Advances in Artificial Intelligence and Machine Learning; Research (Ahead of Print) Received 06-01-2026; Accepted 14-04-2026; Published online 21-04-2026

Deep Learning Framework Using DenseNet-169 Based-features Derived from Thermal Images for Dermatological Disease Identification

Ashu Sharma anshumca2@gmail.com
Department of Computer Science & IT
University of Jammu
India

Tinny Sawhney tinnysawhney@jammuuniversity.ac.in
Department of Electronics
University of Jammu
India

Pawanesh Abrol pawaneshabrol@gmail.com
Department of Computer Science & IT
University of Jammu
India

Parveen Kumar Lehana pklehana@gmail.com
Department of Electronics
University of Jammu
India

Corresponding Author: Parveen Kumar Lehana

Abstract

Accurate and non-invasive diagnosis of dermatological diseases remains a significant clinical challenge due to visual similarity among different skin diseases and variability due to imaging environment. Thermal imaging has emerged as a promising diagnostic technique as it captures physiological variations related to inflammation and vascular abnormalities, independent of illumination conditions. In this study, a deep learning–based framework is proposed for dermatological disease identification using thermal images. The proposed framework employs two strategies, one DenseNet-169 based and another VGG19 for feature extraction, both followed by a classifier consisting of either convolutional neural network (CNN) or bidirectional long short-term memory (Bi-LSTM). The system has pre-processing unit, transfer learning-based deep feature extraction, and classification with CNN for investigating the importance of spatial learning and Bi-LSTM for the sequential learning. Six general categories of dermatological diseases, namely Acne, Allergy, Melasma, Milia, Psoriasis and Vitiligo, were used in the evaluations of experiments. The system performance is measured by common measures like accuracy, precision, recall, F1-score, confusion matrix, and ROC–AUC plots. We can see that CNN classifier outperforming Bi-LSTM classifier in both feature extractors. With DenseNet-169 features, the CNN classifier reaches almost

Citation: Ashu Sharma, et al. Deep Learning Framework Using DenseNet-169 Based-features Derived from Thermal Images for Dermatological Disease Identification. **Advances in Artificial Intelligence and Machine Learning**. 2026. (Ahead of Print). https://dx.doi.org/10.54364/AAIML.2026.62297

perfect performance with an accuracy of 99.68%, which indicates a stable training behaviour, fast convergence, and good discriminative capacity. In comparison, the DenseNet-169 and Bi-LSTM model show significantly worse performance overall, reaching an accuracy of only 73.05% with a low generalization capability. In the case of VGG19 as feature extractor, CNN based classifier achieves a good accuracy of 99.35% while having robust ROC characteristics, while VGG19 and Bi-LSTM configuration achieved a relatively lower performance as 89.61% since they confound the data between classes. In general, the results provide evidence that convolutional classifiers perform better in using the spatially diverse representations that DenseNet-169 and VGG19 generated. The proposed framework could be a scalable and reliable solution for automated thermal image based dermatological diagnosis, with a strong promise for facilitating real world clinical decision support.

Keywords: Dermatological image classification, Deep learning, Thermal imaging, DenseNet-169, VGG19, CNN vs Bi-LSTM, Automated skin disease diagnosis.

1. INTRODUCTION

Precise diagnosis of skin disorders is a major challenge because manual interpretation of large number of skin conditions is cumbersome if not impossible. Many diseases share similar characteristics and hence hamper exact prediction even by experienced dermatologists. The traditional methods of diagnosis like visual inspection paired with dermatoscopy, typically lack objectivity due to inter-observer variability. Biopsy-based methods may provide better prediction, being invasive; restrict their mass testing and early detection. The constraint has resulted in the use of AI (Artificial Intelligence) to automate the diagnostic process where an objective, accurate, early diagnosis may be provided through image-based diagnostic devices [1, 2]. Recently, advanced medical imaging and AI including deep learning technologies have enhanced diagnostic performance capabilities across the whole healthcare spectrum [3, 4]. However, most existing studies rely primarily on visible-spectrum imaging, which may fail to capture underlying physiological and pathological changes that are not apparent to the naked eye. Thermal imaging is thus emerging as a new alternative of dermatological analysis and is a promising technique for overcoming some of the limitations. Thermal imaging records infrared rays radiating from the skin, and thermal patterns correlate with physiologically relevant conditions like skin inflammation, perfusion abnormalities, and hyperactive metabolism, implicating some underlying dermatological disorders. Thermal imaging is contactless and radiation-free and is thus appropriate to patients of all ages. Furthermore, it does not cause discomfort, making it a good choice for frequent monitoring and prompt diagnosis [5, 6]. And thermal imaging has recently been taking up in a lot of medical fields such as breast cancer screening, obesity detection and other applications that demand constant physiological status [7, 8]. Thermal dermatological analysis, however, faces unique challenges. Thermal images often show low dynamic range and mild inter-class temperature variations, so discriminative feature extraction methods capable of modelling mild thermal gradients are required. Earlier studies in this area have been few. For example, a simple CNN combined with handcrafted features was compared with a pretrained model in [9], but the system was extremely sensitive to noise and heterogeneity of the dataset, and was restricted to a limited number of dermatological conditions. Likewise, handcrafted texture-based methods, including entropy and GLCM based frameworks, have shown discriminative ability in colour dermatological images, but are not scalable and robust for complex multi-class

scenarios [10]. These limitations highlight the need for modality-aware deep learning strategies tailored to thermal dermatological data.

Deep neural networks have no manual feature design input for learning rich representations with both low-level patterns and high-level semantic information. The shift to deep learning has achieved immense performance advances in numerous medical imaging applications for skin disease classification [1, 11]. Convolutional neural networks (CNNs) are among the most frequently applied deep learning architectures and have demonstrated potential for image-based disease classification tasks due to their ability to preserve spatial hierarchies and exploit local correlations. CNNs, as a class of deep learning algorithms, have been key drivers in the field of machine learning. They are well established, being of practical use in agriculture, aerial image classification, biometric applications, large-scale geospatial segmentation, and satellite-based photovoltaic power plant detection through DeepLabV3, ResNet and hybrid CNN SVM frameworks [12–19].

CNN based approach has already become popular for the skin disease and skin cancer diagnosis system and it has achieved good discriminative and robustness in various datasets [3, 20]. This robustness in dermatological imaging is the intrinsic ability to obtain the spatial and textural patterns, which lie at the base of predicting skin diseases. A few methods related to transfer learning have been recently introduced to enhance the performance in instances with limited annotated medical data. Because of the large datasets trained on pre-trained deep systems, transfer learning is suitable for adaptation as feature extractors for a range of domains. Since then, dermatological image processing has moved from manually trained feature extraction to CNN-based transfer learning architectures [21]. Major architectures are architectures such as original versions of AlexNet and Inception [22], further architectures as VGG and ResNet [22], and lightweight architectures like MobileNet and ShuffleNet for resource-poor applications [23]. These are often pretrained with ImageNet, and then fine-tuned to the target tasks for hierarchical features for better classification with much higher classification performance [24].

The role of pretrained machine learning models for classification of skin diseases and cancer has recently been effective. Furthermore, InceptionV3, ResNet50, and DenseNet201 substitute the initial classification layers with task-specific fully connected layers to enable multi-class prediction [25]. Other approaches utilize adversarial training based on the fast gradient sign method (FGSM) framework to improve model robustness such as VGG16, VGG19, DenseNet101, and ResNet101 [26]. Existing comparative studies have been implemented on EfficientNet, SENet, and ResNet against various training and data augmentation strategies, indicating that optimization of pre-processing and sampling techniques has a clear impact on diagnostic performance [27, 28]. Custom CNN architectures to classify skin diseases have also been investigated, but not frequently [29–31].

VGG19 has been shown as an effective feature extractor for dermatological image analysis, and other deep architectures have provided satisfactory results (DenseNet-based feature extraction systems) [1, 2, 32]. The regular 3×3 convolutional grid of VGG models allows for precise spatial details to be modeled and serves as a strong baseline. Residual learning models, such as those introduced by ResNet architecture, have shown good results for skin disease classification, due to improving gradient flow and being more suitable for deeper network training [33]. In contrast, dense connectivity mechanisms create direct inter-layer connections which make feature reuse easier, gradient propagation stronger, and representation learning more efficient [34]. DenseNet-169 has been reported for parameter efficiency and reduced overfitting with application in medical

image classification problems [35]. As thermal dermatological images exhibit subtle inter-class temperature differences and low-contrast physiological variations, dense connectivity is proposed as advantageous for sustaining and propagating weak thermal signals. Moreover, recent efforts underline the need for explainable and tuned AI structures for medical diagnostics [12, 36, 37], with interpretation and generalization posing important challenges to clinical AI [38, 39].

The present study proposes a thermal image based multi class dermatological disease classification framework that systematically investigates architecture suitability and representation of learning strategies for thermal images. DenseNet-169 is employed as the primary deep feature extractor due to its dense connectivity mechanism, which facilitates enhanced feature representation and preservation of subtle thermal gradients as reported in [40]. For comparison, VGG19 is also investigated as a uniform convolutional baseline to evaluate whether dense feature connectivity provides measurable advantages in thermal dermatology. Beyond conventional fully connected classification layers, the proposed research introduces a hybrid spatial sequential learning paradigm by feeding extracted deep features into both CNN and Bidirectional Long Short-Term Memory (Bi-LSTM) classifiers. While CNN-based classifiers inherently align with spatial representations, Bi-LSTM networks enable modelling of contextual dependencies across structured feature embeddings. The comparative evaluation will help in assessing the discriminative capabilities of sequential modelling in thermal feature representations. Clinically, the framework will provide a complementary, non-invasive decision-support tool rather than a replacement for dermoscopy or histopathology. Thermal AI-based analysis may support early-stage screening, inflammation monitoring, and scalable deployment due to its non-invasive and radiation-free characteristics. Thus, this work contributes both methodological insight into representation learning for low-contrast physiological imaging and translational value for scalable dermatological diagnostics. The detailed methodology is presented in Section 2, followed by results and discussion in Section 3. Section 4 focuses on conclusions drawn from the investigation.

2. METHODOLOGY

The proposed framework shown in **FIGURE 1** explores the effectiveness of deep feature-based identification of dermatological disease using thermal images. First, the input images are pre-processed to minimize heterogeneity and better identify diagnostic patterns. To extract high-level, discriminative feature representations, pretrained DenseNet-169 and VGG19 models have been used. The extracted features from the pretrained models are classified at a later stage using two different methods, a one-dimensional CNN to capture local spatial relationships within the vector of features and a Bi-LSTM to model the sequential dependencies between different feature dimensions. The proposed framework is evaluated on six dermatological classes using standard performance metrics for reasonable comparison of CNN and Bi-LSTM based classification performance.

A self-collected dataset comprising 3,079 thermal images was used in this study, covering six dermatological diseases: Acne, Allergy, Melasma, Milia, Psoriasis, and Vitiligo. The dataset was split into training, validation, and testing subsets using a 70:20:10 ratio. During splitting it was assured that every class has enough (several hundred) images for training, validation, and testing. The thermal images represent independent diagnostic acquisitions rather than sequential frames from a single imaging session. Because thermal imaging presents only physiological temperature distributions, not identification-linked structural cues, subject memorization bias is inherently min-

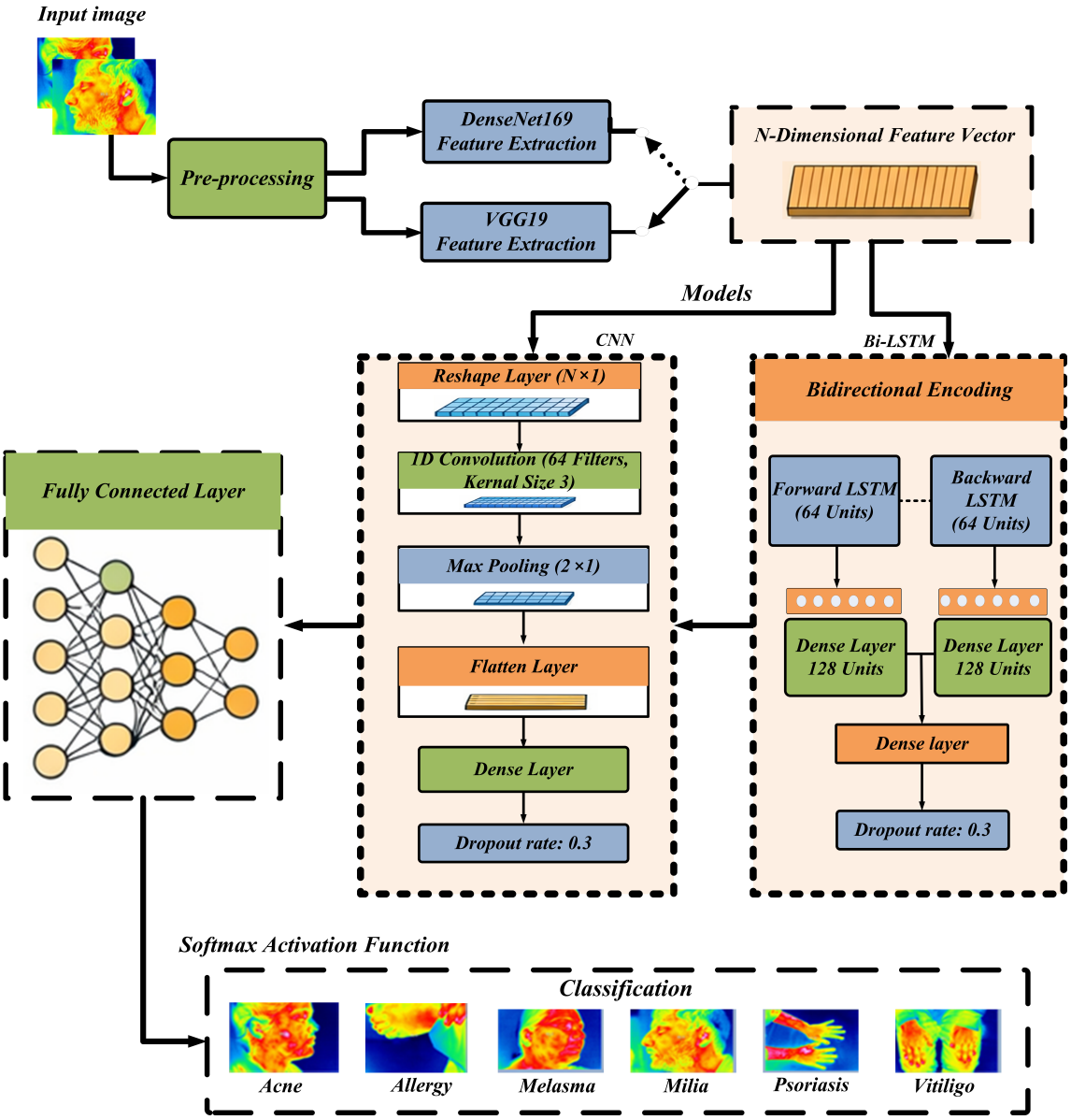

Figure 1: Proposed deep learning–based methodology for multi-class dermatological image classification using DenseNet-169 and VGG19 as feature extractors and CNN, Bi-LSTM as classifiers.

imized. All the thermal images were obtained using a Fluke Ti480 PRO infrared thermal camera with thermal sensitivity, radiometric accuracy, and suitability for clinical-grade surface temperature assessment. They were taken in a controlled indoor space to minimize the effects of ambient temperature fluctuations, airflow disturbances, humidity variations, and external heat sources. For all samples, a uniform imaging distance of about 1 meter was held in the photography to ensure

that the spatial resolution was consistent and thermal measurement would be consistent. To further minimize the effect of environmental variation and background thermal influence, the imaging was implemented under relatively constant room environment at about 71°F (22°C), where radiometric metadata was checked using Fluke SmartView software. This temperature falls within the recommended range (22–25°C) for dermatological thermal imaging, ensuring physiological stability without inducing artificial vasoconstriction or vasodilation effects. Camera conditions, including emissivity parameters of the camera, were kept consistent throughout the acquisition procedure to guarantee uniform temperature calibration.

The thermal images were analysed and exported by SmartView software to allow thermal visualization with increased precision, radiometric data management, and standardized image extraction for computing purposes. The designed controlled acquisition protocol was introduced to overcome the frequent problems related to thermal imaging sensitivity to environmental temperature variations, low spatial contrast. With standardization of capture conditions and uniform SmartView software processing, the dataset makes the thermal patterns extracted in this work mainly represent physiologically relevant dermatological properties; and environmental noise is avoided. Following acquisition, all images were resized to meet the input requirements of the pretrained deep learning models employed in the framework. Intensity normalization and basic noise suppression are applied to reduce sensor-related variations and enhance diagnostically relevant thermal patterns. This ensures a consistent input representation for effective deep feature extraction and classification. After pre-processing, deep feature extraction is carried out using DenseNet-169 and VGG19. From the last dense block of DenseNet-169, features are extracted and global average pooling is applied to obtain a compact 1664-dimensional feature representation for each image. VGG19 is also used to complement feature extraction with its deep convolutional structure that follows a sequence. The final convolutional block features are refined by global pooling to a fixed-length 512-dimensional feature vector per image. The high-level, discriminative and compact features flattened and are fed to classifier.

In CNN, the N-dimensional feature vector is first reshaped to make it compatible with one-dimensional convolution operations. A 1D convolutional layer with 64 filters and a kernel of size 3 x 3 to capture local correlations and interactions between neighbouring feature elements. This is followed by a Max Pooling layer with a pool size of 2 which reduced dimensionality, retaining prominent discriminative features. The resulting feature maps are flattened and passed through dense layers with dropout regularization to learn high-level discriminative patterns while minimizing overfitting. This CNN branch focuses on learning spatial correlations present in the deep feature vectors. Finally, in the output layer, a softmax activation function is used to generate normalized class probabilities for multi-class dermatological disease classification.

Alongside the CNN classifier, a Bi-LSTM classifier is applied to study the sequential learning of the extracted features, the deep feature vectors are characterized as ordered sequences, enabling the Bi-LSTM network to capture contextual dependencies across feature dimensions. The bidirectional framework is designed to model not only the forward and backward paths for the feature sequence, thereby capturing long range dependencies that spatial models may not capture sufficiently. Dense layers are applied after bidirectional encoding to enhance the learned temporal representations. The Bi-LSTM output is then subjected to fully connected layers with dropout regularization to improve model stability and alleviate overfitting. In the last layer, a softmax activation function is used to implement dermatological disease identification. This architecture allows to contrast the

performance of spatial vs sequential convolutional learning in relation to deep features obtained using VGG19 and DenseNet-169. The last output layer generates a model to predict one of the six dermatological diseases Acne, Allergy, Melasma, Milia, Psoriasis, and Vitiligo. Performance of a feature extractor-classifier combination is assessed by a set of basic classification based metrics including accuracy, precision, recall, and F1-score. Confusion matrices are used to examine class wise prediction behaviour as well as inter-class confusion, and Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) metrics are employed to examine threshold-independent discriminative performance. This assessment method helps compare CNN and Bi-LSTM models based on features extracted from VGG19 and DenseNet-169. This enables us to understand how well spatial learning using CNN and sequential learning using Bi-LSTM work for automatic dermatological disease detection and provides useful insights into their effectiveness for medical diagnosis. The DenseNet-169 architecture used in this study is shown in **FIGURE 2**.

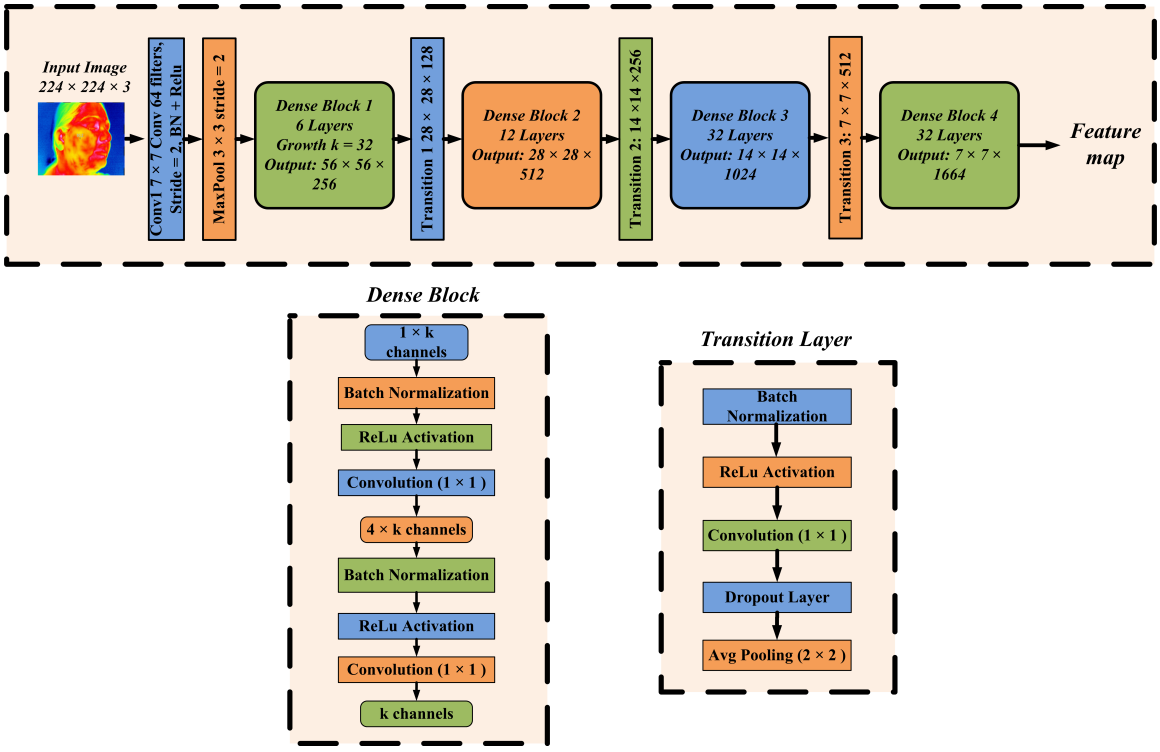

Figure 2: DenseNet-169 deep feature extraction architecture for dermatological disease analysis.

3. RESULTS AND DISCUSSION

This section presents the results of evaluation of the proposed deep learning framework for dermatological disease identification using thermal images. The performance of DenseNet-169 and VGG19 employed as feature extractors is investigated under CNN and Bi-LSTM based models used as classifiers. The investigations focused on training performance, classification accuracy, class-wise performance metrics, confusion matrix, and ROC characteristic. The investigations present not

only overall accuracy but also class-level behaviour, robustness, and model interpretability, under the following sub sections.

3.1 DENSENET169 Feature Extractor-Based Classification

The training and validation curves of Densenet169 feature extractor-based classification are shown in **FIGURE 3**. **FIGURE 3a** and **FIGURE 3b** correspond to the model employing DenseNet-169 as a feature extractor followed by CNN-based classification. The steep decline in loss curves (**FIGURE 3a**) during the initial training phase (epochs less than 10) for both training and validation show fast training. The rapidly converging to very low values and remaining stable throughout the 100 training epochs. The close alignment of the training and validation loss curves indicates effective learning with minimal overfitting. This behaviour is further reflected in the accuracy curves **FIGURE 3b**, where both training and validation accuracies increase sharply and reach near-saturation levels (approximately 99–100%) within the early epochs, maintaining consistent performance thereafter. The trends indicate that the CNN classifier is highly effective in exploiting the densely connected feature representations produced by DenseNet-169, leading to fast convergence and strong generalization.

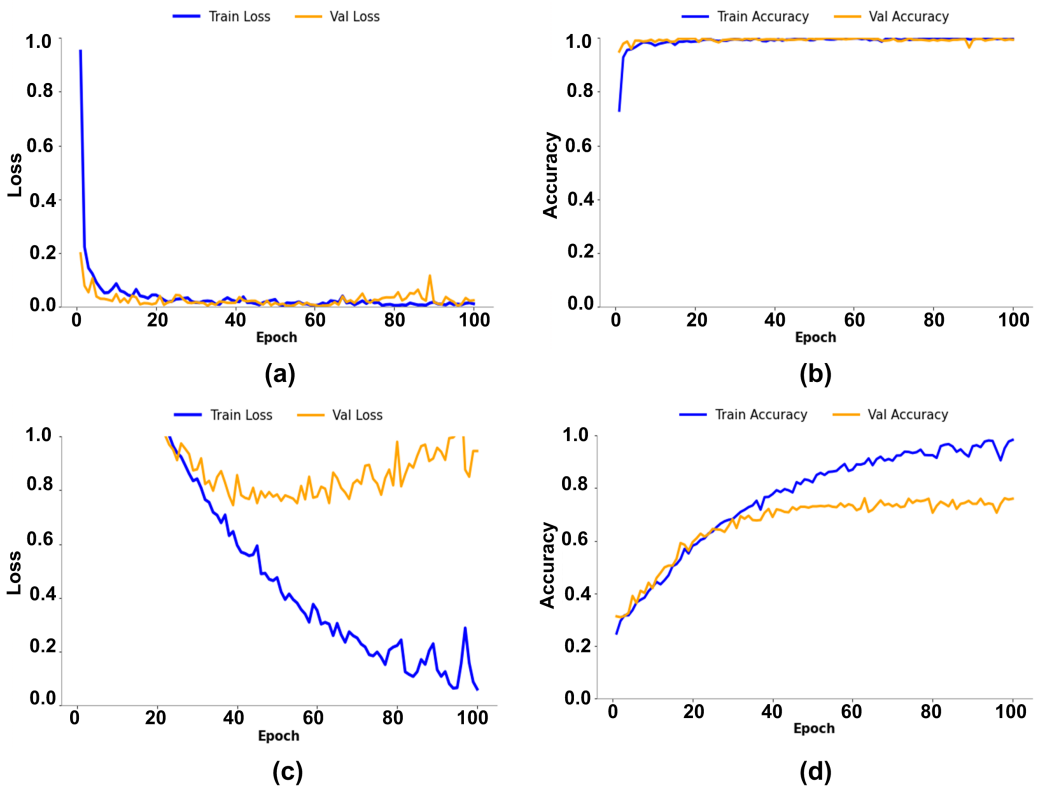

Figure 3: Training and validation loss and accuracy curves for the CNN-based classifier (a–b) and the Bi-LSTM-based classifier (c–d).

In contrast, **FIGURE 3c** and **FIGURE 3d** illustrate the performance of DenseNet-169 features classified using Bi-LSTM. The training loss decreases steadily across epochs (c), whereas the validation loss remains relatively high and exhibits noticeable fluctuations, particularly in later epochs. This divergence between training and validation loss suggests overfitting and reduced generalization capability. The corresponding accuracy curves (d) show that, although training accuracy continues to improve and approaches high values, validation accuracy plateaus at lower level (around 70–75%), resulting in a pronounced gap between the two curves. This indicates that sequential modelling via Bi-LSTM is less suited for the predominantly spatial features extracted by DenseNet-169. Overall, the comparative analysis highlights that, when DenseNet-169 is used as the feature extractor, CNN-based classification achieves superior convergence behaviour, higher accuracy, and more stable generalization compared to Bi-LSTM-based classification.

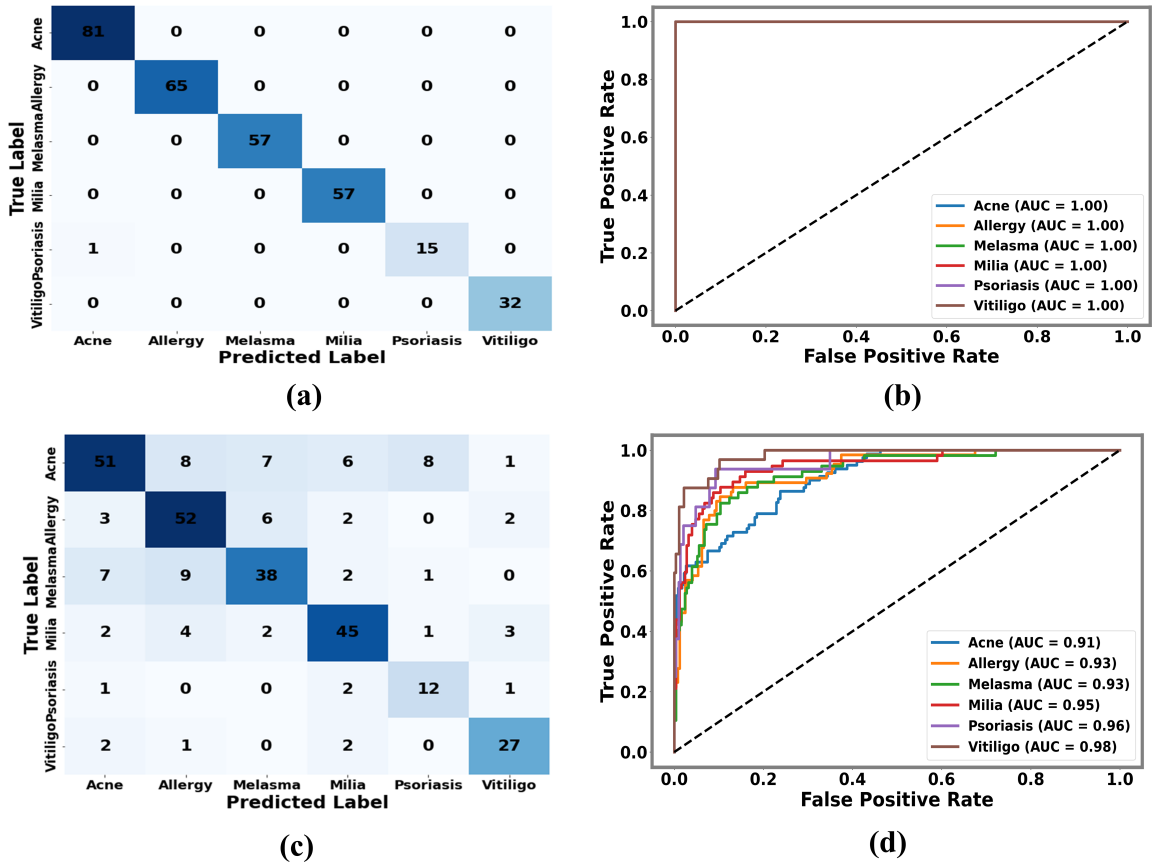

Figure 4: Confusion matrices and ROC curve analysis for DenseNet169-based classification: (a–b) CNN classifier and (c–d) Bi-LSTM classifier.

The results presented in **FIGURE 4** correspond to the performance of the CNN classifier using DenseNet-169 as the feature extractor (**FIGURE 4a, FIGURE 4b**), while the performance of the Bi-LSTM classifier operating on DenseNet-169 extracted features is shown in **FIGURE 4c** and **FIGURE 4d**. The confusion matrix in **FIGURE 4a** indicates that the CNN classifier achieves near-perfect class-wise discrimination when DenseNet-169 features are used, attaining an overall

accuracy of 99.68%. Most samples are correctly classified along the main diagonal, with minimal misclassification across dermatological classes such as Acne, Allergy, Melasma, Milia, Psoriasis, and Vitiligo. This highlights the strong discriminative power of DenseNet-169 densely connected feature representations in combination with convolutional classification. The corresponding ROC curves in **FIGURE 4b** further support this observation, with all classes achieving an AUC of 1.00, indicating an ideal balance of sensitivity and specificity and demonstrating high robustness and reliability.

In contrast, the confusion matrix in **FIGURE 4c** shows increased off-diagonal entries, reflecting higher inter-class confusion, particularly between visually similar conditions such as Acne, Allergy, and Melasma, as well as between Milia and Psoriasis. The Bi-LSTM configuration yields an overall accuracy of 73.05%. In **FIGURE 5**, samples (a-c) for Acne and Allergy were wrongly assigned as well, because of both the overlap in inflammatory thermal signatures and the diffuse heat distribution. The similarity in the surface temperature rise and irregular gradient spread of these two conditions led to an impaired separation between classes resulting in prediction errors despite the robustness of the models. Sequential modelling was initially proposed to capture contextual dependencies among deep feature embeddings because dermatological diseases frequently display distinctive spatial texture variations influenced by inflammatory or microbial processes, however, the experimental performance did not support this assumption. In this respect, the deep features extracted from DenseNet-169 are given global average pooling prior to classification, resulting in flattened feature vectors that do not have inherent spatial or directional ordering. Though these vectors encode hierarchical spatial information internally, their one-dimensional representation does not correspond to a meaningful sequential structure. As Bi-LSTM networks fundamentally capture dependencies with sequence or time order, a lack of ordering induces a representation–model mismatch. Consequently, the inductive bias of Bi-LSTM does not match the statistical structure of the pooled feature vectors well thus its ability to exploit spatial correlations is limited and the generalization becomes reduced class overlap increases.

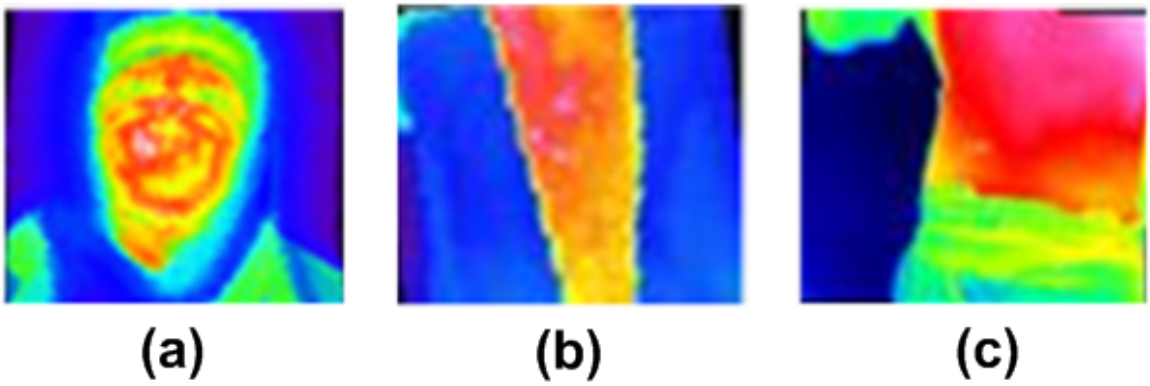

Figure 5: Misclassified cases between Acne and Allergy based on thermal imaging.

The obtained ROC curves in **FIGURE 4d** confirm the findings, suggesting that AUC values, although still high (ranging approximately from 0.91 to 0.98), are consistently lower than those produced by the CNN classifier, demonstrating comparatively reduced discriminative capability. Taken together, the findings clearly show that CNN-based classification achieves better class-wise

accuracy and ROC characteristics than Bi-LSTM-based classification under the feature extractor DenseNet-169. This underlines that convolutional classifier is in accordance with the spatially rich representations learned under DenseNet-169 while sequential modelling through Bi-LSTM introduces limitations that adversely affect class separability and overall diagnostic performance. The observation is also corroborated by class-wise assessment in **FIGURE 6**, which includes precision, recall, and F1-score comparisons between CNN- and Bi-LSTM-based classifiers employing DenseNet-169 features. In all dermatological diseases, the CNN model consistently achieves near-saturated precision and recall, leading to F1-scores that approach unity for most classes. Overall, the results support that DenseNet-169 produces highly discriminative and reusable feature representations that are efficiently leveraged by the CNN to achieve balanced, stable, and robust classification performance.

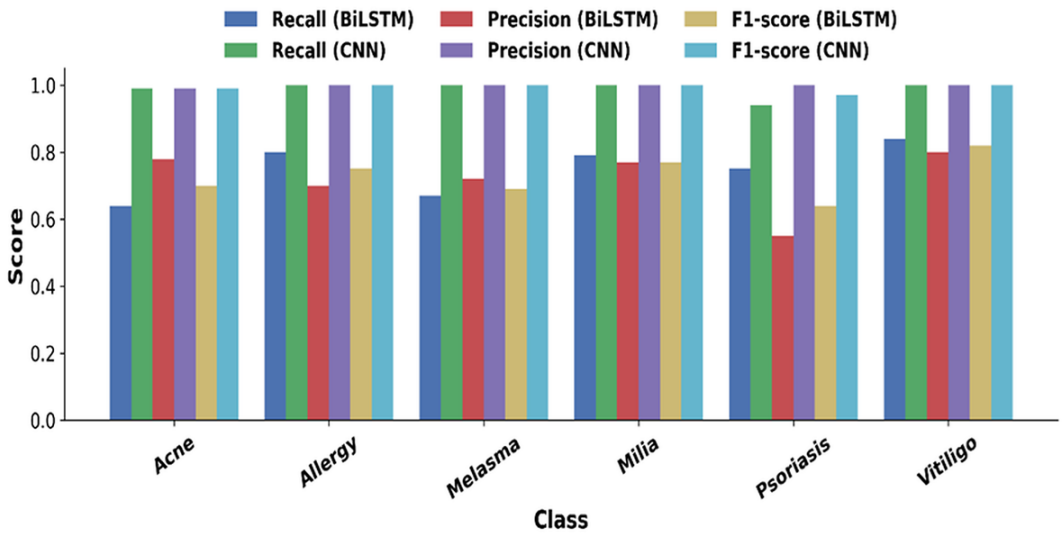

Figure 6: Class-wise comparison of precision, recall, and F1-score for CNN and Bi-LSTM classifiers using DenseNet-169 feature extraction.

On the other hand, the Bi-LSTM classifier shows relatively low recall and F1-score, especially for visually complex classes such as Acne, Melasma, and Psoriasis, where intra-class variability and inter-class similarity are more pronounced. Despite achieving fairly high precision for some classes, the recall value is lower and the model likely fails to identify detections when operating on DenseNet-169 feature maps. These findings support that the densely connected, spatially rich features produced by DenseNet-169 are more naturally compatible with convolutional classification, whereas sequential modelling through Bi-LSTM does not fully leverage the spatial dependencies inherent in such representations. Overall, the histogram clearly shows that when DenseNet-169 serves as the feature extractor, CNN-based classification shows a more consistent and balanced performance across all evaluation metrics, validating its suitability for dermatological image analysis within the proposed framework.

3.2 VGG19 Feature Extractor-Based Classification

The training and validation curves are shown in **FIGURE 7** correspond to the model employing VGG19 as a feature extractor followed by CNN-based classification (**FIGURE 7a and FIGURE 7b**), trained for a total of 100 epochs. In this configuration, both training and validation losses decrease sharply within the initial epochs and stabilize at near-zero values well before the end of training. The intense overlap of the loss curves during the rest of the epochs indicates steady convergence and no obvious overfitting. The accuracy curves have a similar trend, indicating that the levels of training and validation accuracies are increasing at an all-time high rate and saturating (approximately 99–100%) until well beyond the initial stages of training, with similar general consistency up to the 100th epoch. This behaviour evidences that the CNN classifier is able to effectively exploit the spatial features extracted by VGG19 and converge efficiently within the given number of epochs. We demonstrate the training and validation curves in **FIGURE 7c and FIGURE 7d** is the VGG19 feature extractor with Bi-LSTM-based classifier trained for 100 epochs. Unlike CNN-based methods, loss curves show a gradual descent through each epoch, thus showing a slower process of convergence.

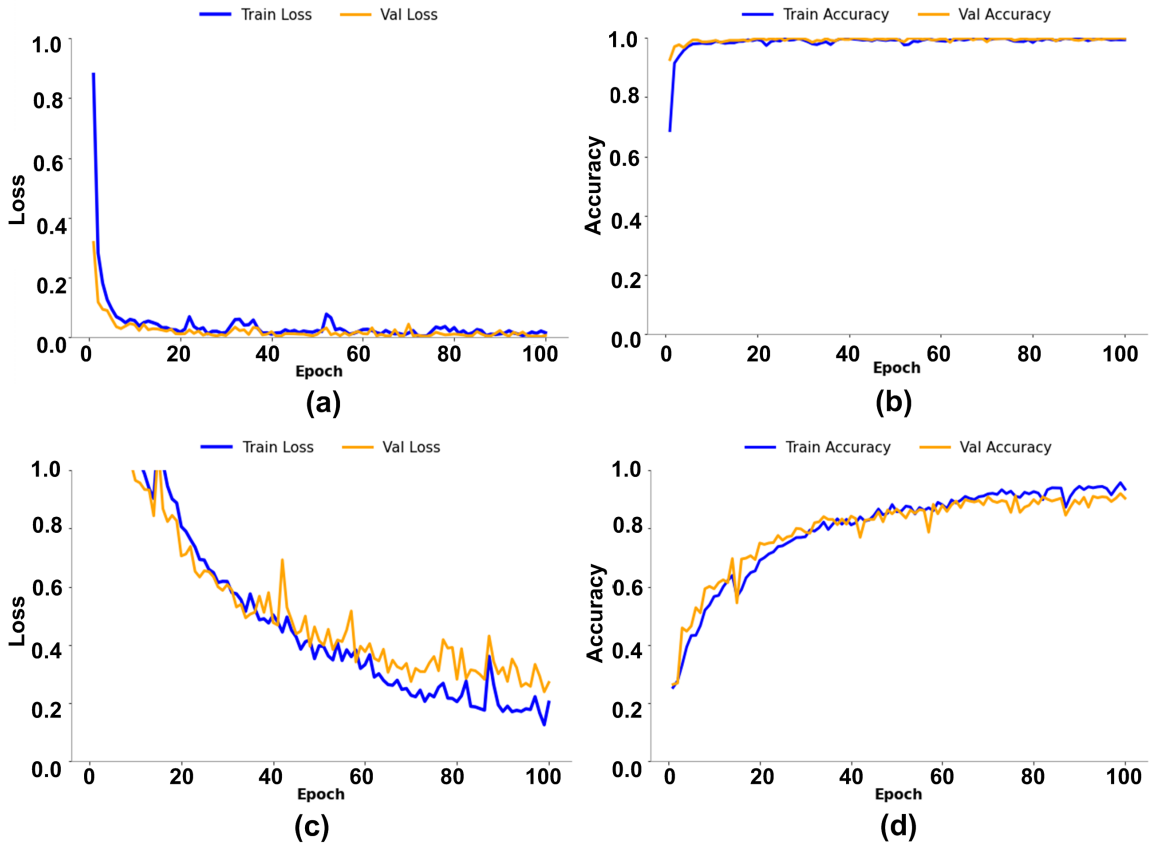

Figure 7: Training and validation loss and accuracy curves for the CNN-based classifier (a–b) and the Bi-LSTM-based classifier (c–d)

The validation loss **FIGURE 7a** is higher than that of the training one consistently and the distribution varies throughout training, which indicates increasing sensitivity to data variability and mild overfitting. The accuracy curves **FIGURE 7b** show a consistent gain over epochs, reaching a level of around 90–95% at the end of the training. The small difference between training and validation accuracies indicates reduced generalization potential comparatively.

Overall, despite both models being trained for the same number of epochs, the CNN-based classifier achieves faster convergence and higher accuracy than the Bi-LSTM-based classifier, indicating its greater suitability for classification using VGG19-extracted spatial features in the present study.

The results shown in **FIGURE 8**, correspond to the performance of the CNN classifier using VGG19 as the feature extractor (**FIGURE 8a and FIGURE 8b**), whereas **FIGURE 8c** and **FIGURE 8d** represent the performance of the Bi-LSTM classifier operating on VGG19 extracted features. The confusion matrix in **FIGURE 8a** demonstrates that the CNN classifier achieves high class wise accuracy of 99.35%, with most samples correctly classified along the diagonal and only negligible off diagonal entries. All classes, including features extracted by VGG19 are effectively exploited by the CNN classifier.

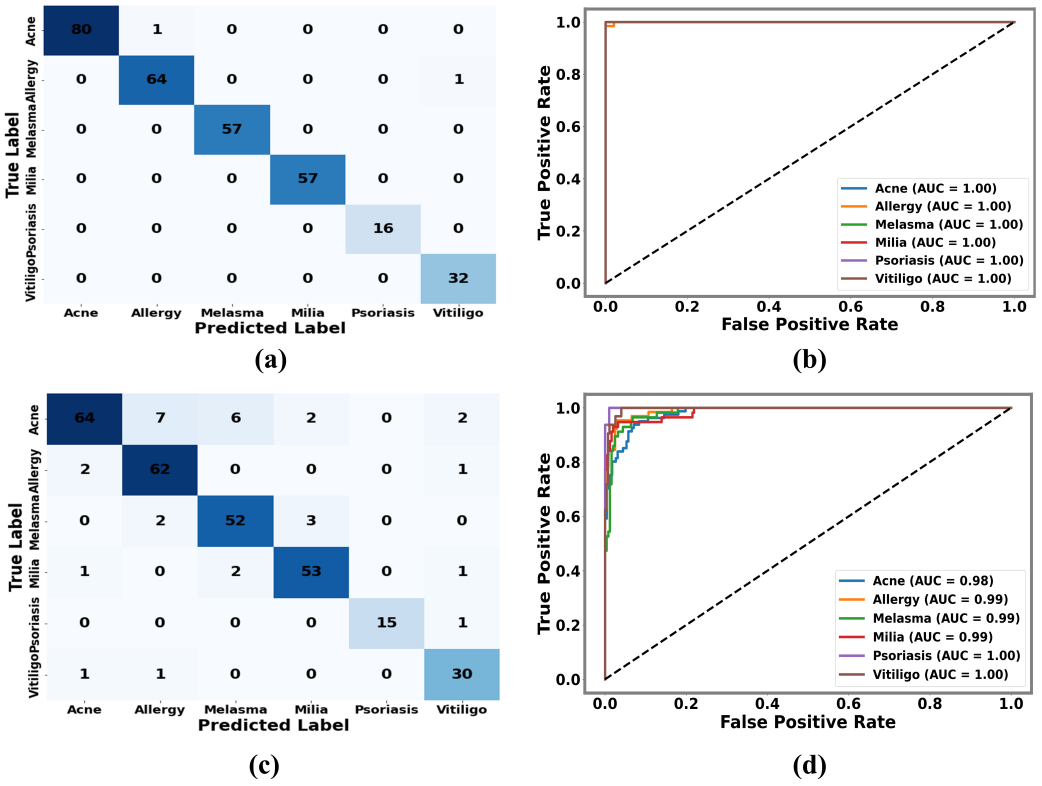

Figure 8: Confusion matrices and ROC curve analysis for VGG19-based classification: (a–b) CNN classifier and (c–d) Bi-LSTM classifier.

The ROC curves in **FIGURE 8b** corroborate its discriminative performance, showing that all classes reach an AUC of 1.00 (optimal compromise trade-off between sensitivity and specificity confirming robustness of CNN based model. **FIGURE 8c** shows a relative decrease in the performance of

the Bi-LSTM classifier with VGG19 features, with an overall accuracy of 89.61%. The high off diagonal values suggest a higher level of inter-class confusion, especially among visually similar conditions such as Acne, Allergy, and Melasma, and also confusion between Milia and Psoriasis. This indicates that the sequential model used by the Bi-LSTM does not exploit well the spatial and hierarchical representation based on VGG19. The ROC curves in **FIGURE 8d** presents AUC values ranging approximately from 0.98 to 1.00, a feature that is indicative of strong discriminative capability. Hence, the comparative study indicates that CNN classification surpasses Bi-LSTM classification when employing VGG19 as a feature extractor in terms of achieved class-wise accuracy and ROC characteristics. These results show that convolutional classifiers fit better with the spatial representation of the features that VGG19 learns, rather than sequential modelling through Bi-LSTM that introduces limitations, increases class confusion, and causes a decrease in diagnostic performance. The histogram of class-wise precision, recall, and F1-score for all classes, as presented in **FIGURE 9**, illustrates that the CNN classifier consistently achieves higher values for these metrics when applied across features extracted from VGG19. For Melasma, Milia, and Psoriasis, CNN achieves almost saturated precision, recall, and F1-score as shown in the histogram, showing highly reliable discrimination. Conversely, the Bi-LSTM classifier exhibits lower recall and F1-score than the CNN, predominantly for Acne and Allergy, indicating an elevated miss rate and reduced robustness for visually overlapping patterns. The performance difference observed seems to be due to the fact that our VGG19 features (predominantly spatial and hierarchical) are largely derived from VGG19. The CNN classifier is more consistent with such a model and can better capture local and global visual cues, but the Bi-LSTM introduces the sequential model which does not directly represent the spatial structure of the VGG19 feature maps. In general, when the VGG19 is used as the feature extractor, it shows better and more consistent performance in precision, recall, and F1 score in the CNN approach and therefore it proves to be a better choice for the proposed dermatological image classification framework. The comparative study of dermatological classification models employing various feature extraction routines and classifier models is presented in **TABLE 1** with a description of such methodologies and performance.

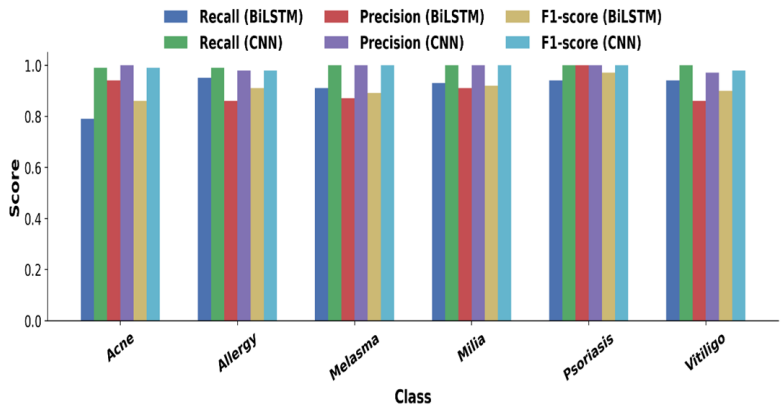

Figure 9: Histogram-based class-wise evaluation of precision, recall, and F1-score achieved by CNN and Bi- LSTM classifiers employing VGG19-extracted deep features.

Table 1: Overall performance comparison of dermatological classification models using different feature extractors and classifier architectures.

Feature Extractor	**Classifier**	**Accuracy**	**Precision**	**Recall**	**F1-score**
DenseNet169	CNN	0.9968	0.9980	0.9896	0.9936
DenseNet169	Bi-LSTM	0.7305	0.7158	0.7466	0.7264
VGG19	*CNN*	0.9935	0.9924	0.9954	0.9938
VGG19	*Bi-LSTM*	0.8961	0.9066	0.9102	0.9063

Overall, the experimental results show that thermal imaging based deep learning has proven to be able to exhibit good discrimination reliability-based classification for various dermatological disorders. Thermal modalities, in contrast to conventional visible-spectrum imaging, can capture underlying biological events such as inflammatory heat patterns, vascular alterations and metabolic activity while also revealing an underlying disease pattern in a physiological sense and may serve as complement to texture signals other than simply the physical appearance of the skin. The high performance of the DenseNet-169 based framework shows its excellent behaviour and its strong ability to simulate minute temperature gradients in low dynamic range medical images of static dimensions. The design of the proposed system is not as a replacement for dermoscopic evaluation but as an adjunctive decision support platform, which could be used for early diagnosis, triage and longitudinal monitoring across non-invasive clinical processes. Due to steady convergence, computational acceptability, and constant generalization over classes, the framework shows practical scalability and possible feasibility for the deployment in tele dermatology and resource limited healthcare environments.

Table 2: Detailed Architecture and Parameter Analysis of the proposed DenseNet169–CNN Framework Identification of Dermatological Disease using thermal images.

Layer Type	***Output Shape***	***Parameters***
Input Image	$[3 \times 224 \times 224]$	–
DenseNet169 (Feature Extractor)	[1664]	14,149,480
Global Avg Pool + Flatten	[1664]	–
Reshape	$[1664 \times 1]$	–
Conv1D (64, $K = 3$)	$[1662 \times 64]$	256
MaxPooling1D (2)	$[831 \times 64]$	–
Flatten	[53184]	–
Fully Connected (fc1)	[128]	6,808,704
Dropout (0.3)	[128]	–
Output (Softmax, 6)	[6]	774
Total (Classifier Only)	–	6,809,734
Total (With Dense-Net)	–	20.96 M

Recent studies were also studied, they found the accuracy on infrared thermal images for the human-augmented Dense-Net framework was 96.35% [40]. In addition, the optimized CNN model from Villa-Pulgarin et al. reported a 98% accuracy on the HAM10000 dataset, demonstrating its excellent

classification ability [41]. The hybrid EViT-Dens169 model integrating vision transformer and CNN features exhibited 97.1% accuracy on the ISIC 2018 dataset [42]. Additionally, the Dense-Net-based model with fused metadata presented a lower accuracy of 81.4% on the PAD/ISIC dataset [43]. Finally, the proposed thermal imaging-based DenseNet-169 model outperforms the state-of-the-art with an accuracy of 99.68%, proving its robustness, reliability, and effectiveness for thermal image-based skin disease detection, as supported by the parametric analysis presented in **TABLE 2.**

4. CONCLUSION

The present research proposed and investigated a deep learning-based framework for dermatological diseases identification using thermal skin images. The DenseNet-169 and VGG19 were employed as deep feature extractors, followed by CNN and Bi-LSTM classifiers. The experimental evaluation showed that both pretrained networks are capable of extracting highly discriminative thermal features; however, their effectiveness is influenced by the following classifier. The CNN classifier consistently outperformed Bi-LSTM classifier in terms of accuracy, precision, recall, F1-score, and ROC–AUC, irrespective of the feature extractor. When DenseNet-169 was combined with CNN classifier, the system was able to extract distinctive features and thereby, resulting in near-perfect classification performance with an accuracy of 99.68%. It also showed stable convergence and strong generalization capability. In contrast, the DenseNet-169, followed by Bi-LSTM configuration exhibited substantially reduced identification accuracy of 73.05%, accompanied by higher misclassification rates and limited generalization capability. The combination of VGG19 and the CNN provided slightly less performance, giving an accuracy of 99.35%. The combination of VGG19 with Bi-LSTM could reach only to an accuracy of 89.61%. It also lacks in convergence and inter-class differentiation capabilities.

Despite the strong performance, certain limitations should be acknowledged. In the present study, deep features were extracted from pretrained convolutional networks and subsequently used for classification via a structured CSV-based feature representation. While this approach improves computational efficiency and experimental reproducibility, it does not inherently provide spatial localization of discriminative regions, thereby limiting interpretability at the lesion level. The current work primarily focused on representation learning and classifier compatibility analysis. Future research may integrate attention mechanisms and heatmap-based visualization techniques (e.g., Grad-CAM) to enhance spatial explainability and clinical transparency. Additionally, the incorporation of multimodal clinical data will be explored, and the evaluation of EfficientNet, MobileNet, and Vision Transformer (ViT) architectures will be considered to further improve scalability and real-world clinical deployment.

5. ACKNOWLEDGEMENTS

The author **Ashu Sharma**, like to express her sincere gratitude to the Govt. of India, Ministry of Science and Technology, Department of Science and Technology, for their support and encouragement under the Wise-Ph.D. scheme (DST/WISE-PhD/PM/2023/15). The authors acknowledge the support of Department of Computer Science & IT and Department of Electronics, University of Jammu for providing research facilities for this study.

6. COMPETING INTERESTS

The authors declare that they have no competing interests.

7. DATA AVAILABILITY

The dataset used in this study was collected firsthand by the authors. Although it is not publicly available, it can be shared upon request.

8. CONSENT FOR PUBLICATION

Informed consent was obtained from all participants for the use of anonymized images and associated data in research and publication.

9. ETHICS APPROVAL

NOC from the ethical committee of University of Jammu had been taken before recording the dermatological images from the patients (Reference No. RA/24/4843).

References

[1] Sadik R, Majumder A, Biswas AA, Ahammad B, Rahman MM. An In-Depth Analysis of Convolutional Neural Network Architectures With Transfer Learning for Skin Disease Diagnosis. Healthcare Analytics. 2023;3:100143.

[2] Kousis I, Perikos I, Hatzilygeroudis I, Virvou M. Deep Learning Methods for Accurate Skin Cancer Recognition and Mobile Application. Electronics. 2022;11:1294.

[3] Tahir M, Naeem A, Malik H, Tanveer J, Naqvi RA, et al. DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopic Images. Cancers. 2023;15:2179.

[4] Zhi S, Li Z, Yang X, Sun K, Wang J. A Multiclass Skin Disease Classification Model Using Dermatoscopy Images With Inception-V2. Appl Sci. 2024;14:10197.

[5] Farooq MA, Javidnia H, Corcoran P. Performance Estimation of State-Of-The-Art Convolutional Neural Networks for Thermal Images-Based Gender Classification. J Electron Imaging. 2020;29:063004.

[6] Bai Y, Liu L. Personal Thermal Preference Prediction Based on Deep Feature Extraction and Machine Learning. SSRN. 2024:35. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4933823

[7] Leo H, Saddami K, Roslidar, Muharar R, Munadi K, et al. Lightweight Convolutional Neural Network (CNN) Model for Obesity Early Detection Using Thermal Images. Digit Health. 2024;10.

[8] Dey S, Roychoudhury R, Malakar S, Sarkar R. Screening of Breast Cancer From Thermogram Images by Edge Detection Aided Deep Transfer Learning Model. Multimed Tools Appl. 2022;81:9331-9349.

[9] Grignaffini F, Troiano M, Barbuto F, Simeoni P, Mangini F, et al. Anomaly Detection for Skin Lesion Images Using Convolutional Neural Networks and Handcrafted Feature Injection. Algorithms. 2023;16:466.

[10] Sharma A, Sawhney T, Abrol P, Lehana PK. Identification of Dermatological Diseases Using Ai-Driven Entropy and Texture-Based Analysis. J Sci Res. 2026;18:123-135.

[11] Balaha HM, Hassan AE, El-Gendy EM, ZainEldin H, Saafan MM. An Aseptic Approach Towards Skin Lesion Localization and Grading Using Deep Learning and Harris Hawks Optimization. Multimed Tools Appl. 2024;83:19787-19815.

[12] Sharma P, Abrol P. Multi-Component Image Analysis for Citrus Disease Detection Using Convolutional Neural Networks. Crop Prot. 2025;193:107181.

[13] Sharma P, Abrol P. Agricultural Advancements Through Machine Learning Technologies. Environ Ecol. 2024;42:775-779.

[14] Sharma P, Abrol P. Analysis of Multiple Component-Based CNN for Similar Citrus Species Classification. In Modern Approaches in Machine Learning Cognitive Science: A Walkthrough. Cham: Springer International Publishing. 2022:221-232.

[15] Mahajan P, Abrol P, Lehana PK. Scene Based Classification of Aerial Images Using Convolution Neural Networks. J Sci Ind Res. 2020;79.

[16] Sawhney T, Sharma A, Abrol P, Kumar Lehana P, Chaahat, et al. Fingerprint Matching for Noisy and Distorted Patterns Using a Siamese Network With ResNet50 and Multihead Attention. IEEE Access. 2025;13:88047-88064.

[17] Sawhney T, Lehana PK. Extraction of Gender Specific Hidden Information From Head Related Transfer Function Using Machine Learning. Adv Artif Intell Mach Learn. 2025;5:4747-4764.

[18] Kleebauer M, Hafdaoui H, Bouchakour S, Häckner B, Lindenmeyer M. Globally Scalable Qgis-Integrated Workflow for Solar Photovoltaic System Segmentation and Capacity Estimation: A Case Study in Algeria. IGARSS 2025-2025 IEEE International Geoscience and Remote Sensing Symposium. IEEE. 2025:8000-8004.

[19] Hafdaoui H, Kleebauer M, Bouzekri A, Belhaouas N, Charki A, et al. Detection of Photovoltaic Power Plants in Satellite Images Using Artificial Intelligence Techniques. Next Res. 2026;6:101385.

[20] Naeem A, Anees T. DVFNet: A deep feature fusion-based model for the multiclassification of skin cancer utilizing dermoscopy images. PLOS One. 2024;19:e0297667.

[21] Sadeghi M, Lee TK, McLean D, Lui H, Atkins MS. Detection and Analysis of Irregular Streaks in Dermoscopic Images of Skin Lesions. IEEE Trans Med Imaging. IEEE. 2013;32:849-861.

[22] Adegun AA, Viriri S. Deep Learning-Based System for Automatic Melanom a Detection. IEEE Access. IEEE. 2019;8:7160-7172.

[23] Moldovan D. Transfer Learning-Based Method for Two-Step Skin Cancer Images Classification. In 2019 E-Health and Bioengineering Conference. EHB. IEEE. 2019:1-4.

[24] Çevik E, Zengin K. Classification of Skin Lesions in Dermatoscopic Images With Deep Convolution Network. Avrupa Bilim Teknol Derg. 2019:309-318.

[25] Wu Y, Lariba AC, Chen H, Zhao H. Skin Lesion Classification Based on Deep Convolutional Neural Network. In 2022 IEEE 4th international conference on power, intelligent computing and systems. ICPICS. IEEE. 2022:376-380.

[26] Sharma P, Gautam A, Nayak R, Balabantaray BK. Melanoma Detection Using Advanced Deep Neural Network. In2022 4th international conference on energy, power and environment. ICEPE. IEEE. 2022:1-5.

[27] Nadipineni H. Method to Classify Skin Lesions Using Dermoscopic Images. 2020. ArXiv Preprint: https://arxiv.org/pdf/2008.09418.

[28] Jojoa Acosta MF, Caballero Tovar LY, Garcia-Zapirain MB, Percybrooks WS. Melanoma Diagnosis Using Deep Learning Techniques on Dermatoscopic Images. BMC Med Imaging. 2021;21:6.

[29] Girdhar N, Sinha A, Gupta S. DenseNet-II: An Improved Deep Convolutional Neural Network for Melanoma Cancer Detection. Soft Comput. 2023;27:13285-13304.

[30] Alom MZ, Aspiras T, Taha TM, Asari VK. Skin Cancer Segmentation and Classification With NABLA—N and Inception Recurrent Residual Convolutional Networks. 2019. arXiv preprint: https://arxiv.org/pdf/1904.11126.

[31] Khan MA, Zhang YD, Sharif M, Akram T. Pixels to Classes: Intelligent Learning Framework for Multiclass Skin Lesion Localization and Classification. Comput Electr Eng. 2021;90:106956.

[32] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014. ArXiv preprint: https://arxiv.org/pdf/1409.1556

[33] Minarno AE, Lusianti A, Azhar Y, Wibowo H. Classification of Skin Cancer Images Using Convolutional Neural Network With ResNet50 Pre-Trained Model. JOIV Int J Inform Visualization. 2024;8:2013-2019.

[34] Ukwuoma CC, Qin Z, Heyat MB, Akhtar F, Smahi A, Jackson JK et al. Automated Lung-Related Pneumonia and COVID-19 Detection Based on Novel Feature Extraction Framework and Vision Transformer Approaches Using Chest X-Ray Images. Bioengineering. 2022;9:709.

[35] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. InProceedings of the IEEE conference on computer vision and pattern recognition. CVPR. IEEE. 2017:4700-4708.

[36] Mazhar F, Aslam N, Naeem A, Ahmad H, Fuzail M, et al. Enhanced Diagnosis of Skin Cancer From Dermoscopic Images Using Alignment-Optimized CNNs and Grey Wolf Optimization. J Comput Theor Appl. 2025;2:368-382.

[37] Amjad K, Asif S, Waheed Z, Guo Y. A Novel Lightweight Deep Learning Framework With Knowledge Distillation for Efficient Diabetic Foot Ulcer Detection. Appl Soft Comput. 2024;167:112296.

[38] Ukwuoma CC, Qin Z, Belal Bin Heyat MB, Akhtar F, Bamisile O, et al. A Hybrid Explainable Ensemble Transformer Encoder for Pneumonia Identification From Chest X-Ray Images. J Adv Res. 2023;48:191-211.

[39] Unnisa K, Prasanna K, Ratkal B. A Review on Skin Cancer Detection Using Deep Learning Algorithms and LBP. IJRASET. 2024;12.

[40] She X, Lu H, Liu Q, Xie P, Xia Q. Dermatological Infrared Thermal Imaging With Human–Machine Interaction Image Diagnostics Interface Using DenseNet. J Radiat Res Appl Sci. 2024;17:100826.

[41] Villa-Pulgarin JP, Ruales-Torres AA, Arias-Garzon D, Bravo-Ortiz MA, Arteaga-Arteaga HB, et al. Optimized Convolutional Neural Network Models for Skin Lesion Classification. Comput Mater Continua. 2022;70:2131-2148.

[42] Halawani HT, Senan EM, Asiri Y, Abunadi I, Mashraqi AM, et al. Enhanced Early Skin Cancer Detection Through Fusion of Vision Transformer and CNN Features Using Hybrid Attention of Evit-DENS169. Sci Rep. 2025;15:34776.

[43] Yin W, Huang J, Chen J, Ji Y. A Study on Skin Tumor Classification Based on Dense Convolutional Networks With Fused Metadata. Front Oncol. 2022;12:989894.