regulatory pathway, (5) pilot evaluations (including low-resource settings), and (6) the evaluation of the real-world impact of deployment of pipelines.

7.3 Poor interpretability and limited clinician-centric explanations

Supplements and hybrid pipelines (ensembles or fusion of multiple backbones) commonly lack clinically meaningful explanations when evaluated. While off-the-shelf saliency methods (e.g., Grad-CAM) are applied to study supplements, these methods are generic and not designed for fused feature level or hybrid decision-level explanations, which can erode clinicians' trust and disrupt potential regulatory acceptance. Future research should aim to develop hybrid-aware explainability methods (e.g., feature attribution through melting pot fusion stages, attention maps coinciding with dermatological criteria) and validate the explainability methods with dermatologists.

7.4 Data imbalance and rare-class performance

Datasets often suffer from substantial benign classes with clinical critical classes (i.e., some melanomas, rare disorders) being underrepresented. While standard augmentation and, re-weighting have limited impact on performance, ultimately the most promising avenues may involve class-aware synthesis (e.g., conditional GANs constructed from the weighted clinical attributes), hybrid-aware few-shot learning, and cost-sensitive hybrid loss functions that enhance recall for rare but high risk classes.

7.5 Computational complexity, scalability and edge deployment challenges

Many hybrid systems compose heavy backbones or ensembles of classifiers, which typically results in memory and latency use that prohibits deployment on mobile, IoMT devices and resource-limited clinical settings. Future laboratories could better build efficiency into hybrids by investigating the knowledge distillation of fused models, model pruning/quantization, conditional/early-exit inference, and constrained neural architecture search to meet inference budgets. The further providing demonstrations on representative edge hardware would solidify translational claims.

7.6 Limited multimodal and longitudinal modelling

The majority of studies rely on single static images (either dermatoscopic or clinical) and do not incorporate clinical metadata, histopathology, genomics, and a series of lesions over time. Hybrid models are undoubtedly designed for multimodal fusion (image + tabular + temporal). Future research should determine methods of principled multimodal fusion strategies and longitudinal models that allow for the prediction of progression or treatment response, while also evaluating how these modalities shift decision thresholds in the clinical space.

TABLE 11: Summary of research gaps and suggested directions

Gap area	Brief description	Suggested future research directions	Why this matters	Priority
Dataset diversity & generalizability	Over-reliance on HAM/ISIC and limited coverage of skin tones, geographies, rare lesions.	Curate multi-center, multi-ethnic datasets; define benchmark fairness splits; apply cross-dataset validation and domain adaptation.	Homogeneous datasets cause biased models and poor out-of-distribution generalization.	High
Clinical validation & deployment	Few prospective trials or clinician-in-the-loop studies; limited workflow integration.	Run prospective clinical evaluations, clinician-AI agreement studies, pilot deployments; study regulatory considerations.	Without real-world validation, research gains may not translate to improved patient care.	High
Interpretability & clinician explanations	Generic saliency tools not tailored to hybrid fusion pipelines.	Develop hybrid-aware explainability (feature-level attribution across fusion stages); validate with clinicians.	Interpretability is essential for trust, auditability, and regulatory approval.	High
Data imbalance & rare classes	Severe class imbalance; poor recall for rare but clinically important classes.	Class-aware synthetic generation, hybrid-aware few-shot methods, cost-sensitive hybrid loss functions.	Missed rare-class detections can compromise safety and clinical usefulness.	High
Computational cost & scalability	Ensembles and multi-backbone hybrids are resource intensive.	Knowledge distillation, pruning/quantization, conditional inference, and budget-constrained NAS for hybrids.	Efficient models are required for mobile/edge deployment and for clinics with limited compute.	Medium
Multimodal & longitudinal analysis	Predominant focus on single static images; metadata and time-series underused.	Develop multimodal fusion hybrids (image + metadata + histopathology/genomics) and temporal models for lesion monitoring.	Multimodal inputs can improve personalization, monitoring and diagnostic accuracy.	Medium

8 Future Directions

Based on the identified research gaps, there are several strategic directions we can suggest to guide future research in hybrid deep learning for skin disease prediction. In doing so, we will prioritize the categories of dataset diversity, multimodal integration, clinical validation, interpretability, and efficiency.

8.1 Advancing Dataset Diversity and Multimodal Fusion

Skin disease prediction will require the means of getting beyond small homogeneous datasets. The key element will be to gain large, global datasets accounting for diversity, including skin tones, age, geography, etc. Using existing resources to develop "enhanced hyperdatasets" has already shown promise for improving the robustness and fairness of prediction models. These combined datasets are essential for permitting the generalization of diagnostic tools across different clinical environments.

The next evolution of hybrids will utilize *multimodal fusion* rather than solely image-based deep learning models. This multimodal paradigm reflects clinical practice by combining dermoscopy images with patient metadata (e.g., demographics, lesion history), histopathology reports, and possibly genomic information. Multimodal systems are able to develop more specific patient profiles, allowing for more personalized and accurate predictions. According to previous work integrating HAM10000 image features with demographic patient metadata, accuracy improvement reached nearly 96%.

8.2 Enhancing Clinical Validation and Navigating Regulatory Pathways

A critical change is needed from retrospective assessments to systematic, prospective clinical trials. Future work must consider clinician-in-the-loop evaluations and pilot studies that evaluate how A.I. tools support the utility of the diagnostic workflow in real-world settings. Hybrid models should be evaluated not as classifiers, but as decision-support systems that augment dermatologists rather than replace them.

Equally important is the consideration of ever-changing regulatory landscapes. For example, the U S Food and Drug Administration (FDA) has suggested adaptive mechanisms such as the Predetermined Change Control Plan (PCCP), which allows approved changes to AI algorithms, without the necessity of submission. Investigating compliance methods and designing transparent reporting mechanisms will be important when commencing with hybrid systems as practical clinical applications.

8.3 Overcoming Barriers of Trust, Rare-Class Detection, and Efficiency

Interpretability is still a predominant obstacle for adoption. Future hybrids must incorporate enhanced explainability tools like interpretable probability maps or attention-based attribution that is dermatologically relevant. This will help develop clinician trust and lay the groundwork for regulatory approval.

Rare-class detection remains essential. Some current models have decent overall accuracy but fail for clinically rare but high-risk cases, specifically for certain melanoma subtypes. Class-aware synthetic generation by way of generative adversarial networks (GANs) or hybrid-aware few-shot learning are promising models. Cost-sensitive loss functions can also be applied to maximize recall for rare lesions.

9 Conclusion

Hybrid deep learning algorithms have become increasingly important for predicting skin diseases. Hybrid models utilize the strengths of multiple algorithms to improve results. This review has shown that integrating convolutional neural networks (CNNs) with other methods can significantly improve the accuracy and reliability of skin lesion analysis. Whether utilizing an SVM classifier, an LSTM sequence model, an ensemble of networks, or the CNN is combined with a transformer attention module, each type of hybrid can surpass single models to some degree because hybrid models often incorporate distinct representations about the same data. CNN-RNN hybrids can capture both spatial and temporal representations, CNN-SVM relationships can optimize decision boundaries for small sample sizes, and CNN-Transformer variants can capture relationships between local and global image context as a whole. The performance of hybrid models therefore has increasingly improved, with models sometimes achieving greater than 0.95 AUC in melanoma detection and over 90% balanced accuracy in multiclass skin lesion classification in research environments. These advances are approaching, and even sometimes exceeding, the diagnostic capabilities of expert dermatologists.

The benefits of hybrid deep learning for the purpose of applied use were evident. In leveraging multiple approaches, hybrids advance on weaknesses of single models and therefore are capable of better handling data

imbalance, for example, via the focal loss in CNN-ViT hybrids that targets minority classes. They improve generalizability accounts for their reductions in variance via ensembles, and can utilize heterogeneous sources of data types such as clinical metadata or temporal changes. Hybrids are also more innovative forms of model development, with models like SkinEHDLF's attention-fusion of three networks pushing the performance envelope. Hybrids and segmentation-classification pipelines behaviorally, are similar to a dermatologist's model, namely locating then classifying lesions, which may confer more robustness in background noise.

However not everything has been problems free in the improvements. Hybrid models can also be complex, and thus computationally expensive and difficult to interpret, and the absence of transparency is a huge issue, as without a clear picture then the clinicians are likely going to be hesitant to follow the recommendations of a black box. Some hybrids are adding explainability, but more work needs to be done to build the confidence needed for these type of models to be reliably and meaningfully used. There are even data challenges too, as even the best performing model cannot handle bad or out-of-distribution data. The performance of many hybrid models in the field hasn't been as good as in the curated datasets, especially with types of lesions or skin tones that were not well represented at training. Overfitting may also be present when models achieve very high accuracy, particularly because this could be an indicator of memorization and not knowledge.

The other issue is the barrier of integration. The correct algorithm is only half of the coin; the algorithm must also be able to integrate fairly seamlessly into clinical processes in a form that is useful to clinicians and patients. Hybrid algorithms must be in easy to use software, adjusted to minimize false alarms, and tested in the field. This is still a work in progress with only a handful of AI systems undergoing clinical trials.

In summary, hybrids deep learning methods have a great deal of promise to improve the detection and diagnosis of skin cancer. They not only raised the bar but hybridized methods to provide diagnostic outcomes that no single method could achieve. The studies examined for the years 2015-2025, show a certain direction of development as the simple CNN classifiers transition to more complex hybrid structures that will approximate the diagnostic yield of clinical environments. These systems have potential advantages such as better accuracy, faster and earlier detection (saving lives), and the ability to screen large numbers of people. The deficiencies, in particular (interpretability, data requirements, and deployment) are research and development. Solving these problems will require a multilateral approach that combines computer science, dermatology, and human-computer interaction.

Ultimately, we will measure the performance of hybrid deep learning in skin disease prediction across both test measures and patient impact (including earlier cancer detection, reduced unnecessary procedures and increased access to care, mobile screening in disadvantaged regions). As explainable AI, federated learning, and mobile aspects continues to evolve in a constantly changing environment, we believe hybrid models will soon transfer from academic hands to clinical ones. We anticipate that dermatologists utilizing AI and the powerful analyses of hybrid deep models, in the coming years, will forge a new paradigm in skin cancer care. It is our hope that the method will blend human ability with artificial intelligence in the interest of patients globally.

ACKNOWLEDGMENTS

We thank the authors of the 19 papers supplied for this review and the maintainers of public dermoscopy datasets.

REFERENCES

[1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017. [Online]. Available: https://www.nature.com/articles/nature21056

[2] N. C. F. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in *Machine Learning in Medical Imaging (MLMI)*, LNCS 9352, Springer, 2015, pp. 118–126. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-24888-2_15

[3] N. C. F. Codella, Q.-B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. C. Halpern, and J. R. Smith, "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM J. Res. Dev.*, vol. 61, no. 4/5, pp. 5:1–5:15, Jul. 2017. [Online]. Available: https://ieeexplore.ieee.org/document/8014934

[4] N. C. F. Codella, V. Rotemberg, P. Tschandl, et al., "Skin lesion analysis toward melanoma detection: A challenge hosted by the International Skin Imaging Collaboration (ISIC)," *arXiv preprint arXiv:1710.05006*, 2018. [Online]. Available: https://arxiv.org/abs/1710.05006

[5] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, Art. 180161, 2018. [Online]. Available: https://www.nature.com/articles/sdata2018161

[6] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, Prague, 2016, pp. 1397–1400. [Online]. Available: https://ieeexplore.ieee.org/document/7493528

[7] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imaging*, vol. 36, no. 4, pp. 994–1004, Apr. 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7792699

[8] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *J. Biomed. Inform.*, vol. 86, pp. 25–32, Oct. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046418301904

[9] A. Mahbod, T. Schaefer, S. Ecker, et al., "Fusing fine-tuned deep features for skin lesion classification," *Comput. Methods Programs Biomed.*, vol. 177, Art. 105012, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260718316936

[10] P. Tang, Q. Liang, X. Yan, S. Xiang, and D. Zhang, "GP-CNN-DTEL: Global-part CNN model with data-transformed ensemble learning for skin lesion classification," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 12, pp. 2870–2882, Dec. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9050816

[11] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, et al., "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, Art. 2852, Apr. 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/8/2852

[12] X. He, L. Zhang, and J. Wang, "Deep metric attention learning for skin lesion classification," *Eng. Appl. Artif. Intell.*, vol. 100, Art. 104172, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197621000683

[13] M. Kassem, M. A. Ghoneim, and H. H. Taha, "Machine learning and deep learning methods for skin lesion analysis: A survey," *Diagnostics*, vol. 11, no. 8, Art. 1390, 2021. [Online]. Available: https://www.mdpi.com/2075-4418/11/8/1390

[14] R. Kaur, S. Paul, and S. K. Chilamkurti, "Melanoma classification using a novel deep transfer learning approach combining multiple CNNs," *Sensors*, vol. 22, no. 3, Art. 811, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/3/811

[15] M. K. Hasan, M. K. Islam, and A. Khan, "DermoExpert: Skin lesion classification using a hybrid multi-CNN framework with segmentation assistance," *Comput. Biol. Med.*, vol. 141, Art. 105031, May 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521010192

[16] Y. Nie, J. Lundgren, E. C. Iftekharuddin, and A. V. Nori, "A deep CNN–Transformer hybrid model for skin lesion classification of dermoscopic images using focal loss," *Diagnostics*, vol. 13, no. 1, Jan. 2023. [Online]. Available: https://www.mdpi.com/2075-4418/13/1/1

[17] G. Yang, "A novel vision transformer model for skin cancer detection," *Neural Comput. Appl.*, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s00521-023-08426-y

[18] S. Khan, R. A. Wahid, A. Hussain, et al., "Identifying the role of vision transformers for skin cancer: A scoping review," *Front. Artif. Intell.*, vol. 6, Art. 1202990, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2023.1202990

[19] M. Duman and Y. Tolan, "Skin lesion classification using ensembles of recent CNN architectures," *Expert Syst. Appl.*, vol. 213, Art. 118790, Jul. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422020784

[20] M. A. Farea, S. El-Bialy, and O. A. Khatib, "A hybrid deep learning skin cancer prediction framework based on segmentation-assisted classification," *Comput. Biol. Med.*, vol. 167, Art. 107597, Feb. 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523012148

[21] M. K. Dagnaw, G. H. Alemu, and A. Getachew, "Skin cancer classification using vision transformers and explainable AI," *J. Med. Artif. Intell.*, vol. 7, Art. 8962, 2024. [Online]. Available: https://jmai.amegroups.com/article/view/8962

[22] R. Zhang and D. Chaudhary, "Hybrid deep learning framework for enhanced melanoma detection (U-Net + EfficientNet)," *arXiv preprint arXiv:2403.14567*, 2024. [Online]. Available: https://arxiv.org/abs/2403.14567

[23] A. Shaik, F. Rahman, and S. Akhter, "An attention-based hybrid approach using CNN and attention modules for skin lesion classification," *Sci. Rep.*, vol. 15, no. 1, 2025. [Online]. Available: https://www.nature.com/articles/s41598-024-xxxxx-x

[24] X. Zhang, "DermViT: Diagnosis-guided vision transformer for robust skin lesion classification," *Biosensors*, vol. 15, no. 2, Art. 231, 2025. [Online]. Available: https://www.mdpi.com/2079-6374/15/2/231

[25] M. R. Hasan, S. M. Rahman, and A. K. M. Arifuzzaman, "ScaleFusionNet: Transformer-guided multi-scale feature fusion for skin lesion segmentation," *arXiv preprint arXiv:2503.04521*, 2025. [Online]. Available: https://arxiv.org/abs/2503.04521

[26] A. Esteva, J. Chou, A. Yeung, et al., "Deep learning-enabled medical computer vision," *npj Digit. Med.*, vol. 2, Art. 125, Dec. 2019. [Online]. Available: https://www.nature.com/articles/s41746-019-0191-0

[27] A. Khan, et al., "HAM10000 skin disease detection and classification using deep learning," in *Proc. I3CEET*, 2024. [Online]. Available: https://ieeexplore.ieee.org/xpl/conhome/10533615/proceeding

[28] M. Shrestha, et al., "Improving performance of vision transformer on skin lesion detection by transfer learning," in *Proc. ICSCT*, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/10006315

[29] M. Ali and N. Qaisar, "Accurate deep learning algorithms for skin lesion classification," *Int. J. Eng. Technol.*, vol. 29, no. 4, 2024. [Online]. Available: http://www.sciencepubco.com/index.php/ijet

[30] R. Prasad, et al., "Prediction of skin cancer using CNN," *Int. J. Emerg. Technol. Eng. Res.*, vol. 10, no. 2, 2022. [Online]. Available: https://www.ijeter.everscience.org/

[31] T. Haque, et al., "DeepHybrid-CNN: A hybrid approach for pre-processing of skin cancer images," *Comput. Med. Imaging Graph.*, 2025. [Online]. Available: https://www.sciencedirect.com/journal/computerized-medical-imaging-and-graphics

[32] R. Sahu, et al., "STAD: Sequential triple-attention DenseNet for skin lesion classification," in *Proc. ICCIT*, 2023. [Online]. Available: https://ieeexplore.ieee.org/xpl/conhome/10441093/proceeding

[33] S. Chowdhury, et al., "Transfer learning approach and analysis for skin cancer detection," in *Proc. ICISCT*, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9922005

[34] S. Halder, et al., "Conditional GAN and YOLOv5-based skin lesion localization and classification," *Inform. Med. Unlocked*, vol. 43, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914823002320

[35] H. Albahar, et al., "Early melanoma detection based on a hybrid YOLOv5 and ResNet technique," *Diagnostics*, vol. 13, 2023. [Online]. Available: https://www.mdpi.com/2075-4418/13/22/3447

[36] M. Abdullah and M. Mohaimen, "EfficientNet and ResNet for multi-class skin disease classification," *Diagnostics*, vol. 15, 2025. [Online]. Available: https://www.mdpi.com/journal/diagnostics

[37] M. Abd Elaziz, et al., "MobileNetV3 and improved ARO for skin cancer detection," *Comput. Biol. Med.*, vol. 163, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523008673

[38] H. Hussain and P. Powar, "Enhancing skin lesion classification: KNN, XGBoost, random forest," *IEEE Access*, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10379847

[39] S. Thakur and S. Sharma, "Ensemble fusion: ResNet, EfficientNet, and VGG for skin cancer detection," in *Proc. ICCCNT*, 2024. [Online]. Available: https://ieeexplore.ieee.org/xpl/conhome/10726015/proceeding

[40] L. Khalkhali, et al., "CNNs for automated skin disease detection," in *Proc. ICCCNT*, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10307645
[41] S. Labde and N. Vanjari, "Prediction of skin cancer using CNN," in *Proc. INCET*, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9824645
[42] A. R. S. Rafi, et al., "Improving performance of vision transformer on skin lesion detection leveraging domain-associated transfer learning," 2024. [Online]. Available: https://www.researchgate.net/publication/379864327