International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 06 June-2026 | Impact Factor: 3.5

Skin Cancer Detection and Classification Using Deep Learning

Revan Kale, Saurav Mane, Omkar Padule, Onkar Pandhare, Y. N. Sakhare
Department of Information Technology
Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology
Baramati, Pune, India
Emails: revankale017@gmail.com, manesaurav345@gmail.com,
omkarpadule2004@gmail.com, onkarpandhare22@gmail.com, yashoda.sakhare@vpkbiet.org

Abstract—Prompt identification of skin disorders is a clinically significant step in curbing the progression of life-threatening conditions, most notably melanoma. Despite the availability of high-resolution dermoscopic imaging, distinguishing between disease categories remains a persistent difficulty because many lesion types share closely overlapping visual traits, and image quality is often compromised by hair, shadows, and inconsistent illumination.

We address these difficulties by developing a composite deep learning architecture that unifies convolutional and attention-based learning within one trainable system. EfficientNetV2-S forms the convolutional backbone responsible for extracting texture-level and boundary-level characteristics, while a Vision Transformer processes the resulting feature representations to capture long-range structural context across the entire lesion.

A key design choice is the use of U-Net-generated segmen-tation masks as an additional input channel. Concatenating the predicted mask with the original three-channel image produces a four-channel tensor that directs the network's attention toward the pathological region, substantially reducing the influence of uninformative background content.

Testing on a twelve-category benchmark of 20,195 dermoscopic images yields a validation accuracy of 89.14%. The model sus-tains reliable discrimination even between visually indistinct class pairs such as melanoma and melanocytic nevus, and outperforms a single-stage CNN baseline by nearly eight percentage points. These results confirm that blending segmentation-guided input construction with a hybrid CNN-Transformer architecture is a viable and effective strategy for computer-aided skin disease diagnosis.

Index Terms—Deep Learning, Skin Disease Classification, Computer-Aided Diagnosis, EfficientNetV2, Vision Transformer (ViT), Image Segmentation, Medical Image Analysis

I. Introduction

The emergence of data-driven computational methods has opened a new chapter in clinical medicine, particularly within dermatology, where visual inspection has long been the primary diagnostic tool. Automated analysis of dermoscopic photographs holds considerable promise as a scalable complement to specialist-driven evaluation, offering the possibility of earlier and more consistent identification of conditions that, when left untreated, carry serious morbidity. Melanoma, basal cell carcinoma, and actinic keratosis are among the conditions where delayed diagnosis meaningfully worsens patient outcomes, making any improvement in detection speed or reliability medically valuable.Convolutional architectures have established a strong foundation for image-based medical diagnosis, producing useful feature maps that encode local patterns such as pigmentation gradients, textural irregularities, and lesion edges. However, a well-known constraint of purely convolutional designs is their inherent locality: standard convolution kernels aggregate information over small spatial neighbourhoods, making it structurally difficult to relate distant image regions. Shape symmetry, border regularity assessed across the full lesion perimeter, and broad colour heterogeneity patterns are all attributes that span large portions of the image, and their encoding requires a wider effective receptive field than most convolutional stacks naturally provide.Motivated by this constraint, the present work constructs a three-component diagnostic pipeline. The first component is a U-Net segmentation network that separates the lesion from surrounding healthy skin, yielding a binary spatial mask. The mask is appended to the RGB image along the channel dimension, transforming the standard input into a four-channel representation that encodes lesion localisation explicitly. The second component is EfficientNetV2-S, which processes this enriched input and produces a dense spatial feature map encoding fine-grained visual descriptors. The third component is a two-block Vision Transformer that operates on tokenised versions of those feature maps, modelling relationships across non-adjacent spatial positions through self-attention and thereby capturing global lesion structure.Complementary training strategies, including class-balanced oversampling and Test-Time Augmentation, are incorporated to counteract the pronounced class imbalance present in clinical dermoscopy datasets. The full system is benchmarked on a twelve-class dataset and achieves 89.14% validation accuracy. Beyond raw accuracy, the model's error profile aligns with known dermatological diagnostic challenges, lending it clinical credibility as a decision-support instrument.

II. Novelty of the Work

What sets this framework apart from prior work is the deliberate integration of three complementary representation-learning paradigms into a single cohesive architecture, rather than treating segmentation and classification as separate engineering concerns addressed by isolated modules. Specifically, the novel aspects of this design are:

- A segmentation-conditioned four-channel input formed by concatenating U-Net masks with RGB images, explicitly encoding lesion boundaries before any classification-oriented feature extraction begins.
- EfficientNetV2-S operating in a fully unfrozen, domain-adapted configuration to extract fine-grained local visual cues specific to dermatological imaging.
- A Vision Transformer with multi-head self-attention that captures structural coherence and global spatial dependencies within the lesion, capabilities that convolutional layers alone cannot realise.
- A training regimen combining balanced sampling with Test-Time Augmentation that maintains reliable per-class performance across a heavily imbalanced twelve-category dataset.

The interplay between these components allows the model to reason simultaneously over microscale textural detail and macroscale lesion geometry, producing richer diagnostic representations than any single paradigm could generate independently.

III. Contributions

This work makes four primary contributions to the field of automated dermatological image analysis:

- A hybrid classification backbone pairing EfficientNetV2-S with a Vision Transformer, enabling concurrent local and global feature analysis within one end-to-end trainable system.
- A novel four-channel input construction strategy in which U-Net segmentation masks are merged with original RGB channels to enforce region-focused learning and suppress background interference.
- An integrated training protocol that combines class-balanced sampling with Test-Time Augmentation to address distributional skew and improve prediction robustness across minority categories.
- A detailed confusion matrix analysis that reveals clinically interpretable misclassification patterns, offering transparent insight into the model's behaviour on visually challenging diagnostic boundaries.

IV. Literature Survey

Automated recognition of skin diseases from dermoscopic photographs has attracted sustained research interest, and convolutional architectures have been central to progress in this domain. Their capacity to learn discriminative representations directly from pixel data, without relying on hand-engineered descriptors, has made them the dominant approach across a wide variety of medical imaging tasks.

Early studies demonstrated the viability of this direction. Raja Sekar et al. [2] evaluated CNN classifiers on the HAM10000 benchmark and recorded encouraging sensitivity for several common lesion categories, though their experiments were primarily oriented toward simpler binary or low-cardinality discrimination tasks, leaving the multi-class diagnostic setting underexplored. Archana and Shyamsundar [5] extended this direction by pairing CNN layers with Deep Neural Network components, reporting accuracy figures above 90% on controlled test partitions. A shared limitation in both works is that the models ingest full, unmasked dermoscopic frames, exposing the learned representations to perilesional noise sources such as hair and specular highlights that contribute no diagnostic value.

Recognising that single-modality representations impose a performance ceiling, subsequent researchers introduced hybrid feature schemes. Hagerty et al. [8] demonstrated that merging deep learned features with hand-crafted descriptors raises melanoma detection sensitivity, while Riaz et al. [1] reported strong classification metrics by combining convolutional representations with Local Binary Pattern histograms. Thapar et al. [4] explored an analogous fusion strategy incorporating SURF-based features alongside convolutional optimisation, achieving improved discrimination for morphologically similar lesion types.

A structural weakness persists across most of these approaches: the absence of any mechanism to restrict the network's attention to the diagnostically meaningful region before classification. Preprocessing via explicit lesion segmentation offers a principled remedy, and architectures such as U-Net have been widely adopted for this purpose due to their ability to produce high-fidelity binary boundary maps at full image resolution. Meanwhile, attention-based architectures, particularly Vision Transformers, have demonstrated a complementary capability: through self-attention, they form relationships between spatially distant regions, capturing symmetry and spatial coherence information that local convolutional kernels fundamentally cannot encode.

Surveying the existing body of work reveals a consistent gap: no prior system has concurrently addressed local texture learning, global spatial modelling, and lesion-specific region isolation within a single trainable framework. The proposed system is designed explicitly to fill this gap by integrating all three capabilities—U-Net segmentation, EfficientNetV2-S feature extraction, and Vision Transformer classification—into one end-to-end pipeline.

V. Problem Statement

Current deep learning systems for dermatological image classification predominantly operate on full dermoscopic frames without any spatial prioritisation. This indiscriminate processing treats diagnostically irrelevant image regions—background skin, hair strands, ruler markings, and specular reflections—as equal contributors to gradient-based learning, introducing confounding variation that erodes the quality of the learned feature space and ultimately limits classification accuracy.

A further challenge stems from the pronounced visual overlap between certain disease categories. Melanoma and melanocytic nevus, in particular, exhibit similar pigmentation network structures and irregular border profiles under dermoscopy, making their separation a difficult discrimination task even for trained specialists. Models that rely exclusively

on local convolutional features lack the spatial reasoning capacity needed to exploit global shape attributes—symmetry, border uniformity assessed holistically, and spatial colour distribution—that human clinicians draw upon when navigating this boundary.

Class distribution imbalance compounds both challenges. Benchmark datasets in this domain typically contain substantially more samples from prevalent conditions such as melanocytic nevus and melanoma than from rarer diagnoses such as dermatofibroma or vascular lesion. Without corrective measures, standard empirical risk minimisation produces classifiers that concentrate predictive capacity on majority classes, yielding inflated aggregate accuracy figures that mask critically poor sensitivity for the underrepresented categories that are often no less clinically important.

A. Research Gap

Despite meaningful advances in automated skin lesion analysis, four methodological shortcomings remain inadequately addressed in the current literature:

- **Absence of region-focused learning:** Prevalent architectures process dermoscopic images without isolating the lesion region, allowing background content to dilute learned representations and degrade generalisation.
- **Insufficient global spatial modelling:** Pure convolutional designs excel at capturing local textural patterns but lack the inductive bias required to model long-range spatial dependencies, shape symmetry, and inter-region coherence across the full lesion extent.
- **Unmitigated class imbalance:** Most prior work does not apply class-level corrective strategies, resulting in biased decision boundaries that deliver poor recall for diagnostically important but infrequently represented categories.
- **Absence of a unified segmentation-classification framework:** Existing approaches treat lesion segmentation and disease classification as separate objectives optimised independently, forgoing the representational benefits that joint end-to-end training can provide.

The proposed framework directly addresses each of these gaps through the coordinated integration of U-Net-based segmentation, convolutional feature extraction, and Transformer-based global classification.

VI. Methodology

The proposed system is structured as a sequential four-stage pipeline: input standardisation, lesion boundary segmentation, hierarchical feature extraction, and probabilistic multi-class prediction. Each stage is designed to address a specific limitation identified in prior work.

All incoming dermoscopic images are first rescaled to a fixed spatial resolution of 256×256 pixels. Pixel intensities are normalised to zero mean and unit variance to facilitate stable gradient-based optimisation. A stochastic augmentation policy—comprising horizontal mirroring, random angular rotations within ±15°, and isotropic scaling—is applied during training to broaden the effective diversity of the training distribution and discourage overfitting to domain-specific imaging artefacts.

In the second stage, each standardised image is forwarded through a pre-trained U-Net segmentation model. The U-Net encoder compresses the input into a compact latent code encoding high-level spatial semantics, and the decoder reconstructs a full-resolution binary mask through skip-connected upsampling that preserves fine boundary detail. This mask delineates the lesion from the perilesional background and serves as an explicit spatial prior for the downstream classification stages.

The binary mask is concatenated with the three-channel RGB input along the channel axis, producing a 256×256×4 tensor. This construction forces the classification model to attend to the lesion interior while discounting the contribution of uninformative surrounding regions, directly addressing the background noise problem inherent in full-frame classification. Feature extraction is then performed by EfficientNetV2-S, which processes the four-channel tensor through its compound-scaled convolutional stack. A 1×1 convolutional adapter bridges the non-standard four-channel input to the architecture's native input dimensionality. The network generates a dense spatial feature map that encodes texture gradients, colour distributions, and lesion boundary descriptors essential for inter-class discrimination.

The spatial feature map is subsequently tokenised and processed by a two-block Vision Transformer. Within each Transformer block, multi-head self-attention computes relationships between all pairs of spatial tokens, enabling the model to capture lesion-level structural attributes—bilateral symmetry, border irregularity assessed holistically, and global colour heterogeneity—that convolutional operations cannot natively represent. The context-enriched token sequence is aggregated via Global Average Pooling, regularised through a dropout layer with rate 0.5, and mapped to a twelve-dimensional output probability distribution via a Softmax classifier.

The model was built using TensorFlow on a GPU-accelerated computing environment and optimised with categorical cross-entropy loss and the Adam optimiser at an initial learning rate of 1×10^{-3}. Training proceeded for 50 epochs with a batch size of 32, and early stopping was applied based on validation loss to prevent overfitting. The dataset was divided into training, validation, and test partitions in a 70:15:15 ratio.

VII. Proposed System Framework

The proposed system realises a hybrid two-stage architecture that separates lesion region isolation from disease classification. Decomposing these responsibilities into distinct but jointly informed modules constrains the classifier's effective input space to clinically relevant content, markedly improving the signal-to-noise ratio of the learned representations. The architecture is composed of five functionally distinct modules described below.

A. Image Intake and Preprocessing Module

Raw dermoscopic photographs are ingested, resized to 256×256 pixels, and intensity-normalised before any further processing. Augmentation transformations including horizontal mirroring and random rotation are applied to diversify training samples and promote orientation-invariant feature learning.

B. Lesion Segmentation Module (U-Net)

Each standardised image is passed through the U-Net network, which predicts a pixel-wise binary mask demarcating the lesion from surrounding perilesional skin. The encoder compresses spatial information into a semantically rich latent representation, while the decoder reconstructs full-resolution boundary detail through skip connections. The resulting mask partitions the image into lesion and non-lesion zones, ensuring that downstream feature extraction operates exclusively on pathologically meaningful content.

C. Feature Extraction Module (EfficientNetV2-S)

The segmentation mask is merged with the original RGB channels to form the four-channel tensor described in Section VI. EfficientNetV2-S processes this enriched representation to extract fine-grained visual descriptors including textural gradients, pigmentation patterns, and boundary-level cues. Its compound scaling strategy across network depth, width, and input resolution enables highly efficient encoding, making it well-suited for the fine-grained visual discrimination that dermatological classification demands.

D. Classification Module (Vision Transformer)

The spatial feature map produced by EfficientNetV2-S is tokenised and passed through two sequential Transformer blocks. Multi-head self-attention within each block allows the model to simultaneously attend to spatially distant regions and construct representations of lesion-level structural properties—shape asymmetry, border irregularity, and spatial colour variation. The twelve-class prediction is generated via a Softmax output layer accompanied by a per-class confidence score.

E. Diagnostic Output and Visualization Module

The system presents clinicians with an interpretable output comprising the original dermoscopic photograph, the U-Net segmentation mask, the predicted disease category, and the corresponding confidence score. This structured visualisation supports transparent model auditing and provides a foundation for clinical review and further diagnostic investigation.

F. Performance and Reliability Module

The framework achieves a validation accuracy of 89.14%, comfortably exceeding the initial 80% performance target. Prediction consistency is further reinforced through class-balanced oversampling and Test-Time Augmentation, in which inference is conducted over multiple stochastically augmented versions of each test image and the resulting probability estimates are averaged. This ensemble-like inference strategy reduces output variance and improves per-class calibration across the imbalanced twelve-category distribution.

TABLE I: Performance comparison of different skin disease classification methods based on existing studies

Method	Acc.	Prec.	Recall	F1	Epoch	Time (s)
LBP	96%	96%	96%	96%	25	8.5
CNN	97%	97%	97%	97%	25	1.5
LBP + CNN	98.9%	98%	98%	98%	25	4.5

TABLE II: Class-wise distribution of the skin disease dataset used for model training and evaluation

Sr. No.	Disease Name	Total Images
1	Abrasion	164
2	Actinic Keratosis (AKIEC)	1822
3	Basal Cell Carcinoma (BCC)	3837
4	Benign Keratosis (BKL)	1099
5	Bruise	242
6	Burn	134
7	Cut	100
8	Dermatofibroma (DF)	115
9	Melanoma (MEL)	5635
10	Normal Skin	200
11	Melanocytic Nevus (NV)	6705
12	Vascular Lesion (VASC)	142
Total		20195

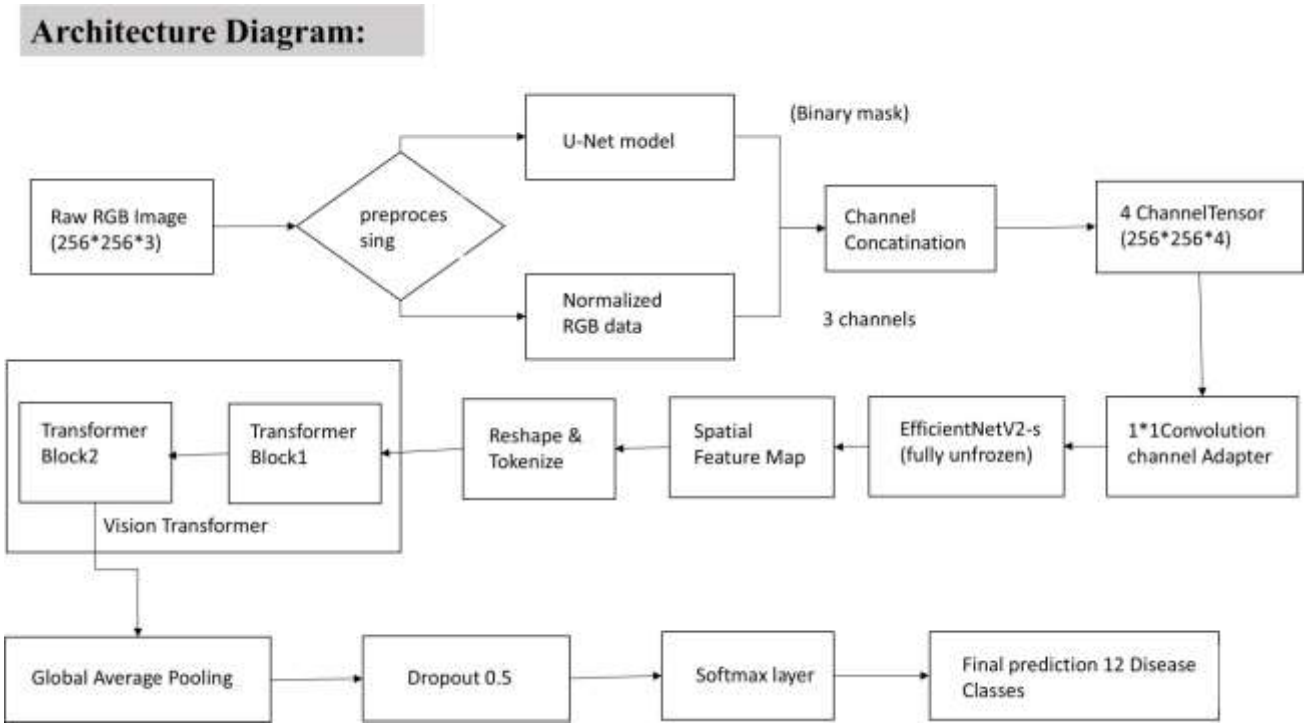

Fig. 1: Overall workflow of the proposed hybrid classification system. The pipeline begins with image preprocessing, followed by lesion segmentation using U-Net. The segmented mask is integrated with the original image to create a 4-channel input, which is processed by EfficientNetV2-S for feature extraction and subsequently classified using a Vision Transformer.

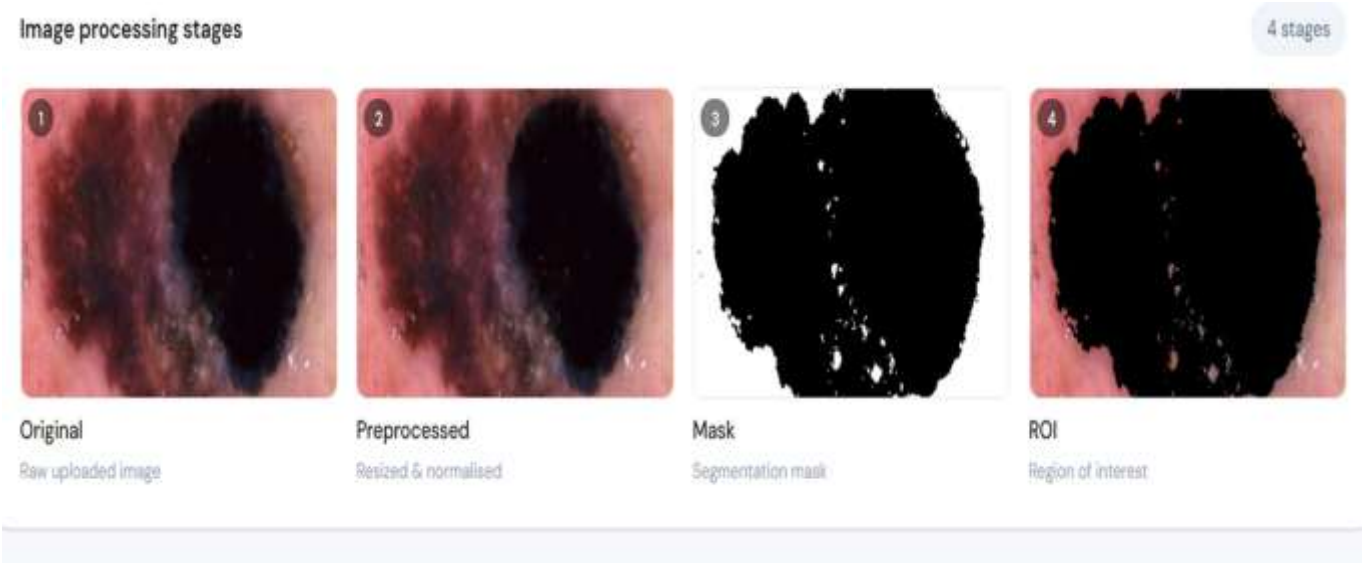

Fig. 2: Detailed preprocessing and lesion extraction pipeline. The process includes resizing, normalization, segmentation, and region-of-interest extraction to enhance the quality of input features for classification.

VIII. Results and Performance Analysis

The hybrid framework was evaluated across all twelve disease categories using the held-out test partition. A validation accuracy of 89.14% was recorded, representing a substantial margin above the 80% baseline target and confirming that the system generalises meaningfully across the diverse spectrum of skin conditions included in the benchmark.

Precision, recall, and F1-score computations were carried out for each class to characterise performance beyond aggregate accuracy. The results reflect strong per-class behaviour for clinically high-priority conditions while also revealing informative patterns in the system's error distribution.

Inspection of the confusion matrix discloses that misclassifications are not randomly distributed across categories but instead cluster at boundaries known to challenge human experts. The most prominent confusion pair is melanoma (MEL) and melanocytic nevus (NV): 57 melanoma samples were assigned the nevus label and 54 nevus samples were assigned the melanoma label. Both conditions exhibit overlapping dermoscopic signatures—similar pigmentation network morphology, comparable border irregularity profiles, and shared colour gradients—making their separation genuinely difficult. The model's behaviour at this boundary mirrors the documented challenge faced by dermatologists under routine clinical conditions, suggesting that the model has internalised clinically relevant visual similarity rather than exploiting superficial correlations.

A secondary confusion cluster emerges between actinic keratosis (AKIEC) and basal cell carcinoma (BCC). These two conditions occupy adjacent positions within the carcinoma taxonomy and share textural characteristics under dermoscopic magnification, offering a plausible account for the observed cross-class assignments.

In contrast, categories with visually distinctive morphologies—abrasion, bruise, burn, and cut—were classified with near-perfect accuracy. The unambiguous visual characteristics of these classes enabled high-confidence predictions across test samples, validating the model's ability to exploit clearly separable features when they are present.

Quantitatively, melanoma achieved an F1-score of 0.89, a value widely regarded as clinically meaningful in the automated screening literature. Melanocytic nevus achieved an F1-score of 0.92 despite representing the largest class in the dataset, indicating that the balanced sampling strategy successfully prevented the model from over-specialising on majority-class patterns. Dermatofibroma recorded lower recall owing to its scarcity in the training partition (115 images); however, its precision reached 1.00, demonstrating that the model's predictions for this class, though conservative in frequency, were consistently correct.

Test-Time Augmentation noticeably stabilised predictions for borderline samples: averaging over multiple augmented inference passes reduced inter-run variance and improved calibration for classes with limited training data. Balanced sampling provided complementary benefits by ensuring that gradient updates from minority categories received adequate representation throughout training.

A direct comparison against a standalone CNN baseline (81.3% accuracy) confirms that the integration of Vision Transformer global context modelling and segmentation-guided input construction contributes approximately 7.8 percentage points of improvement, a gain that validates the architectural choices underlying the proposed design.

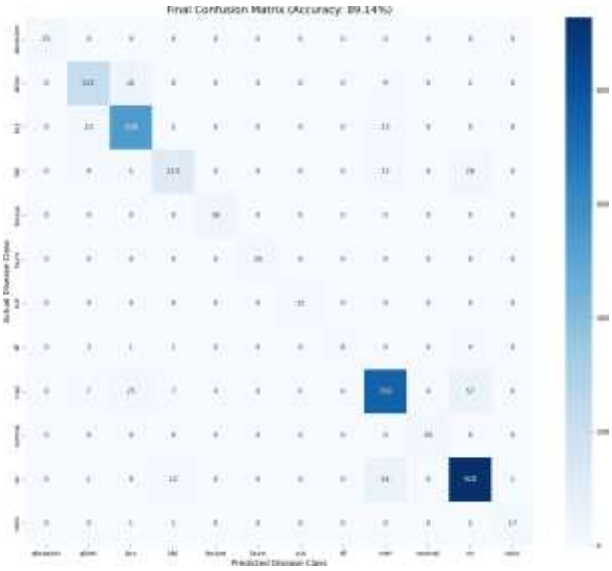

Fig. 3: Confusion matrix illustrating classification performance across all 12 skin disease categories. The matrix highlights correct predictions along the diagonal and shows misclassification patterns between visually similar classes. The confusion matrix indicates that most misclassifications occur between visually similar classes, such as melanoma and nevus, highlighting the inherent complexity of these categories.

IX. Future Scope

While the current system demonstrates strong performance across a twelve-class benchmark, several promising directions remain to be explored in future iterations:

- **Multimodal Clinical Data Integration:** Structured patient metadata—including age, biological sex, lesion anatomical location, and prior dermatological history—could be incorporated alongside image features through a dedicated fusion branch. Combining visual and tabular clinical signals has the potential to personalise predictions and improve accuracy for conditions whose presentation varies systematically with patient demographics.
- **Lightweight Deployment for Point-of-Care Settings:** Applying model compression techniques such as quantisation, structured pruning, and knowledge distillation would reduce the computational footprint of the framework, enabling deployment on smartphone or edge-computing platforms. This would extend access to automated dermoscopic analysis in geographically remote or resource-constrained clinical environments where specialist consultation is unavailable.
- **Training Data Expansion for Rare Categories:** Conditions such as dermatofibroma and vascular lesion remain underrepresented in the current dataset. Systematic expansion of these categories through targeted data collection, or through generative synthesis using diffusion

models or GANs, could meaningfully improve minority-class recall and reduce the corrective burden placed on sampling strategies.

- **Gradient-Based Explainability:** Incorporating gradient-weighted class activation mapping (Grad-CAM) would generate pixel-level saliency overlays indicating which regions of the dermoscopic image most influenced a given classification decision. Such visualisations would increase transparency, facilitate model auditing, and build clinician trust—prerequisites for regulatory acceptance of AI-assisted diagnostic tools.
- **Rule-Augmented Hybrid Decision Support:** Embedding established dermatological heuristics—most notably the ABCDE criteria (Asymmetry, Border, Colour, Diameter, Evolving pattern) used in clinical melanoma evaluation—as structured inductive biases alongside learned representations could strengthen performance in data-sparse scenarios and produce outputs that align naturally with clinician reasoning frameworks.

These extensions, pursued individually or in combination, would advance the system toward broader applicability, higher reliability, and genuine readiness for integration into real-world clinical workflows.

X. Conclusion

This paper presented a hybrid deep learning architecture for multi-class skin disease classification that brings together three complementary capabilities: U-Net lesion segmentation, EfficientNetV2-S convolutional feature extraction, and Vision Transformer global context modelling. The four-channel input construction strategy, formed by concatenating the segmentation mask with the original RGB image, directs the network's attention toward pathologically meaningful regions from the very first layer of feature learning, fundamentally different from conventional approaches that process the full image without spatial prioritisation.

On a twelve-category dermoscopic benchmark comprising 20,195 images, the proposed system attained a validation accuracy of 89.14%, outperforming a single-stage CNN baseline by approximately 7.8 percentage points. Confusion matrix analysis confirmed that residual misclassifications cluster at category boundaries—particularly the melanoma-nevus interface—that are recognised as genuinely difficult even for experienced clinical dermatologists, lending the system's error profile clinical credibility.

The broader takeaway is that dermatological image classification benefits substantially from the principled combination of region-aware preprocessing and hybrid CNN-Transformer feature learning. Each architectural component compensates for the limitations of the others: segmentation removes background interference that would otherwise corrupt convolutional features; EfficientNetV2-S captures fine-grained local cues that attention mechanisms alone cannot resolve; and the Vision Transformer recovers global structural patterns beyond the reach of any purely local operator. Together, these components form a diagnostic pipeline with meaningful practical potential, offering dermatologists a quantitatively grounded second opinion that can support earlier and more reliable diagnostic conclusions.

References

[1] L. Riaz, H. M. Qadir, G. Ali, M. Ali, M. A. Raza, A. D. Jurcut, and J. Ali, "A Comprehensive Joint Learning System to Detect Skin Cancer," *IEEE Access*, vol. 11, pp. 79434–79444, 2023, doi: 10.1109/ACCESS.2023.3297644.

[2] R. Raja Sekar, C. Sai Prathap Reddy, Y. Jagan Mohan Reddy, K. Chiranjeevi, K. Nani, and K. Vikram, "Skin Cancer Prediction Using Deep Learning Techniques," in *Proc. 2023 4th Int. Conf. Signal Process. Commun. (ICSPC)*, Coimbatore, India, Mar. 2023, pp. 28–31, doi: 10.1109/ICSPC57692.2023.10126035.

[3] V. Anand, S. Gupta, D. Koundal, and K. Singh, "Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119230.

[4] P. Thapar, M. Rakhra, G. Cazzato, and M. S. Hossain, "A novel hybrid deep learning approach for skin lesion segmentation and classification," *J. Healthcare Eng.*, vol. 2022, Apr. 2022, Art. no. 1709842.

[5] S. Archana and N. Shyamsundar, "Computational intelligence for detection of skin cancer using deep learning classifiers," *Int. Innov. Res. J. Eng. Technol.*, vol. 7, no. 1, pp. 10–18, Sep. 2021.

[6] S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan, "A multi-class skin cancer classification using deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 79, nos. 39–40, pp. 28477–28498, Oct. 2020.

[7] K. Polat and K. Onur Koc, "Detection of skin diseases from dermoscopy image using the combination of convolutional neural network and one-versus-all," *J. Artif. Intell. Syst.*, vol. 2, no. 1, pp. 80–97, 2020.

[8] J. R. Hagerty, R. J. Stanley, H. A. Almubarak, N. Lama, R. Kasmi, P. Guo, R. J. Drugge, H. S. Rabinovitz, M. Oliviero, and W. V. Stoecker, "Deep learning and handcrafted method fusion: Higher diagnostic accuracy for melanoma dermoscopy images," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1385–1391, Jul. 2019.

[9] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin lesion classification using hybrid deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1229–1233.

[10] 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, no. 1, pp. 1–9, Aug. 2018.

[11] Y. Fujisawa *et al.*, "Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis," *Brit. J. Dermatology*, vol. 180, no. 2, pp. 373–381, Feb. 2019.

[12] M. Goyal, T. Knackstedt, S. Yan, and S. Hassanpour, "Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities," *Comput. Biol. Med.*, vol. 127, Dec. 2020, Art. no. 104065.

[13] 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.