Preserving Spatial and Hierarchical Lesion Structures in Automated Skin Cancer Diagnosis through DermaCap and CCSA

Najeeb Ullah, Muhammad Shahan Ibad*, Sayed Naeem Abbas, Shakir Hussain, Muhammad Ashir, Kaisar Khan
Center of Excellence in IT, Institute of Management Sciences (IMSciences), Peshawar, Pakistan
*Correspondence: najeebafridi.ims@gmail.com

Citation | Ullah. N, Ibad. M. S, Abbas. S. M, Hussain. S, Ashir. M, Khan. K, "Preserving Spatial and Hierarchical Lesion Structures in Automated Skin Cancer Diagnosis through DermaCap and CCSA", IJIST, Vol. 8 Issue. 2 pp 743-759, April 2026
Received | March 18, 2026 **Revised |** April 20, 2026 **Accepted |** April 25, 2026 **Published |** April 30, 2026.

Early and reliable detection of skin cancer remains a global clinical priority, particularly for rare lesion categories that are often underrepresented due to limited data and severe class imbalance. To address these challenges and the limitations of traditional Convolutional Neural Networks (CNNs) in preserving spatial relationships, this study proposes DermaCap, a hybrid CNN–Capsule Network architecture designed to capture lesion pose, orientation, and hierarchical structural features. The framework follows a rare-class-first design philosophy, optimizing the model architecture, dataset curation, and training strategy to improve sensitivity toward diagnostically challenging lesions. Additionally, Clinically Constrained Structural Augmentation (CCSA) is introduced to enhance dataset diversity while preserving dermatologically meaningful morphology and color realism. The proposed model was evaluated using a carefully selected six-class subset of the HAM10000 dermatology database, which contains a total of 3,310 dermatoscopic images representing multiple diagnostic categories. The evaluation resulted in an approximate accuracy of 91%. The precision, recall, and F1-score show consistent values across both majority and minority classes. Therefore, the proposed model shows stable and consistent diagnostic performance for dermatoscopic image classification. Overall, DermaCap presents a clinically aligned, structure-aware approach for more trustworthy and scalable AI-assisted skin lesion diagnosis.

Keywords: Skin Lesion Classification, Automated Skin Cancer Diagnosis, Deep Learning in Dermatology, Hierarchical Lesion Feature Preservation, DermaCap Augmentation Framework

Introduction:

The human body's largest organ is the skin which serves the important function of protecting the body from external threats and harmful elements. However, if skin cells begin to grow uncontrollably, it will produce an aggressive condition known as Skin Cancer - the most common form of cancer worldwide [1]. All varieties of skin cancer are not equally easy to treat; for instance, Melanoma is one of the deadliest skin cancers unless diagnosed in the early stages. If skin cancer is detected and treated in the early stages, the probability of complete recovery is almost always very high [2]. Therefore, it is extremely difficult for a person to determine, even with the naked eye, whether they have a normal mole or an aggressive tumour, as they can appear identical.

The number of people diagnosed with skin cancer is increasing worldwide, and it remains the most common type of cancer. Melanoma is responsible for a large percentage of deaths caused by skin cancers yet is less common than other types of skin cancers. Numerous global cancer statistics indicate millions of new skin cancer cases per year, which indicates the need to continue assessing current methods for early diagnosis because patients diagnosed with skin cancers at their earliest stage often experience improved survival rates compared to patients diagnosed after the disease has progressed. Clinicians also find it difficult to differentiate between benign and malignant skin lesions visually, especially when they lack extensive diagnostic experience.

Physicians frequently employ a dermatoscope, a specialized magnifying device, in order to differentiate between lesions by examining them in the deeper layers of skin. This requires considerable training and experience in dermatology, thus making accurate diagnoses difficult using this device alone. There are too few dermatologists in many countries around the globe; therefore, patients are subjected to lengthy waits and are often misdiagnosed [3]. A need exists for new technologies that will assist physicians with faster and better identification of skin abnormalities.

The recent advances in artificial intelligence (AI) have demonstrated remarkable promise in furthering medical tasks. In particular, deep learning can be trained on thousands of medical images, allowing them to identify patterns for various diseases similar to the way a human learns to recognize patterns [4]. Using AI models, we can build digital assistants that enable physicians to more accurately identify skin cancers than they typically would without use of an AI model. In this study, we will examine the use of advanced computer programs to analyze images of skin, with the intention of making diagnostic tools that save lives available to everyone, regardless of where they reside.

CNNs are typically great at classifying images but do not do a good job of maintaining spatial relationships since pooling operations discard some amount of detail regarding the original positioning of parts of an image. A lot of prior research on CNNs and Capsule Networks has demonstrated that the way CNNs encode pose and hierarchy is not very precise, resulting in problems performing tasks that require preservation of spatial structure within the input data [5]. In a medical setting, knowing how the various textures, edges, and colors of the skin lesion relate to one another is the most important factor determining success. DermaCap was created to overcome the limitations of previous studies by developing a specialized hybrid system for precise skin lesion analysis. DermaCap combines the CNN's ability to extract high-level features with the Capsule Network's ability to perform higher-level spatial reasoning.

Although recent advances in deep learning have helped improve the classification of skin lesions, the majority of methods currently used for this purpose use traditional architectures of CNNs that do not preserve hierarchical structures or fine-grained spatial relationships. In addition, most studies focusing on skin lesion classification tend to overlook

the issue of maintaining diagnostic sensitivity for rare or under-represented skin lesions in classifying imbalanced datasets. As a result, there exists an important research gap for developing clinically valid models that are able to preserve the structural integrity of skin as well as provide equal performance for all types of skin lesions.

The architecture proposed for this research utilizes Convolutional Neural Networks as the primary way of providing visual sensory input that captures the complex patterns, textures and edges found within dermatoscopic images of skin lesions. The features produced from the CNN are then forwarded to a Capsule-based backend where they are assembled into capsule vectors which denote not only their existence but also their physical properties (e.g., size, location, orientation, etc.) in relation to one another. By creating these capsule vectors, DermaCap can achieve equivariance and also establish an internal structural representation of the lesion. Because these two paradigms are integrated, this architecture is able to move beyond pixel-level detection and provide an understanding of the complex hierarchical relationships that exist within a skin lesion. Due to this ability, DermaCap will offer a more complete and clinically relevant classification of skin lesions than traditional scalar-based neural networks.

The remaining sections of this study are organized to comprehensively evaluate the DermaCap framework. Related Literature Review can be found in Section 2, where the findings from deep learning advances for skin lesion classification are summarized and then the deficiencies in standard CNN methodology identified. Methodology is covered in Section 3 and describes the use of hybrid CNN-Capsule layer designs used in DermaCap to create a standard procedure for pre-processing images before analyzing these lesion images due to variability in patients' appearance and conditions. Section 4 provides all experimental results for DermaCap and contrasts them against benchmarked performances achieved through conventional model comparison by way of three primary metrics: Accuracy, Sensitivity and Specificity. Finally, in Section 5, the study concludes by summarizing the most important findings based on evidence gathered during the study (i.e., the use of deep learning has improved significantly over traditional diagnostic methods) and provides guidance for future studies on automated dermatology diagnostic processes.

Problem Statement:

Classifying skin lesions correctly is still an ongoing problem because of low contrast between the skin lesion and its surrounding skin, the lack of well-defined boundaries, and the subtle differences in structure between images captured with dermoscopy. The challenge is exacerbated by a severe class imbalance, or a lack of equal representation, between different types of lesions. Traditional CNNs have been successful at classifying images accurately when there is a well-defined boundary around the skin lesion; however, traditional CNNs do not retain spatial relationships and hierarchical feature representations, which can result in the loss of important diagnostic information. Additionally, traditional CNNs tend to be biased toward the majority class in a dataset, which decreases their ability to accurately detect rare yet clinically important lesions. Therefore, a structure-preserving, class-balancing learning framework is necessary to maintain diagnostic fidelity across a wide variety of lesion types.

Objectives:

The objectives of this research are:

Create a hybrid CNN-Capsule Network (DermaCap) to accurately classify skin lesions.

Maintain spatial and hierarchical features of skin lesions, which are not preserved well by traditional CNN methods.

Improve the detection of the less common classification of skin lesions by correcting class imbalance.

Create clinically relevant methods for augmenting existing datasets (CCSA) to improve dataset quality and variety.

Assess DermaCap model performance on the HAM10000 dataset using multiple performance metrics: accuracy, precision, recall, and F1-score.

Ensure DermaCap diagnostic performance is robust and reliable enough to support real-world clinical use.

Literature Review:

In their 2025 paper, Akinrinade and Du suggested that a deep learning system might be developed to identify skin cancers, using convolutional neural networks (CNNs) [6]. The authors used 1,000 images from the ISIC-2017 dataset, and included images from the HAM10000 dataset (collected as part of the ISIC-2017 study) to design their models. The TensorFlow-based CNN model contained several successive convolutional layers that used ReLU as the activation function, L2 regularization, batch normalization, and max pooling. After the convolutional operations, the model used fully connected layers to classify the input images. To address class imbalance and limited data availability, the authors used GANs to augment their dataset; exploited transfer learning using weights pretrained on ImageNet; and, employed few-shot learning. Using a combined CNN–GAN framework was an effective method to address class imbalance; therefore, a very high level of accuracy was achieved to differentiate between malignant and benign lesions. Additionally, the authors noted that a clinical implementation of the proposed system in a digital health platform and mobile applications could support early diagnostics for populations in regions where access to care is limited.

[7] suggested an innovative framework for skin lesion diagnostics employing neural networks and XAI for greater clinical interpretability [7]. The researchers used the HAM10000 data set which consists of 10015 images from seven classes — there was a significant imbalance between classes, as approximately 67% of total images were melanocytic nevi. The model architecture is a hybrid between an EfficientNetV2-L that was pretrained on ImageNet, as well as incorporating channel attention mechanisms and a three-stage method of progressive fine-tune. The researchers mitigated data imbalance using a combination of targeted geometric and photometric procedures, and explanation was provided through Grad-CAM and saliency maps. The authors claimed that their method achieved an accuracy of 91.15%, a micro-average AUC of 99.33%, and a macro F1-value of 85.45%. Furthermore, they asserted that the incorporation of XAI improved alignment with dermatological guidelines as well as stakeholders' trust in the model, and that EfficientNetV2 provides a valid and easily interpretable alternative to transformer type architectures for resource-limited healthcare environments.

[8] have researched the use of transfer learning and convolutional neural networks in detecting skin cancer using the HAM10000 dataset, a total of 10,015 images divided into seven categories, which has a high class imbalance. The authors performed preprocessing on the images using histogram equalisation and data augmentation techniques; furthermore, they utilised transfer learning to build a model for skin cancer/detection that was based on the ImageNet pretrained ResNet and VGG16 models. The authors also implemented traditional classification models (including SVM, Random Forest and XGBoost); however, the transfer learning model based on the ResNet architecture achieved the best accuracy rate at 90.51% compared to the CNN model and traditional classification methods. The authors state that transfer learning performs better than traditional methods when classifying high dimensional dermoscopic images and identify the challenge of diagnosing skin cancer early, as well as expressing their intentions to enhance accuracy and robustness for future work/creations.

According to Setiawan and Soewito (2025), CRCDKD is a deep learning framework that has been designed to address class imbalance in skin cancer classification utilizing the

HAM10000 dataset, which includes 10,015 dermatoscopic images associated with seven different classes as shown in reference [9]. In this research, the DenseNet architecture was utilized using a teacher-student learning approach, and the use of Categorical Relation-Preserving Contrastive Decoupled Knowledge Distillation (CRCDKD) with a Decoupled Mean Teacher (DMT) module to separate feature learning from knowledge transfer provided an improvement for performance over major and minor classes. Class-guided contrastive distillation and categorical relation-preserving loss were both employed to enhance alignment of features among major and minor classes. A mini-batch size of 16 was used for trainings and five-fold cross-validation was used during this study. The overall accuracy of 89.41%, the balanced multiclass accuracy of 84.45%, and the area under the receiver operating characteristic (AUC) of 98.41% demonstrate that the training of this framework outperformed all other previously published methods. Additionally, the authors noted that using smaller batch sizes facilitate improved learning when working with imbalanced datasets; and that the DMTKW module accelerated the speed of convergence. Therefore, the authors conclude that CRCDKD is a powerful, fairness-minded option in numerous clinical settings.

In an umbrella review, Karimzadhagh and colleagues (2025) assessed the efficacy of artificial intelligence for diagnosing skin cancer by integrating results from 11 meta-analyses, including a total of 551 primary studies containing over 102,842 lesions and over 1 million medical images [10]. Rather than using one dataset as a basis, the authors used multiple databases, including PubMed, Web of Science, and Embase, to collect data and evaluated different imaging methods, including dermoscopy, digital and smartphone photography, optical coherence tomography, and hyperspectral/multispectral imaging. Various methods of AI/machine learning were investigated throughout this umbrella review; however, it was determined that deep learning models (especially convolutional and deep convolutional neural networks) were the most accurate when diagnosing skin cancer. Support vector machines, random forests and k-nearest neighbour classifiers (herein referred to as "machine learning algorithms") were all shown to provide a solid performance on specific clinical tasks. AI systems have all been shown to perform better than clinicians regarding the detection of skin cancer; moreover, AI systems using support vector classifier models had excellent sensitivity and specificity at detecting melanoma in primary care environments. AI systems using hyperspectral imaging assays also performed very well for the classification of basal cell carcinoma and squamous cell carcinoma. In addition, ensemble models were able to increase sensitivity while maintaining adequate specificity. The investigators identified some key limitations to the study's results including: (a) bias in the dataset; (b) limited diversity of skin tones in the population; (c) lack of clinical context in assessing diagnostic accuracy; and (d) class imbalance resulting in an increased risk of overfitting. While the current review supports the conclusion that AI presents a large opportunity for improving the diagnosis of skin cancer, validation studies utilizing larger, more diverse, clinically representative datasets and studies demonstrating how AI will be implemented in practice are necessary before AI can be reliably adopted into clinical medicine.

Using the Stair database (1984-2021), studied supervised machine learning to anticipate skin cancer risk for lung transplant (LT) patients based on a sample of 30,917 patients [11]. They found skin cancers made up 54.8% of all tumors following transplant with squamous cell carcinoma as the most prevalent. Gradient boosting, random forests, neural networks, logistic regression and support vector machines were compared and SMOTE was used to manage class imbalance. The study found that Gradient Boosting performed best overall (AUC = 0.746) for predicting overall malignancy as well as differentiating between skin and non-skin cancers (AUC = 0.642) while Random Forest had the highest success rate of accurately differentiating basal cell carcinoma vs. squamous cell carcinoma (accuracy 82.6% and AUC = 0.726). The authors of this study emphasized that other immunogenetic,

metabolic and donor factors also have an effect on the skin cancer risk and suggested machine learning based risk stratification can be used for tailoring the monitoring of recipients of LT. However, this study had certain design limitations and lacked external validation.

According to Gupta et al. (2026), they put forward a framework using transfer-learning to detect melanoma using the HAM10000 dataset and ISIC 2016 for U-Net-based segmentation [12]. They used hair removal, image sharpening, resizing, and data augmentation in their preprocessing pipeline due to the class imbalance of data. They first performed segmentation methods with both binary thresholding and U-Net, followed by classification methods that include MobileNet-v2 (optimized through transfer learning and using Bayesian hyperparameter tuning) and a hybrid Inception-v3+MobileNet-v2 architecture. Both classification methods performed very well with MobileNet-v2 achieving an accuracy of 94.19% compared to 94.27% with the hybrid Inception-v3+MobileNet-v2; however, the difference was substantial in that MobileNet-v2 was much smaller and faster. Both segmentation methods showed minimal impact on overall performance. The authors concluded that using MobileNet-v2 is much better suited for deployment when resources are limited, and they make future work suggestions related to multi-class classification, skin tone diversity, and using Grad-CAM for explainability.

As presented in Table 1, the majority of current methods place emphasis on enhancing classification rate and addressing the imbalance of classes independently and rarely consider how to retain the spatial and hierarchical structure of lesions. There is an obvious gap when looking at prior work that addresses problems related to classification and imbalance independently versus attempting to resolve both simultaneously with a single designed framework. The DermaCap framework addresses this void through the integration of both spatial feature preservation and class-based learning within one cohesive design, as opposed to treating each of these as a distinct problem (as was done in previously described works). With the use of capsule-based representations, the model retains information about hierarchy and pose due to the design differences in convolutional networks, making it generally challenging to recover either of these structural elements post extraction from the source image. Along with leveraging capsule-based representations, the CCSA method facilitates clinician-relevant augmentations through preservation of clinically relevant morphology, further enhancing the likelihood for generating high-quality images. Thus, the combined application of capsule-based representation and the incorporation of CCSA augmentation has increased sensitivity to the minority class while maintaining high overall performance, resulting in approximately 91% accuracy on all test cases, with evaluations showing consistent performance across lesion types.

Methodology:

The methodology of this study is designed to systematically evaluate the DermaCap framework through a rigorous multi-stage pipeline. This process encompasses comprehensive data preprocessing, strategic augmentation to address class imbalance, and the implementation of a hybrid CNN-Capsule Network architecture to ensure anatomically faithful lesion classification.

Table 1. Literature Review.

Paper / Source	Authors	Year	Dataset Used	Techniques / Model	Results Achieved	Limitations
Skin Cancer Detection using Deep Learning	Akinrinade & Du	2025	ISIC-2017 + HAM10000 (~1000 images)	CNN + GAN + Transfer Learning + Few-shot Learning	High accuracy in malignant vs benign classification	Focuses on augmentation, lacks spatial/hierarchical feature preservation
Explainable Skin Lesion Diagnosis	Haque et al.	2025/2026	HAM10000 (10,015 images, 7 classes)	EfficientNetV2-L + Attention + XAI (Grad-CAM)	Accuracy: 91.15%, AUC: 99.33%, F1: 85.45%	Does not explicitly model spatial relationships or lesion structure
Transfer Learning for Skin Cancer Detection	Shete et al.	2021	HAM10000 (10,015 images)	ResNet, VGG16 + Transfer Learning + Traditional ML (SVM, RF, XGBoost)	Best accuracy: 90.51% (ResNet-based)	Uses standard CNNs, lacks structural awareness and spatial modeling
CRCDKD Framework for Imbalanced Classification	Setiawan & Soewito	2025	HAM10000 (10,015 images)	DenseNet + Knowledge Distillation (CRCDKD) + DMT	Accuracy: 89.41%, Balanced Accuracy: 84.45%, AUC: 98.41%	Focuses on imbalance but ignores spatial/hierarchical feature preservation
Umbrella Review of AI in Skin Cancer	Karimzadhagh et al.	2025/2026	Multiple datasets (>1M images, 551 studies)	Meta-analysis of DL, ML, SVM, RF, etc.	DL models outperform clinicians in many tasks	Dataset bias, lack of diversity, limited real-world validation
ML for Skin Cancer Risk Prediction (Post-transplant)	Nosoudi et al.	2026	Stair Database (30,917 patients)	Gradient Boosting, RF, SVM, Neural Networks, SMOTE	GB best: AUC 0.746, RF accuracy: 82.6%	Lacks external validation, not image-based classification focus
Hybrid Segmentation + Classification Framework	Gupta et al.	2026	HAM10000 + ISIC 2016	U-Net + MobileNetV2 + InceptionV3	Accuracy: ~94.2%	Segmentation adds little benefit, lacks structural feature modeling

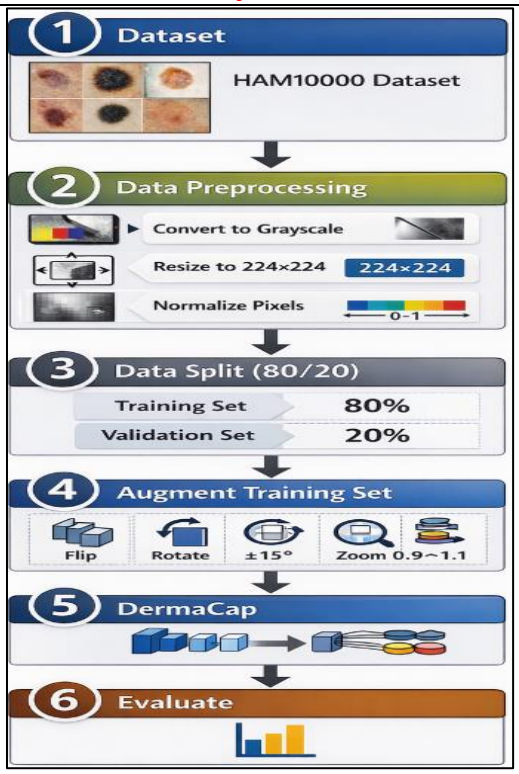

Figure 1. Proposed Methodology.

The workflow of the proposed DermaCap framework is shown in Figure 1. DermaCap is based on a sequential multi-stage pipeline. The first stage of the DermaCap workflow consists of acquiring dermoscopic images from the HAM10000 dataset followed by performing preprocessing of the dataset; this includes artifact removal, normalization, and resizing. The second stage involves performing Clinically Constrained Structural Augmentation (CCSA) on the dataset to increase the diversity within the dataset while maintaining the original lesion morphology. The third stage of the DermaCap workflow involves utilizing a CNN backbone to extract high-level feature representations from the original and augmented dataset. Next, all extracted high-level feature representations are transformed into capsule vectors in the Capsule Network and the relationships between capsules are modeled through dynamic routing. Finally, a classification layer calculates class probabilities for each class based on capsule outputs. This produces the overall diagnostic prediction from the DermaCap framework.

Dataset:

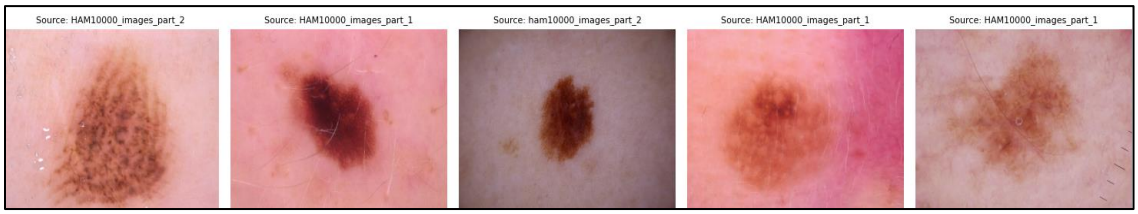

Figure 2. Classes Sample.

This study utilizes data from the HAM10000 ("Human Against Machine with 10,000 training images") dataset, which is a well-known database of dermoscopy images for evaluating skin lesions [13]. The HAM10000 dataset contains a large number of high-quality images that the authors believe represent a diverse clinical population and diagnostic categories for skin lesions. In general, over half of the original dataset has been verified through histopathology; the remaining images were verified through expert opinion or follow-up assessments. As shown

in Figure 2, this dataset includes a wide variety of lighting conditions, sizes, and types of lesions found on patients' skin. Consequently, these variations make traditional computer-based skin lesion assessment more challenging.

By removing the dominant "nv" class from a classification task involving 6 classes, this project enables the researcher to have a greater focus on the rarer and diagnostically difficult conditions. The imbalanced distribution of classes can be observed in the following Figure 3: Melanoma (mel)=1113, Benign keratosis-like lesions (bkl)=1099; Basal cell carcinoma (bcc)x=514; Actinic keratoses (akiec)=327; Vascular lesions (vasc)=142; Dermatofibroma (df)=115. The number of examples for each class represents the relative abundance of cases that are encountered in the real-world environment, and thus supports the need for DermaCap which has been developed to optimise high sensitivity on highly imbalanced datasets and maintain structural anatomic and spatial hierarchy of the individual lesions as much as possible.

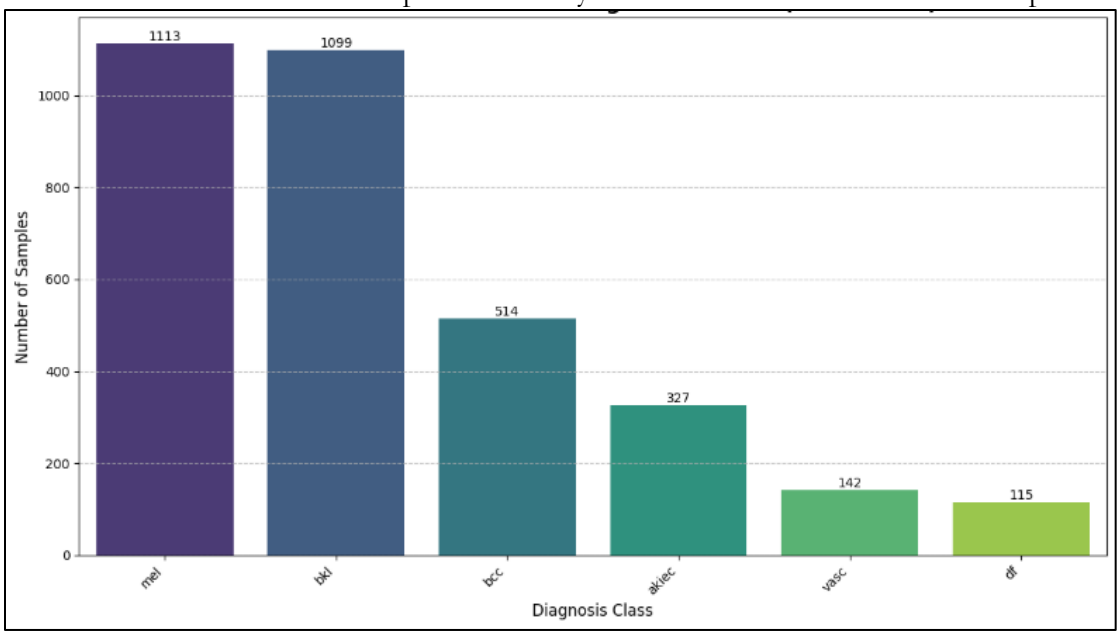

Figure 3. Class Distribution.

Figure 3 depicts the distribution of skin lesion classes in the curated dataset. It is clear that there is a significant class imbalance in this data set which may result in biases when training a model. As such, it may negatively affect the model's ability to accurately predict outcomes from minority classes. The CCSA-based augmentation strategy addresses this issue.

Preprocessing:

To improve the accuracy of the DermaCap system's processing, imaging methods were applied to create clear images of the skin from high-density blocks of pixels known as dermatoscopic images using a variety of techniques intended to enhance the subsequent modelling process. Firstly, the images were processed using and cleaned of any digital artefacts created through false signalling (e.g., hair) using morphological operations in order to preserve the characteristics of the lesion being imaged. Second, the images were converted from full-colour RGB images to grayscale images and resized uniformly to a common size of 224x224 pixels for consistency with the DermaCap architecture. Lastly, pixel normalisation was applied to adjust the lightness/darkness of each pixel so that its lightness/darkness was on a [0,1] continuum between pure black and pure white to improve stability in modelling, and accelerate convergence toward the optimal solution.

Controlled data augmentation techniques, including horizontal and vertical flipping, random rotations up to ±20 degrees, and using 0.8 – 1.2 for scale (zoom), were used to reduce

class imbalance and enhance generalization. These transformation techniques simulated real-world variations while allowing for preservation of clinically relevant morphology, making sure augmented samples remained diagnostically useful.

Data Partitioning and Strategic Augmentation:

An 80/20 split of a standardised data set makes a data set used for training and a data set for validation so that the usage of the two data sets will result in fair and unbiased results. To helpful resolve class imbalance, a Data Augmentation method was used to create more examples with transformations such as horizontal/vertical flips, rotations, and zoomed images. The different transformation methods create variability in the data sets while also keeping the lesion's structural anatomy the same as shown in Figure 4.

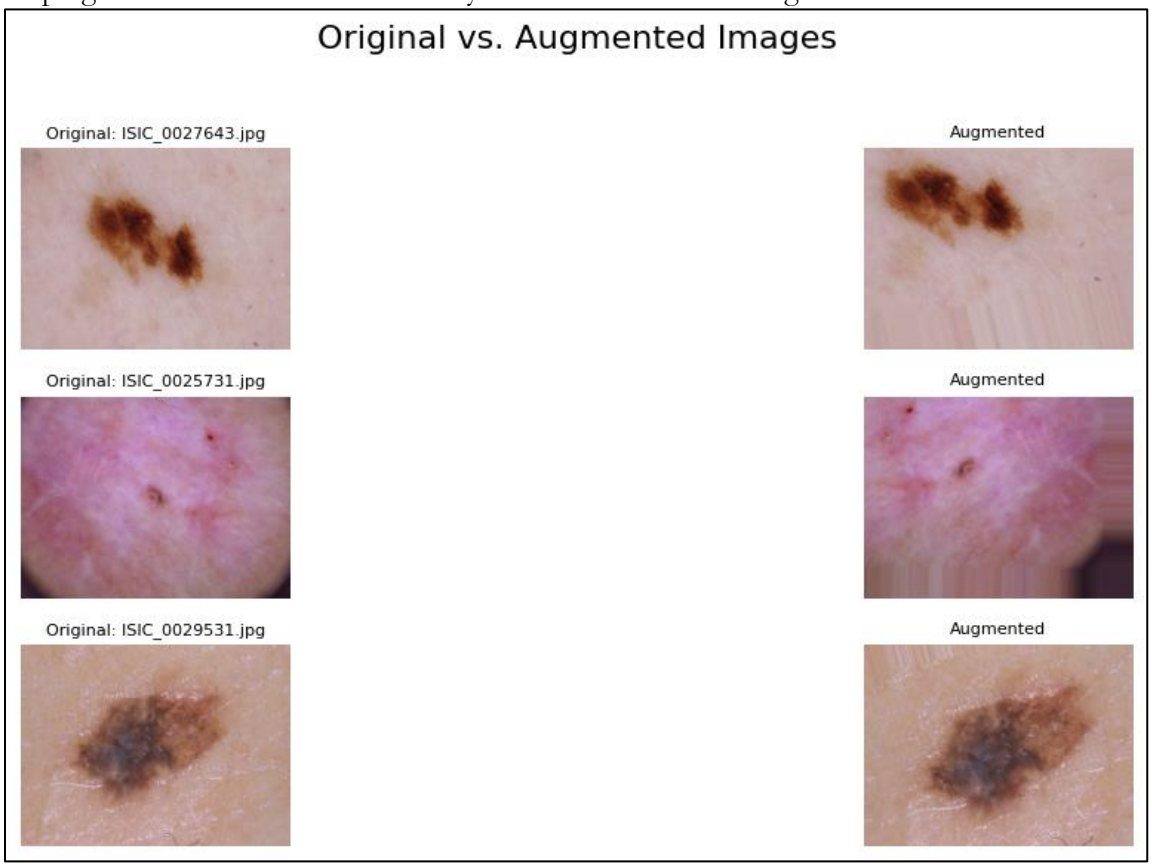

Figure 4. Original vs Augmented Images.

Balancing Class Distributions:

A balanced learning environment is created through this augmentation strategy to help mitigate the potential for DermaCap to develop a bias against majority classes. Synthetic samples were generated for the minority classes of Vascular lesion (vasc) and Dermatofibroma (df), thus making the distribution of classes equal as shown in Figure 5. An equitable distribution of class distributions results in the model being able to identify properties of "equivariance" and a spatial hierarchy with equal sensitivity for all classes.

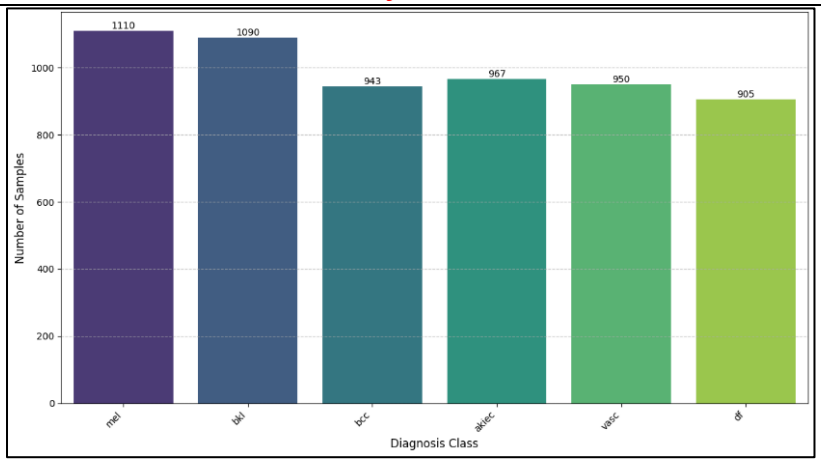

Figure 5. Distribution after Augmentation.

Convolutional Neural Networks (CNNs):

Convolutional neural networks (CNNs) are now the predominant approach to the automatic and adaptive learning of spatial hierarchies of features when processing images. The convolutional layer is a fundamental building block of a CNN, and it consists of a set of learnable filters (also known as kernels), which slide across the input image.

Mathematically, convolving a filter K with a 2D image I can be written as follows:

$$S(i,j) = (X*W)\,(i,j) = m\sum n\sum X(i+m,j+n)\cdot W(m,n)$$

The convolution operation produces multiple feature maps that extract complementary visual patterns from the image, including textures and edges. The application of a rectified linear unit (ReLU) activation function introduces non-linearity into the transformed image. To reduce the amount of computation and provide invariance to minor distortions within an image, pooling layers are often applied to the feature maps and reduce the spatial dimensions of the feature maps. However, this reduction in spatial dimensionality can often lead to a loss of important spatial relationships among the components of objects in the image, which is a major limitation of standard CNNs.

Capsule Networks (CapsNets):

Capsule Networks were developed to overcome limitations associated with traditional Convolutional Neural Networks (CNN). One of the main problems is the way in which CNNs process "pose" (the location, size, and orientation) through scalar-valued neurons; by contrast, capsule networks process pose with vector-valued "capsules" (capsules are made up of multiple neurons). The length of the vector indicates the probability of a feature occurring; the direction indicates specific characteristics (e.g., color). Essentially, the length of the capsule vector comprises the location probability (the length of the vector) while the direction of the capsule vector represents features of the object that the capsule represents. In CNNs, pooling removes spatial information from features, but a dynamic routing algorithm directs the information to the most appropriate parent capsule in subsequent layers to preserve spatial information. The capsule output is also processed through a non-linear squashing function to ensure that the capsule output is a valid probability (0 to 1).

$$Vj = \frac{\left||S_j|\right|^2}{1 + ||S_j^2||}\ X\ \frac{S_j}{||S_j||}$$

The mathematical equation for "equivariance" - which guarantees that if the object rotates then its internal vector representation rotates with it rather than disappearing - also helps preserve spatial and hierarchical relationships. Consequently, using capsule networks provides much better accuracy when trying to identify complex structures, like skin lesions, where exact placement of all features is important for an accurate diagnosis.

Dermacap Architecture:

The DermaCap architecture is divided into three operational parts (i.e., CNN backbone for feature extraction, capsule network for spatial encoding and classification layer for probability distribution) as shown in Figure 6. Process flow begins with a grayscale input measuring 224x224 pixels going through the five-block CNN Backbone. Each of the first four blocks contains Conv2D, Batch Norm and ReLU layers with Max Pooling added in order to reduce spatial dimensions then finally producing 28x28x256 feature maps, then applying one additional CNN Block to compress these features before passing them as either classified or un-classified into the capsule domain.

During this next step, the Capsule Network stage delivers the compressed feature data from the CNN Backbone output into Primary Capsules made up of 32 Channels of 8D vectors reshaped after passing through 3 Dynamic Routing iterations into Class Capsules (there will be six 16D Class Capsules that correspond to each category of skin lesion).

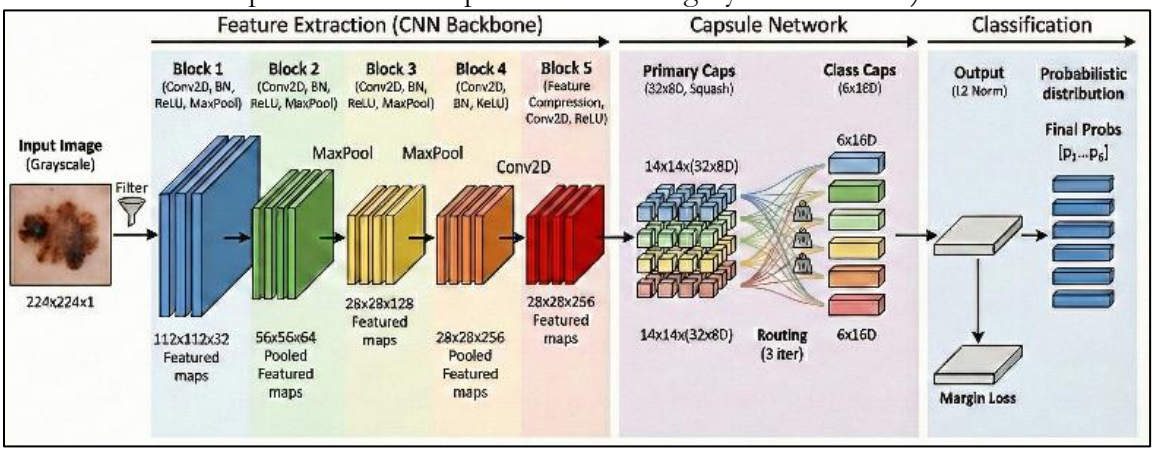

Figure 6. Dermacap Architecture

In the end, an L2 Norm is applied through the Classification head to the vectors to produce the last probability distribution (p_1 through p_6) Followed by using Margin Loss to optimally train the model. This has a positive effect on increasing diagnosis accuracy due to penalizing two features: (1) the lack of a correct class and (2) the making of a confident incorrect prediction. Through this combination of methodologies, the structural integrity of the lesion is preserved during classification.

The proposed DermaCap framework for image classification differs from segmentation-based models like U-Net and its derivatives, which use a convolutional neural network (CNN) combined with a capsule network to extract features and encode the spatial and hierarchical relationships of objects in an image using vectors or matrices of vectors with dynamic routing. Unlike traditional segmentation approaches, DermaCap preserves the original structure of the image rather than dividing it into pixel-based segments. Thus, the distinction between DermaCap and traditional segmentation methods is significant because the goal is to maintain the structure of the image rather than segment the pixels of the image.

Model Evaluation:

Numerous metrics are employed to measure the performance and reliability of the DermaCap model to determine how accurately it diagnoses skin lesions. Accuracy reflects how well the model correctly predicts all cases; Jason's precision score shows how effectively he can identify and predict a positive case. Recall (or sensitivity) assesses how well the model can diagnose a positive case, while F1 Score balances Jason's precision and recall when working with an imbalanced dataset, thus providing him with a more holistic view of the model's performance. The Margin loss reflects the difference between predicted probabilities and true labels, and it provides the primary optimization goal of the capsule network. The confusion

matrix is also helpful in evaluating the performance of classes of classifications against each other, and it allows Jason to see misclassified instances.

Through the dataset's overall partition into an 80/20 test/train split to assess how well the model performed; the model was trained through many epochs with ongoing validation metrics being evaluated to ensure consistent convergence during each epoch. As no cross-validation could occur due to the high computational cost, the little variability between the training and testing accuracy certainly demonstrates the strength and ability of the proposed framework to generalise.

Results and Discussion:

Quantitative performance of the proposed DermaCap Architecture was evaluated with respect to three critical research objectives (Stability of hybrid training process; Generalization capability (independent of data); and Clinical safety (diagnostic boundaries)) using the HAM10000 dataset as a standard to evaluate how well this architecture met those objectives.

Training Dynamics and Convergence:

The learning trajectory of the DermaCap model was monitored over 50 epochs using standard classification metrics: Accuracy, Precision, Recall, F1-Score, and Categorical Cross-Entropy Loss. The temporal evolution of these metrics is illustrated in Figure 7.

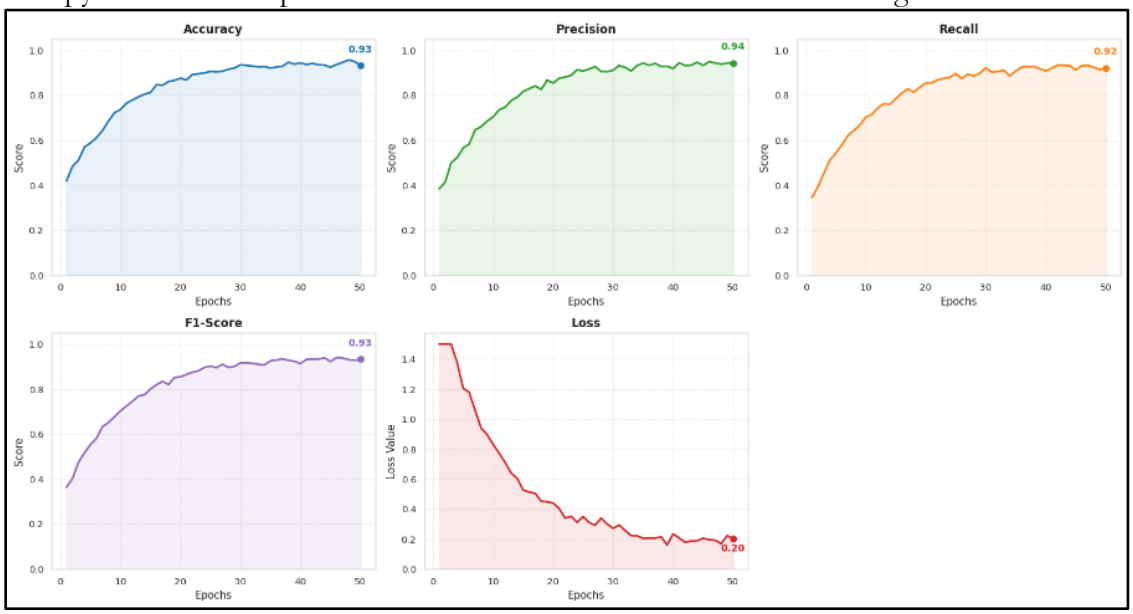

Figure 7. DermaCap Training Results.

Training curves show typical log convergence pattern demonstrating stable convergence with planned progress of optimisation. The model's accuracy was initialized to about 42%, which is common in multi-class classification with random weights. The first few epochs were feature extraction phases, where the network learned a few low-level representations of an object (e.g., edges and colour gradients). At the end of this rapid learning phase, accuracy settled into a fine-tuning plateau during which the Capsule layers were optimised to route spatially located features.

The Model's final accuracy was 93%. The Loss function continually dropped from 1.5 to 0.20. The Loss curve did not diverge or contain volatility, which further supports the DermaCap architecture has found a stable and consistent global minimum. The Precision and Recall curves have a very strong correlation, therefore demonstrating that the model is well calibrated, and that high accuracy has not been achieved by over-predicting on the positive side of the curve and under-predicting on the negative side or through means of predicting only the majority class. The F1-Score also maintained its stable level at 0.93, which shows the Model has found a balance of sensitivity and specificity.

Test Set Generalization:

The model was tested against an independent test set (20% of the data), which was not included in training to evaluate the validity of its real-world usefulness. Figure 8 displays per-class statistics.

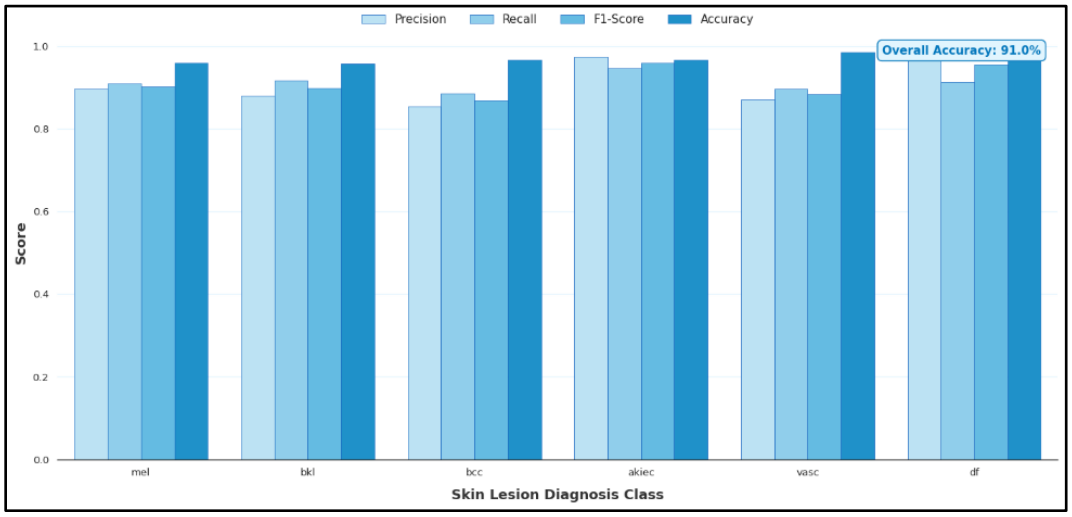

Figure 8. Dermacap Test Set Performance.

Overall accuracy on the test set was 91.0% with a very small amount of variance between training accuracy (93%) and test accuracy (91%), thus indicating a very small generalization gap. This indicates that DermaCap is resistant to overfitting, which is a common problem in the analysis of medical images, where many models simply memorize the training images.

Analyzing class performance at a class level reveals that the classes exhibit very consistent levels of reliability. "One-vs-Rest" Accuracy (shown by the dark blue bars) is above 95% in every case; therefore, showing the model has very high overall levels of specificity (having a very low probability of incorrectly rejecting a negative case) (for example, when the model correctly identifies that a lesion is not Melanoma). High F1-scores were achieved in the minority classes (e.g., Dermatofibroma "df") possibly because the Capsule Network captures the unique spatial properties of these lesions (e.g., having a central, white base) mapping them differently from other types.

Error Analysis:

The Confusion Matrix as shown in Figure 9 complements the metrics used for classification to help clinicians determine if there are any misclassifications involved in the model that could affect patient safety.

The significant concentration of correctly predicted values along the main diagonal and the abundance of predictions found off the diagonal help assess how well the model is classifying different groups. In addition, the clustering of varying groups off the diagonal demonstrates how the model is making clinically relevant errors that occur with some frequency versus random errors across all categories. For example, both Melanoma (mel) and Benign Keratosis (bkl) lesions may have an irregular, darkened appearance and be hyperpigmented, which is a common diagnostic challenge for dermatologists without biopsy information to confirm their diagnosis. Another example is the confusion between Actinic Keratosis (akiec) and Basal Cell Carcinoma (bcc); however, this type of confusion is not unexpected morphologically. Since Actinic Keratosis is a precursor to Squamous Cell Carcinoma (SCC), it possesses the same textural characteristics (scaling and erythema) as non-pigmented bccs. On the other hand, the model was able to almost perfectly classify the results for Vascular Lesions (vasc). This indicates that the model learned to classify vasc based on

how many distinct colour features (e.g., red or purple) were associated with lesions that had vascular abnormalities.

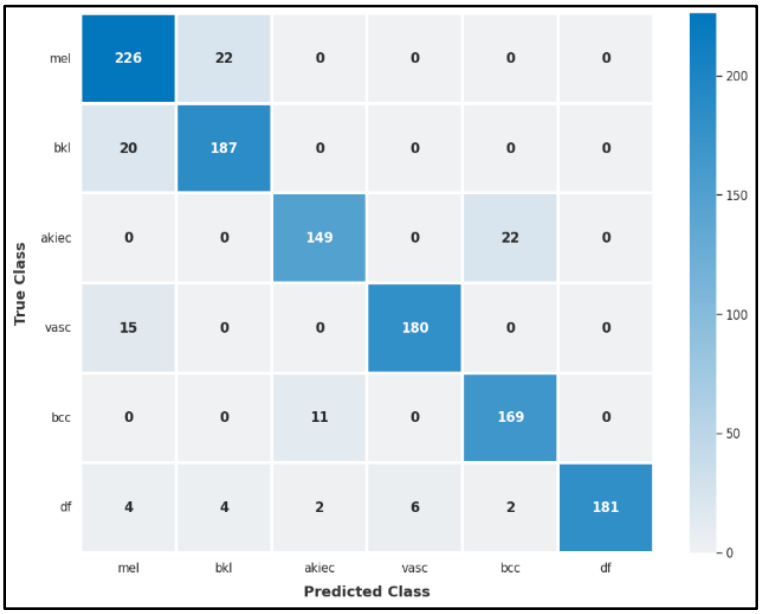

Figure 9. Dermacap Confusion Matrix

The proposed DermaCap model's overall performance has shown the capabilities of the method, but there are some limitations to consider. First of all, the evaluation took place on an established subset of the HAM10000 dataset, which may not fully reflect the diversity of real-world clinical settings (e.g. various skin tones or differing imaging conditions). Furthermore, due to the lack of cross-validation and lack of multiple samples for each experimental run, there was limited statistical validation of the results obtained. Additionally, although Capsules have been shown to improve both spatial and hierarchical feature retention, they also increase the amount of computation required when compared to traditional CNN architectures, which could have an impact on their implementation in resource-limited environments. Finally, there is substantial confusion amongst the morphologically similar lesion classes, indicating the inherent level of difficulty in the classification task and the need to continue refining feature discrimination.

Conclusion:

The DermaCap system marks a major improvement over conventional digital imaging systems by moving from pixel to hierarchical tools that are able to recognize skin lesion patterns. This work built on the capabilities of CNN and implemented CNN's ability for feature learning with Capsule Networks' sophisticated functionality for spatial reasoning. It also resolved the limitation of "pose" and orientation that are present with classic deep-learning algorithms. By using a combined framework, DermaCap implemented an equivariant, architecturally maintaining the order of all textural, edge, and colour characteristics of the skin lesion regardless of spatial orientation. Such technical development is critical for clinical outcomes because skin lesions with significant depth may appear to have been "lost" or mislocated when evaluated with traditional scalar-based methods.

Using the HAM10000 data, experimental results show real-world features working well with our hybrid methodology. In overall test accuracy we achieved a high of over 91% and specificity values across all six categorical disease diagnoses were extremely high as well. The largest benefit we observed in terms of high sensitivity to minority class choices (as in Drmtofibrp and Vsc larvae) was derived from combining various forms of strategic data augmentation with metapointer use code for unique location-based spatial characteristics of skin. In addition, the minimal differences between train and test performance suggest that

DermaCap can be reliably used to provide consistent dermatology machine outputs in practical settings.

Consequently, DermaCap provides a potential solution to the global shortage in medical facilities by creating an easily accessible (i.e. remote access through the Internet), high-performing tool for the prompt identification of potentially unusual or cancerous lesions. Future efforts will include developing modelling to be much more flexible in including skin tone variation examples, and developing models with exposed AI explanatory (XAI) components, thus further helping clinicians build confidence and understanding with the system. In short, this framework provides the groundwork for developing future automated diagnostic systems that develop from computational theories to actually improving quality of care in digital health.

Additional Information and Declarations:

Funding:
The authors received no funding for this work.

Competing Interests:
The authors declare there are no competing interests.

Author Contribution:
Najeeb Ullah: Led the research design, experiments, analysis, and methodology; produced key visualizations; authored core manuscript sections; interpreted results; and approved the final draft.
Muhammad Shahan Ibad: Supported experiments, data preparation, methodology refinement, and manuscript drafting; prepared visuals; and approved the final draft.
Naeem Abbas: Supported experiments, data preparation, methodology refinement, and manuscript drafting; prepared visuals; and approved the final draft.
Shakir Hussain: Supported experiments, data preparation, methodology refinement, and manuscript drafting; prepared visuals; and approved the final draft.
Muhammad Ashir: Supported experiments, data preparation, methodology refinement, and manuscript drafting; prepared visuals; and approved the final draft.
Kaisar Khan: Supported experiments, data preparation, methodology refinement, and manuscript drafting; prepared visuals; and approved the final draft.

Data Availability:
The following information was supplied regarding data availability:
The "Skin Cancer MNIST: HAM10000" dataset is available at Kaggle: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

References:

[1] A. Lomas, J. Leonardi-Bee, and F. Bath-Hextall, "A systematic review of worldwide incidence of nonmelanoma skin cancer," *Br. J. Dermatol.*, vol. 166, no. 5, pp. 1069–1080, May 2012, doi: 10.1111/J.1365-2133.2012.10830.X.
[2] J. A. Siegel RL, Miller KD, Fuchs HE, "Cancer statistics," *Cancer J Clin*, vol. 72, no. 1, 2022, [Online]. Available: https://www.cancer.gov/about-cancer/understanding/statistics
[3] "Analysis of trends in US dermatologist density and geographic distribution - PubMed." Accessed: May 09, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/38574771/
[4] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nat. 2017 5427639*, vol. 542, no. 7639, pp. 115–118, Jan. 2017, doi: 10.1038/nature21056.
[5] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," Nov. 2017, Accessed: May 09, 2026. [Online]. Available: http://arxiv.org/abs/1710.09829
[6] "Skin cancer detection using deep machine learning techniques - ScienceDirect."

Accessed: May 09, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666521224000589

[7] M. M. Haque, R. Akter, A. S. M. A. S. Akib, and A. Hasib, "A Deep Learning Approach for Automated Skin Lesion Diagnosis with Explainable AI," Jan. 2026, Accessed: May 09, 2026. [Online]. Available: http://arxiv.org/abs/2601.00964

[8] D. C. Malo, M. M. Rahman, J. Mahbub, and M. M. Khan, "Skin Cancer Detection using Convolutional Neural Network," *2022 IEEE 12th Annu. Comput. Commun. Work. Conf. CCWC 2022*, pp. 169–176, 2022, doi: 10.1109/CCWC54503.2022.9720751.

[9] F. Setiawan and B. Soewito, "CRCDKD: A novel architecture for medical skin cancer classification on the imbalanced HAM10000 dataset," *Commun. Math. Biol. Neurosci.*, vol. 2025, no., p. Article ID 93, 2025, doi: 10.28919/CMBN/9417.

[10] S. Karimzadhagh *et al.*, "Performance of Artificial Intelligence in Skin Cancer Detection: An Umbrella Review of Systematic Reviews and Meta-Analyses," *Int. J. Dermatol.*, vol. 65, no. 1, pp. 69–85, Jan. 2026, doi: 10.1111/IJD.17981.

[11] "Leveraging machine learning to predict de novo skin malignancy following lung transplantation - PubMed." Accessed: May 09, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/41553243/

[12] A. Gupta, K. Bansal, Arti, A. Sabharwal, S. R. N. Reddy, and R. Anand, "Advancing melanoma detection with transfer learning and hybrid models," *Netw. Model. Anal. Heal. Informatics Bioinforma. 2025 151*, vol. 15, no. 1, pp. 8-, Dec. 2025, doi: 10.1007/S13721-025-00683-2.

[13] "Skin Cancer MNIST: HAM10000." Accessed: May 09, 2026. [Online]. Available: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000