

Neural Computing and Applications

Autoimmune Blistering Skin Disease Classification: Classical Machine Learning, Deep Learning, and Hybrid Approaches

--Manuscript Draft--

Manuscript Number:	NCAA-D-24-01150R2
Full Title:	Autoimmune Blistering Skin Disease Classification: Classical Machine Learning, Deep Learning, and Hybrid Approaches
Article Type:	Original Article
Keywords:	Autoimmune Blistering Skin Disease; Classification; Machine Learning; deep learning
Abstract:	<p>Contrary to common belief, skin diseases can be life-threatening and have a serious impact on patients' lives. Autoimmune blistering skin diseases (AIBD) of the skin are one of them. The diagnosis based on clinical examination is subjective and irreproducible as it depends upon the expertise and experience of the treating dermatologists. A correct etiological diagnosis of AIBD requires extensive investigations that are costly and time-consuming, leading to delays in confirmation of the diagnosis. Early diagnosis and treatment of AIBD are very important for a better prognosis. Artificial intelligence techniques, particularly convolutional neural networks (CNNs), are being used across several medical domains, including dermatology, and have the potential to assist dermatologists in making early-stage decisions. This paper explores various AI approaches. The proposed models utilized traditional approaches, deep learning techniques, and hybrid methods to classify AIBD by analyzing clinical images. The models were compared using performance parameters and their shortcomings and potential improvement strategies were analyzed based on the experimental results. In the traditional approach, the random forest attained the highest accuracy of 57%, while a hybrid model, Xception integrated with support vector machine (SVM), achieved an overall accuracy of 81%. However, in deep learning techniques, certain models performed very well and showed accuracy that ranged from 84% to 98%, particularly after additional fine-tuning processes.</p>
Response to Reviewers:	<p>Point-by-Point Response to the Reviewers' Comments</p> <p>Reviewer #3:</p> <p>Comment 1: Dataset Limitations: The small sample size (100 images) significantly limits the generalizability of the findings. Expanding the dataset with diverse, high-quality images would be essential to improve robustness and reliability. As acknowledged in our manuscript that, the sample size of 100 images in our study is relatively small, which may limit the generalizability of our findings. To the best of our knowledge, there is no publicly available dataset specifically focused on autoimmune blistering skin diseases due to the rarity of these conditions. For our pilot study, we collected most of the images from publicly available online repositories. However, this study was designed as a foundational step toward developing a practical model for autoimmune blistering skin diseases, which are rare and challenging to study due to limited data availability. We are actively working to expand our dataset by collecting clinically diverse and high-quality images from dermatology clinics in real clinical settings. The primary objective of this work was to identify the most suitable approach, evaluate the strengths and limitations of different methodologies, and establish best practices for developing a robust classification model. This has also been discussed in our manuscript. In future work, we plan to extend this research by applying the model to a larger, more diverse dataset collected from real clinical settings, which will enhance its generalizability and clinical applicability.</p> <p>Comment 2: Reproducibility: Providing open access to the dataset (after addressing ethical considerations) and detailed implementation code or algorithm would enhance the reproducibility of this work. Thank you for the feedback. Regarding reproducibility, the dataset used in this study was sourced from publicly available online repositories, supplemented by a small</p>

number of clinical images collected from the dermatology clinics at PGIMER, Chandigarh, after obtaining ethical approval. Reference for the publicly available images is provided in the data availability section of manuscript. However, the clinical images collected from PGIMER contain sensitive patient information and are subject to ethical and privacy restrictions, preventing their public release. Despite this limitation, we have provided a detailed description of our methodology, including the skin disease classification algorithm, which can be tested using publicly available online images. We remain committed to transparency and reproducibility and will continue to explore ways to share additional resources, such as anonymized datasets, in compliance with ethical guidelines and institutional approvals.

The detailed implementation of the algorithm is provided in the supplementary material, and the revised manuscript includes a mention of these details (Page no 11).

Comment 3: Comparative Analysis:

While the study compares classical, deep learning, and hybrid methods, the rationale for the observed performance differences across models could be better articulated. This analysis would offer insights into model behavior and guide future work.

Thank you for your insightful comment. We have revised the manuscript to include rationale for the observed performance differences across models has been discussed in detail in the Results and Discussion section (Page no 32).

Comment 4: Visualization and Presentation:

Some visualizations, such as performance plots and edge detection examples, need better formatting for clarity. Separating key plots from supplementary material improves accessibility for readers. Also, the figures need to be more visible.

Thank you for the feedback. We have revised the manuscript; all figures have been reformatted for clarity, and the Figures 8, 9, and 10, which contain multiple plots, divided into subparts for better visibility. Only figures that are not referenced or used further have been placed in the supplementary material to enhance accessibility for readers.

Point-by-Point Response to the Reviewers' Comments

Reviewer #3:

Comment 1: Dataset Limitations:

The small sample size (100 images) significantly limits the generalizability of the findings. Expanding the dataset with diverse, high-quality images would be essential to improve robustness and reliability.

As acknowledged in our manuscript that, the sample size of 100 images in our study is relatively small, which may limit the generalizability of our findings. To the best of our knowledge, there is no publicly available dataset specifically focused on autoimmune blistering skin diseases due to the rarity of these conditions. For our pilot study, we collected most of the images from publicly available online repositories. However, this study was designed as a foundational step toward developing a practical model for autoimmune blistering skin diseases, which are rare and challenging to study due to limited data availability. We are actively working to expand our dataset by collecting clinically diverse and high-quality images from dermatology clinics in real clinical settings. The primary objective of this work was to identify the most suitable approach, evaluate the strengths and limitations of different methodologies, and establish best practices for developing a robust classification model. This has also been discussed in our manuscript.

In future work, we plan to extend this research by applying the model to a larger, more diverse dataset collected from real clinical settings, which will enhance its generalizability and clinical applicability.

Comment 2: Reproducibility:

Providing open access to the dataset (after addressing ethical considerations) and detailed implementation code or algorithm would enhance the reproducibility of this work.

Thank you for the feedback. Regarding reproducibility, the dataset used in this study was sourced from publicly available online repositories, supplemented by a small number of clinical images collected from the dermatology clinics at PGIMER, Chandigarh, after obtaining ethical approval. **Reference for the publicly available images is provided in the data availability section of manuscript.** However, the clinical images collected from PGIMER contain sensitive patient information and are subject to ethical and privacy restrictions, preventing their public release. Despite this limitation, we have provided a detailed description of our methodology, including the skin disease classification algorithm, which can be tested using publicly available online images.

We remain committed to transparency and reproducibility and will continue to explore ways to share additional resources, such as anonymized datasets, in compliance with ethical guidelines and institutional approvals.

The detailed implementation of the algorithm is provided in the supplementary material, and the revised manuscript includes a mention of these details (Page no 11).

Comment 3: Comparative Analysis:

While the study compares classical, deep learning, and hybrid methods, the rationale for the observed performance differences across models could be better articulated. This analysis would offer insights into model behavior and guide future work.

Thank you for your insightful comment. We have revised the manuscript to include rationale for the observed performance differences across models has been discussed in detail in the Results and Discussion section (Page no 32).

Comment 4: Visualization and Presentation:

Some visualizations, such as performance plots and edge detection examples, need better formatting for clarity. Separating key plots from supplementary material improves accessibility for readers. Also, the figures need to be more visible.

Thank you for the feedback. We have revised the manuscript; all figures have been reformatted for clarity, and the Figures 8, 9, and 10, which contain multiple plots, divided into subparts for better visibility. Only figures that are not referenced or used further have been placed in the supplementary material to enhance accessibility for readers.

[Click here to view linked References](#)1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Autoimmune Blistering Skin Disease Classification: Classical Machine Learning, Deep Learning, and Hybrid Approaches

1
2
3
4
Abstract:
5

6 Contrary to common belief, skin diseases can be life-threatening and have a serious impact on
7 patients' lives. Autoimmune blistering skin diseases (AIBD) of the skin are one of them. The
8 diagnosis based on clinical examination is subjective and irreproducible as it depends upon the
9 expertise and experience of the treating dermatologists. A correct etiological diagnosis of AIBD
10 requires extensive investigations that are costly and time-consuming, leading to delays in
11 confirmation of the diagnosis. Early diagnosis and treatment of AIBD are very important for a
12 better prognosis. Artificial intelligence techniques, particularly convolutional neural networks
13 (CNNs), are being used across several medical domains, including dermatology, and have the
14 potential to assist dermatologists in making early-stage decisions. This paper explores various AI
15 approaches. The proposed models utilized traditional approaches, deep learning techniques, and
16 hybrid methods to classify AIBD by analyzing clinical images. The models were compared using
17 performance parameters and their shortcomings and potential improvement strategies were
18 analyzed based on the experimental results. In the traditional approach, the random forest attained
19 the highest accuracy of 57%, while a hybrid model, Xception integrated with support vector
20 machine (SVM), achieved an overall accuracy of 81%. However, in deep learning techniques,
21 certain models performed very well and showed accuracy that ranged from 84% to 98%,
22 particularly after additional fine-tuning processes.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5.1. Introduction
6

7 Dermatological disease may have major consequences in the lives of the sufferers. The skin, as
8 the body's largest organ, shields against trauma. Apart from protection from the external
9 environment, it performs many functions such as regulating body temperature, synthesis of vitamin
10 D, maintaining fluid-electrolyte balance, etc. autoimmune blistering skin diseases (AIBD) are rare
11 skin conditions in which the immune system that normally protects us, mistakenly attacks the skin
12 and causes blisters and erosions. Several types of AIBD exist and they characterized based on the
13 presence of autoantibodies against specific target antigens. Diagnosis relies on clinical features
14 and laboratory investigations- basic and advanced. The first step in the laboratory diagnostic
15 process is detection of presence of tissue-bound autoantibodies by direct immunofluorescence.
16 Identification of target antigens against which the antibodies are formed require enzyme-linked
17 immunosorbent assay (ELISA), dermatology mosaic (BIOCHIP) or immunoblot. However, these
18 extensive investigations require specialized facilities and expertise that are available only in a few
19 designated research institutions. Moreover, these tests are cost-intensive and time-consuming. The
20 limited availability of these investigations for certain patients due to financial constraints or the
21 lack of facilities in certain regions can result in delays in confirming a diagnosis. Consequently,
22 this can result in treatment postponements and significantly impact both patients' outcomes and
23 their overall quality of life. Diagnoses that are solely based on clinical examination are subjective,
24 irreproducible, and prone to incorrect results due to overlapping clinical presentations. The
25 correctness of clinical diagnosis depends hugely on the expertise and experience of the treating
26 dermatologists.

27 The chances of misdiagnosis of skin diseases are very high at primary care centres due to lack of
28 dermatologists. This limitation makes the identification of skin diseases very challenging and
29 usually causes delays in the diagnosis process. Early diagnosis of AIBD is very important for
30 timely treatment intervention, and prevention of further progression. AI (artificial intelligence)
31 techniques have made significant contributions in various medical fields, including dermatology.
32 Several machine learning and deep learning techniques are being applied for the classification of
33 skin diseases, consistently demonstrating good results.

34 In this paper, we used a comprehensive approach for the development of AI models using multiple
35 techniques, i.e. classical machine learning, deep learning, and hybrid approaches. Each approach
36 has its own advantages and shortcomings.

37 Classical machine learning requires domain expertise, human intervention, and specific algorithms
38 to manually extract relevant features from the input. The selection of important features from the
39 input dataset is a very difficult task as it requires a good understanding of the domain and
40 knowledge of different feature extraction methods [1]. The classical machine learning models are
41 more suitable where we have a limited dataset and require limited computation resources. In this
42 approach, we utilized a random forest classifier (RFC) and support vector machine (SVM).

1
2
3
4 Deep learning, which is another approach of machine learning, is based on artificial neural
5 networks (ANN). An ANN comprises many artificial neurons arranged into an input layer, hidden
6 layers (the number of these layers depends upon the specific task) and finally, an output layer. A
7 neural network with one or two hidden layers is considered a shallow network, whereas a deep
8 neural network consists of several hidden layers for learning complex mapping from large input
9 dataset. Deep neural networks offered an end-to-end solution by enabling the models to
10 automatically extract relevant features present in input data without any requirement of any domain
11 experts or specific algorithms. Various deep learning models have been developed over time.
12
13

14 A convolutional neural network (CNN) is a subtype of deep learning architecture specifically used
15 for tasks related to image analysis, including classification, object detection, and image
16 segmentation. Convolutional neural networks (CNNs) consist of multiple layers, and the most
17 commonly used layers are convolutional layers, pooling layers, and fully connected layers at the
18 end. In addition to these basic layers, a CNN may also include, normalization layers, regularization
19 layers, and skip connections to make the network more suitable for a particular task. Generally,
20 CNNs stack multiple convolutional layers, often accompanied by activation and pooling layers
21 [2]. However, the precise configuration, such as the specific order or arrangement, number, and
22 type of layers required for a CNN, can vary significantly and depend upon several factors, such as
23 the specific task, type of dataset, and available computational resources. Designing and training a
24 CNN from scratch necessitates a sizable dataset to learn meaningful representation and expertise
25 to decide the number, arrangement, types of layers, choice of hyperparameter, regularization
26 technique, and other factors to ensure its effectiveness [3]. It is an iterative process of experiment,
27 evaluation, and continuous refinement. The preparation of a large dataset, which is a prerequisite,
28 is very difficult to obtain, specifically in the medical domain, and it requires significant resources,
29 time, and effort for data collection and labeling. Numerous pre-trained CNNs are currently
30 available, and the techniques of transfer learning using these networks are widely used in the field
31 of the medical domain [4-8]. Large-scale datasets, such as ImageNet, have been used to train most
32 of these networks [9]. These networks are already trained to learn a wide range of visual patterns
33 and features present in the input dataset. Instead of designing and training a CNN from scratch, we
34 can utilize these pre-trained networks for the extraction of useful features from small datasets.
35 Transfer learning has proven effective in classifying medical images [10-12]. In our work, we
36 utilized pre-trained models.
37
38

39 In the hybrid approach, we integrated both classical machine learning and deep learning methods
40 to construct a comprehensive solution. A hybrid model utilizes the strengths while mitigating the
41 limitations of an individual technique in order to achieve high performance. The development of
42 artificial-based methods for automatic diagnosis of images can assist dermatologists in making
43 decisions.
44
45

46 A dataset of 100 AIBD images was created for this work, sourced from both an online database
47 [13] and a real clinical setting. We started this work as a foundational step for the creation of a
48 practical model that can be used in a real clinical setting. We are in the process of gathering and
49

1
2
3
4 annotating clinical images of AIBD from PGIMER, Chandigarh, specifically from the Department
5 of Dermatology OPD. Concurrently, we have initiated the process of model development using
6 the aforementioned techniques.
7

8 We assessed these approaches using evaluation metrics and gained valuable insights into their
9 strengths and weaknesses. The main objective of this work was to identify the most appropriate
10 approach and best practices for developing the final model. Experimental results showed that some
11 pre-trained models achieved accuracies of over 80%, and with additional fine-tuning, their
12 accuracy could be improved further, up to 98%. The problems and solutions encountered during
13 this work offer valuable strategies and insights for other researchers in similar domains.
14

15 The main contributions of this study are as follow.
16

- 17 • Creation of a dataset: After obtaining ethical approval, a dataset was created by including real
18 clinical images alongside publicly available images of bullous pemphigoid (BP), pemphigus
19 vulgaris (PV), and pemphigus foliaceus (PF).
- 20 • Feature extraction: Features were extracted using a gabor filter bank and various edge detection
21 filters.
- 22 • Classical machine learning: Classical machine learning algorithms were employed for the
23 classification of BP, PV, and PF.
- 24 • Implementation of CNN Models: Five distinct pre-trained CNN models with varying hyper-
25 parameters, focal loss function, and optimizers were implemented, resulting in 25 unique CNN
26 architectures. These architectures were analyzed to select the top-performing models, which
27 were then further fine-tuned to improve accuracy. Overfitting was addressed through data
28 augmentation, dropout layers, and the focal loss function.
- 29 • Integration of machine learning approaches: Classical machine learning models were
30 integrated with deep learning models.
- 31 • Comparative analysis: A systematic comparative analysis was conducted to evaluate classical
32 machine learning, deep learning, and hybrid approaches based on various performance metrics.

33 The remainder of this paper is organized as follows: related work is discussed in Section 2,
34 materials and methods are described in Section 3, the results are presented in Section 4, and finally,
35 Section 5 concludes with the discussions.
36

37 2. Related Work 38

39 Sobel and Feldman [14] presented a traditional approach that is based on support vector machine
40 for the classification of skin cancer. The proposed model was tested on the ISIC2017 and HIS2828
41 datasets. The model showed accuracy of 88.10% and 82.11% on these datasets, respectively.
42

43 AL-Enezi [15] proposed a hybrid method for categorizing melanoma, eczema, and psoriasis. They
44 prepared a dataset of 80 images from online resources. Alex Netwas used to extract relevant
45

1
2
3
4 features, and SVM algorithm was used for classification. The experimental design showed an
5 accuracy of 100% across all disease categories.
6

7
8 In their paper, Mohammed et al. [16] used classical machine learning for the classification of
9 benign vs. malignant lesions from 400 dermoscopic images (200 benign and 200 malignant)
10 collected from the PH2 database. They used the Active Contour algorithm for segmentation, and
11 then from segmented images, statistical features were extracted. These features were subsequently
12 optimized using NSGA-II. Artificial neural network (ANN) and support vector machine (SVM)
13 were then trained on the selected features. The results showed that the ANN achieved the best
14 results.
15

16 Dhivyaa et al. [17] developed a skin lesion classification model using a conventional approach.
17 They used decision trees and random forest algorithms. All the images underwent pre-processing,
18 which includes segmentation using thresholding-based methods, edge and contour methods, as
19 well as region-based segmentation algorithms. After segmentation, features such as asymmetry,
20 border irregularity, color, and diameter (ABCD) were extracted. The model was tested using the
21 ISIC 2017 and HAM10000 skin lesion datasets, and the results showed an accuracy rate of 97%.
22

23 Senan et al. [18] presented an algorithm that utilized image processing and machine learning
24 techniques for the early diagnosis of melanoma from dermoscopic images. They used the PH2
25 dataset, which contains 200 images of melanocytic lesions. All images were pre-processed to
26 remove unwanted noise using a Gaussian filter, and subsequently, active contour-based methods
27 were used for segmentation, followed by, morphological operations to refine the results. The
28 algorithm then utilized a fuzzy color histogram (FCH) for feature extraction. Four classifiers
29 namely SVM, K-NN, ANN and FFNN were implemented. The results showed that all classifiers,
30 except SVM, achieved 100% accuracy.
31

32 Kassem et al. [19] discussed various skin image databases and techniques for classification of skin
33 diseases using traditional and deep learning approaches.
34

35 Han et al. [20] proposed web data augmentation technique to address the problem of small dataset
36 sizes. During the augmentation, a seed image from the dataset was selected for retrieving similar
37 images from the Google search engine, and the images with high similarity from the results were
38 added to the training dataset. Subsequently, a pre-trained network was further fine-tuned on the
39 augmented dataset.
40

41 Gavrilov et al. [21] utilized pre-trained Inception v3 for the classification of benign vs. malignant
42 skin tumors. A dataset was prepared from the ISIC (International Skin Imaging Collaboration)
43 Archive. In this study, five identical networks with different weights were combined to form an
44 ensemble network, which made decisions based on the majority voting principle. The proposed
45 model achieved an AUC-ROC of 0.96.
46

Pangti et al. [22] utilized a CNN for the classification of 40 dermatological diseases, including bullous pemphigoid. The dataset comprised 17,408 images collected from Indian dermatologists and public databases. They developed an app, and the results showed a top-1 and top-3 accuracy of 75.07% and 89.62% respectively in a real clinical scenario, which included 5,014 patients from different categories of healthcare facility settings.

Sriwong et al. [23] proposed three distinct strategies for categorizing skin conditions based on dermoscopic images. In the first strategy, a pre-trained CNN was used for the classification. In the second approach, which is hybrid, pre-trained CNN was used for feature extraction, and SVM was used for the final classification. The third approach, which is similar to the second, uses additional patient metadata along with features that were extracted from the pre-trained CNN.

Guha and Rafizul Haque [24] used three different models, support vector machine (SVM), VGG, and Inception-ResNetv2 for classification of seven types of skin diseases. Their performance was evaluated, and experimental results showed that the Inception-ResNet-v2 outperformed the other two models.

Janoria et al. [25] presented a hybrid approach for classification of melanoma and benign tumors. They extracted features using pre-trained CNN using the transfer learning approach. These features were then used to train classical machine learning classifiers, including ensemble methods. The proposed methods were tested on a publicly available dataset, and KNN achieved the highest accuracy of 99%.

Nair et al. [26] used CNN to extract features from gangrene disease images, which were then employed to train an SVM for classification. This hybrid approach achieved 85% accuracy.

Purwar et al. [27] presented a hybrid approach for the detection of thalassemia, pre-trained CNN was used for the extraction of features from the blood smear images of 20 patients, and these features were combined with clinical features. PCA was used for the features section, and the reduced set of features was used for the training of naive bayes, random forest (RF), and K-NN classifiers. They obtained an accuracy of 99% using both KNN and naive bayes.

Palivelal et al. [28] proposed a model for segmentation and classification of skin lesions using a combination of image processing and deep learning techniques. Initially, the images underwent pre-processing which performed resizing and equalization. Next, the images were segmented using Watershed transformation. The segmented results then served as training data for a U-Net model. The model's performance was evaluated on the ISIC dermoscopic image datasets, which comprised a substantial number of 23,000 images. However, the results indicated the presence of overfitting, evident from a substantial disparity between the training and validation accuracy.

Islam et al. [29] combined pre-trained CNN with recurrent neural networks to diagnose COVID-19 from a dataset of 6939 chest x-rays categorized into COVID-19, pneumonia, and normal. They found that VGG-19-RNN achieved the best performance of AUC (area under the curve) of 99.99%.

1
2
3
4 Jimenez-del-Toro et al. [30] conducted a review on the analysis of histopathology images, covering
5 both traditional and deep learning approaches, and highlighted the challenges involved in the
6 processing of whole slide images (WSI).
7

8 Singh et al. [31] discussed AIBD and explored current research on the diagnosis of skin diseases
9 using AI techniques. In their review, they examined methods for the segmentation and
10 classification of skin diseases utilizing both traditional machine learning and deep learning. The
11 review highlighted various studies that achieved good accuracy using deep learning and transfer
12 learning techniques. We further extended this study by applying traditional machine learning, deep
13 learning, and hybrid approaches to classify AIBD using clinical images.
14

15 Although AI has been extensively applied to various skin diseases and numerous machine learning
16 models have been developed and clinically validated, its application to Autoimmune Blistering
17 Diseases (AIBD) is limited due to non-availability of datasets. Karen et al. [32] presented and
18 summarized studies that employed AI to diagnose Autoimmune Blistering Diseases (AIBD).

19 Schielein et al. [33] collected images of bullous pemphigoid, along with other skin diseases, for
20 binary classification, distinguishing between normal and outlier conditions for each class using
21 pretrained CNN.
22

23 Ahmed et al. [34] applied classical machine learning models and Convolutional Neural Networks
24 (CNN) for the classification of Pemphigus Vulgaris, a type of Autoimmune Blistering Disease
25 (AIBD). Using a dataset of 158 images collected from online sources, the CNN achieved an
26 accuracy of 93%, while classical machine learning models achieved an accuracy range of 70-80%.
27

28 In a similar study, Dubey et al. [35] employed CNN and machine learning models on a dataset of
29 262 images of pemphigus vulgaris, which were collected from online resources.
30

31 **3. Materials and Method**

32 This section mainly consists of three parts: datasets, data augmentation techniques, and model
33 development approaches.
34

35 **3.1 Dataset**

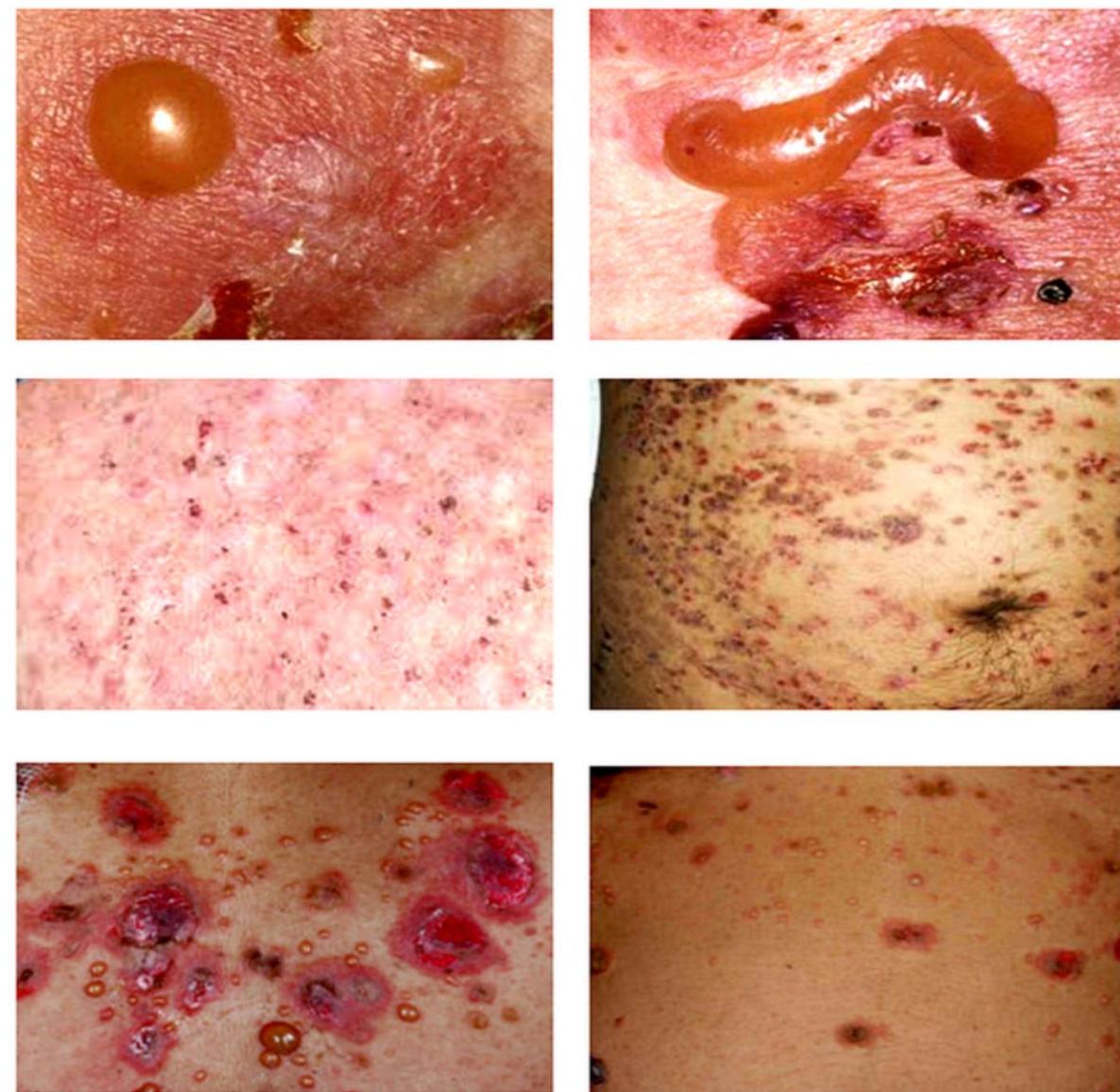
36 There are several databases available on skin diseases. Some of the commonly used databases
37 include: DermNet NZ, PubMed, Skin Disease Database (SKINDIS), Online Mendelian
38 Inheritance in Man (OMIM), DermIS, American Academy of Dermatology (AAD) Dermatology
39 Image Database, VisualDx, ISIC (International Skin Imaging Collaboration) Archive,
40 HAM10000, PH2 Dataset, and Dermofit Image Library.
41

42 To my knowledge, there isn't a publicly available dataset that specifically focuses on autoimmune
43 blistering skin diseases due to the rarity of such diseases. For our pilot study, we collected images
44 of autoimmune blistering skin disease (AIBD) from DermNet NZ and a few images from the
45 Dermatology OPD of PGIMER Chandigarh. This dataset comprises a total of 100 images, with 46
46 47

1
2
3
4 images of bullous pemphigoid (BP), 30 images corresponding to pemphigus vulgaris (PV), and 22
5 images associated with pemphigus foliaceus (PF). To ensure the dataset's quality and clarity, we
6 have removed any images that were unclear. Fig. 1 shows some sample images from the dataset.
7
8

9 It is important to acknowledge that this dataset does not capture the entire range or diversity of
10 AIBD conditions. However, it serves as an initial step, acting as a foundation for utilizing AI for
11 the identification of AIBD.
12
13

14 Considering the limited number of images available in our dataset, the accuracy of our machine
15 learning model could be affected by several factors, and dataset size is a crucial parameter. The
16 optimal dataset to train a deep network depends on various parameters, such as the difficulty level
17 of the task, the chosen learning algorithm, and bayes error.
18
19

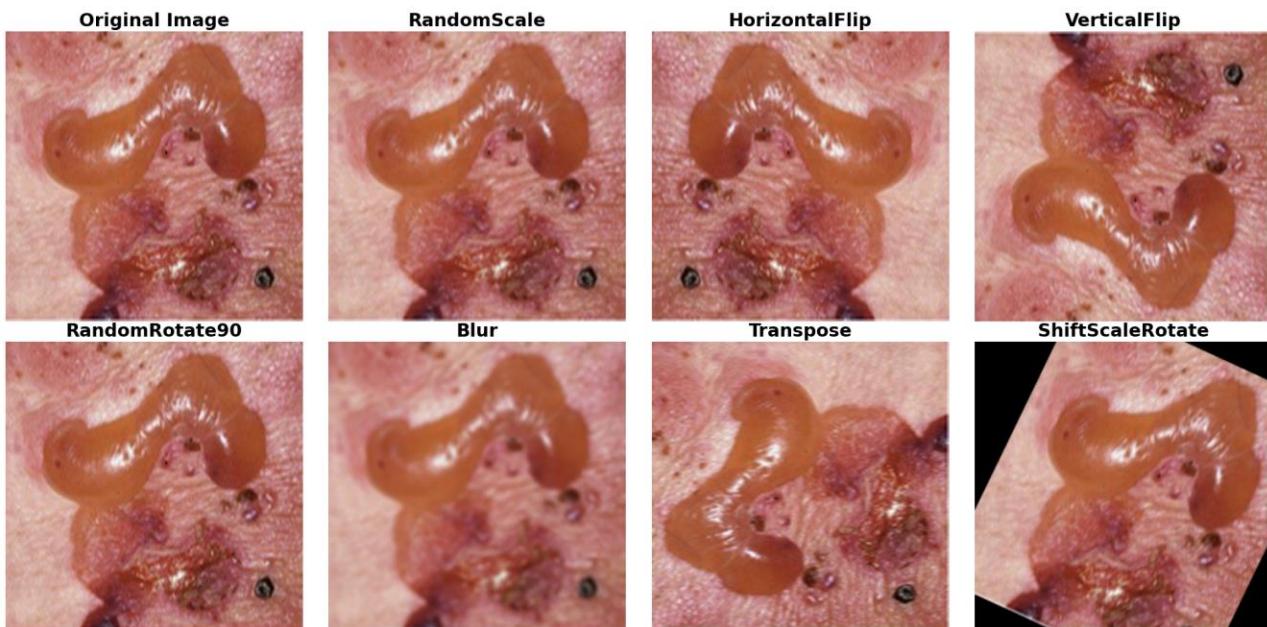


60 Fig. 1 Sample images from the dataset
61
62

1
2
3
4 **3.2 Data augmentation:**
5

6 Data augmentation techniques are a valuable strategy, particularly when working with limited
7 datasets. These techniques apply various transformations to the existing images and effectively
8 increase the dataset size. In our study, we utilized a Python library called Albumentations [36],
9 which provides a range of transformation operations. This library supports probabilistic image
10 transformations, enabling random variations for each image. Clinical images captured using
11 cameras can vary due to factors such as changes in lighting, orientation, noise, camera movement.
12 Probabilistic transformations simulate real-world scenarios by introducing a wide range of
13 variations that aren't predefined, thereby aiding model generalization.
14
15

16 When training a neural network on an imbalanced dataset, where certain classes have a
17 significantly larger number of samples than others, it can lead to biased learning. To address this
18 issue, we employ a strategy that mitigates the impact of class imbalance. We applied selective
19 augmentation to balance class distributions. Additionally, we used focal loss to address class
20 imbalance from a model perspective. During training, this loss function dynamically adjusts
21 weights to focus more on hard-to-classify examples, resulting in improved performance. The
22 transformations we applied, along their probabilities, include randomscale, horizontalflip,
23 verticalflip, randomrotate90, blur, transpose, and shiftscalerotate. Fig. 2 shows the results of these
24 transformations on a sample image.
25
26



53 Fig. 2 Results of augmentation transformations
54
55
56
57
58
59
60
61
62
63
64
65

3.3 Model development approaches

In this paper, we utilize three distinct methods for model development, including classical machine learning, deep learning (specifically CNN), and a hybrid techniques. Figure 3 illustrates these approaches. A detailed implementation algorithm is provided in the supplementary material.

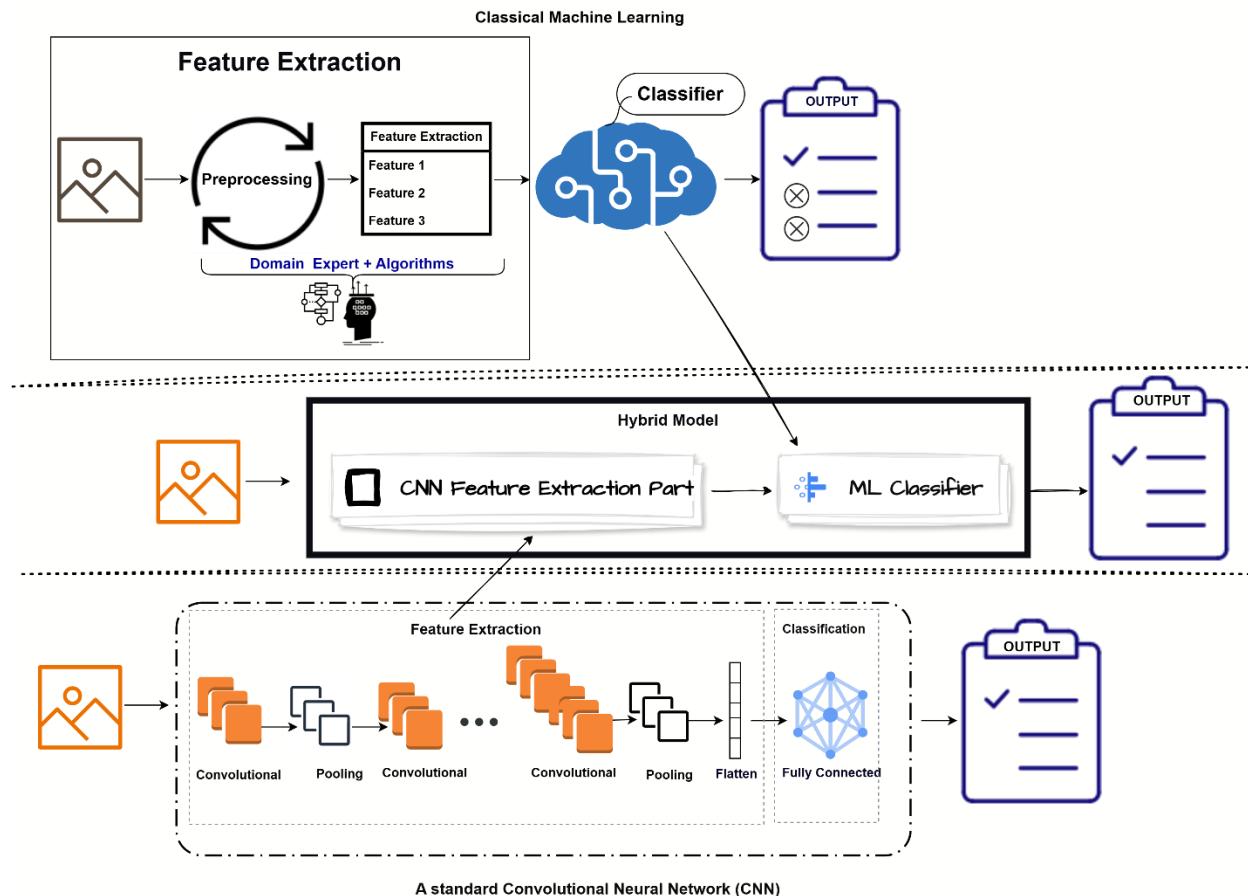


Fig. 3 Classical machine learning, hybrid model, and CNN architecture

3.3.1 Classical machine learning

Classical machine learning approaches are particularly well-suited for small dataset sizes. The main steps involved in this approach are outlined below.

i. Pre-processing:

Prior to training the model, a crucial step involved enhancing and standardizing the collected dataset for subsequent model training, which enables the model to learn from a high-quality dataset. Due to variations in image capturing devices, resolutions, and lighting conditions, the dataset may contain noisy images of different sizes. Therefore, pre-processing techniques were employed to remove various forms of noise and artifacts, as well as to standardize pixel values. In

1
2
3
4 our study, we used contrast limited adaptive histogram equalization (CLAHE) to minimize the
5 effects of different lighting conditions and contrast. CLAHE, which is based on adaptive histogram
6 equalization, addresses the issue of over-amplification of contrast [37].
7
8

9 ii. Feature extraction:
10

11 The next crucial step involved extracting features from the dataset. These features capture the
12 distinctive visual content within the images, enabling differentiation between different diseases.
13 The extraction of appropriate features is critical, as it strongly influences the performance of the
14 classification model.
15
16

17 As images in a dataset may vary in orientation and complexity, a single filter may not capture all
18 relevant information. We utilized a combination of filters, including gabor filter bank, canny,
19 sobel, prewitt, roberts, scharr, gaussian, and median filters, to extract diverse features from the
20 images. This approach created a feature vector that highlights different aspects such as textures,
21 edges, and gradients.
22
23

24 Multiple filters were used for edge detection, as each filter designed to highlight edges in specific
25 orientations (vertical, horizontal, or diagonal) and the effectiveness of these filters also depends
26 on several factors such as noise in the image, edge thickness, and the overall complexity of the
27 edges.
28
29

30 For example, the canny filter works best for images having low noise and well-defined edges. The
31 scharr filter can detect edges in images with higher noise levels, making it useful for noisy or less
32 defined images. The sobel filter emphasizes vertical and horizontal edges, while the roberts filter
33 is effective at detecting diagonal edges. The combination these features increased the likelihood
34 of capturing relevant edge information across diverse image conditions. Figure 4 displays the
35 results of edge detection methods applied to two different images.
36
37

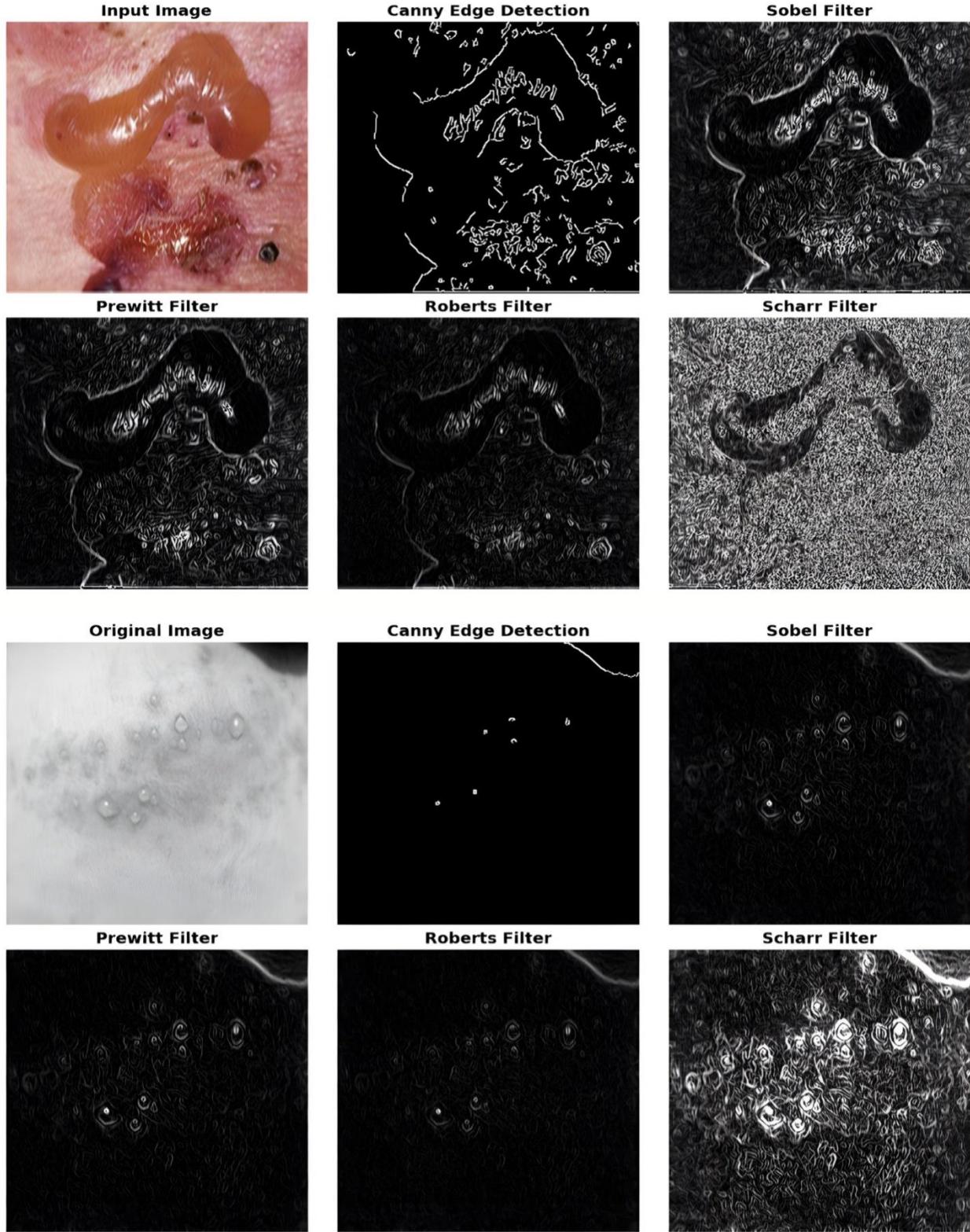


Fig. 4 Edge detection results using filters on two images

1
2
3
4 The filters we used are detailed below:
5
6 a. Gabor filter
7

8 We utilized the gabor filter to extract texture and edge features from the skin disease. This filter's
9 ability to analyse and highlight texture variations enabled it to capture important textural
10 information present in the images [38-39].
11
12

13 b. Canny edge detection
14

15 This is a multistep filter based on thresholds, designed to extract precise edge features. This filter
16 is known for its ability to provide precise and well-connected edges, making it suitable for
17 applications that require high-quality edge detection results [40].
18
19

20 c. Sobel filter
21

22 The sobel filter is computationally efficient and performs reasonably well in detecting edges,
23 especially in relatively smooth images, but it may not capture fine details as effectively as
24 compared to more advanced edge detection techniques.
25
26

27 d. Prewitt filter
28

29 The prewitt filter is another image processing technique commonly used for edge detection, similar
30 to the sobel filter. However, this filter uses a slightly different kernel (-1, 0, 1 for the horizontal
31 gradient and -1, 0, 1 for the vertical gradient) compared to the sobel filter. Like the sobel filter, it
32 may not be as effective in capturing fine details.
33
34

35 e. Roberts filter
36

37 Roberts filter is another edge detection algorithm. This filter used a pair of small kernels of size 2
38 x 2 pixels that are rotated 90 degrees from one another [41], as shown in Fig. 5. Due to the small
39 size of the kernels, it is computationally efficient in capturing coarse edge features and is sensitive
40 to noise.
41
42

43

0	1
0	-1
-1	0
1	0

44
45 Fig. 5 Roberts filter kernels
46
47
48
49
50
51
52
53

54 f. Scharr filter
55

56 Scharr filter, due to its optimized kernel, captures edge features with more accuracy compared to
57 other edge detection filters such as prewitt, sobel, and roberts filters [42-43]. It is particularly useful
58
59
60
61
62
63
64
65

1
2
3
4 in scenarios where precise edge localization is required and performs well even in the presence of
5 noise.
6

7 g. Gaussian filter
8
9

10 It is a linear filter that is commonly used to blur images and eliminate noise. This filter convolved
11 the image with a two-dimensional gaussian function. The gaussian filter is not only used for noise
12 reduction but is also useful for edge detection purposes. In the literature, numerous methods of
13 image processing based on gaussian principles are listed [44].
14

15 h. Median filter
16
17

18 The median filter is a non-linear method employed to eliminate impulse and salt and pepper noise
19 from an image. In this filter, kernel slides over the entire image and during hovering, the pixel
20 being processed is replace by the median of the gray levels in neighborhood of that pixel. This
21 filter can also be applied to color images [45]
22

23 iii. Training of models
24
25

26 After pre-processing and extraction of relevant features using the techniques mentioned above,
27 random forest classifier (RFC) and support vector machine (SVM) were trained on these features.
28 These classifiers are very powerful classical machine learning algorithms.
29

30 The RFC is an ensemble of multiple decision trees that employs the concept of bagging for
31 classification. It is known for its speed, ease of implementation, and its capability to access the
32 significance of each feature [46-47].
33

34 The SVM constructs the hyperplanes using support vectors (data points that are closer to the
35 hyperplane) with the maximum margin for the classification. The hyperplanes are decision
36 boundaries that classify data points, and their dimension depends on the number of features. SVM
37 can also be used to learn nonlinear mappings present in the data by using appropriate kernel
38 functions [48].
39

40 **3.3.2 Deep learning**
41
42

43 Deep Learning, which is based on learning data representations, has capability to automatically
44 learn relevant hierarchical features from input data using multiple layers.
45

46 Initially, we started by designing CNN architecture from scratch and conducted various
47 experiments by tuning multiple hyper-parameters including number of layers and neurons,
48 learning rate, loss function, and optimizer. However, despite these efforts, the model did not
49 demonstrate any significant improvement. The main reason for the lack of improvement was the
50 limited size of the dataset.
51

52 Building a CNN from scratch is time-consuming and challenging. An alternative approach in deep
53 learning is transfer learning. Transfer learning involves utilizing models that have been pre-trained
54

on large datasets. The optimized weight of these models can be applied to new but related tasks. Our dataset consists of clinical images, which differ significantly from natural images. Surprisingly, transfer learning from natural images to medical images has proven to be feasible, despite the inherent dissimilarity between the source and target domains [49]. The knowledge obtained from pre-existing natural images can be effectively transferred and applied to enhance the performance of CNNs in the analysis of medical images. Transfer learning approaches provide opportunities to overcome data limitations and enhance model accuracy in the field of medical image analysis [50].

There are various ways in which a pre-trained model can be repurposed for different tasks. Pretrained models can be used to extract features from a target dataset, which can then be utilized to train a classifier. Alternatively, the architecture of pretrained models, whether initialized with random weights or pretrained weights, can be employed. The weights of these layers can be further adjusted for the target dataset through a process known as fine-tuning.

Typically, fine-tuning a pre-trained model involves freezing certain layers, specifically the earlier ones that capture general features that are applicable across different tasks, and adjusting the last few convolutional layers responsible for extracting features relevant to the new task. The necessity to fine-tune additional layers depends on several factors, including the volume of data in the target dataset, the similarity between the target and source datasets, and the available computational resources. If there is significant dissimilarity between the tasks, it may be necessary to fine-tune more than just the final layers. The fine-tuning process begins with the last layers and gradually progresses to earlier layers until the model achieves the desired performance. This iterative process involves experimentation and refinement based on the model's performance. Adjustments to layer weights enable the model to adapt pre-trained parameters to new data.

The fully connected layers and the final classification layer of the pre-trained model were replaced with a global average pooling layer and a new layer consisting of three neurons and a SoftMax function to provide classification probabilities for bullous pemphigoid (BP), pemphigus vulgaris (PV), and pemphigus foliaceus (PF). We can also experiment with using fully connected layers instead of the GAP layer. The GAP layer reduces the total number of parameters and reduces the overfitting problem. To further prevent overfitting, a dropout rate of 0.4 was applied. Dropout, a regularization technique, randomly deactivates neurons during training to reduce reliance on specific neurons. Additionally, early stopping was used as another regularization technique to avoid unnecessary training and overfitting.

Initially, during training of CNN using categorical cross-entropy loss function, we observed that the models performed reasonably well in terms of accuracy, but the loss value was not satisfactory. This indicated that the models were struggling to classify certain examples, leading us to suspect potential issues such as overfitting due to the limited size of our dataset, an imbalanced class distribution, or the complexities of learning the underlying patterns.

To improve model performance, we opted for focal loss function which is specifically used to handle class imbalance. It adjusts the loss value based on the difficulty of the data, focusing more on challenging examples that are often misclassified. The utilisation of the focal loss function improved our models' performance. The focal loss function has two parameters: alpha (α) and gamma (γ) as shown in equation 1. Alpha (α) addresses the problem of class imbalance, while Gamma (γ) controls the emphasis given to hard examples during training.

$$\text{Focal Loss} = -\alpha(1 - p_t)^\gamma \log p_t \quad (1)$$

The focal loss minimizes the loss for well-classified examples as it is not necessary to focus on such examples where the model predicts a high probability value. However, it increases the loss for hard-to-classify examples, encouraging the model to make further efforts to minimize it as much as possible.

Learning rate, which is a hyperparameter, was set to 0.0001. The learning rate defines the extent of parameter adjustments made by the model during each iteration. We selected 300 epochs to observe the trend of accuracy and the loss curve during training. By analysing the accuracy and loss values, we identified the epoch at which the model begins to converge and detected signs of overfitting. Making an informed decision based on these curves helped us identify more suitable hyperparameters.

We conducted two experiments using the transfer learning approach. In the first experiment, we used the five pre-trained networks solely for features extraction. We then performed an evaluation and comparison of these pre-trained models to identify the most fitting one, while in second experiment, after selecting the appropriate models, we proceeded with the process of fine-tuning.

To analyze the performance of different optimizers, we tested all pre-trained models with five different optimizers; Adam (Adaptive Moment Estimation), RMSprop (Root Mean Square Propagation), Adadelta, Adamax, and Nadam. This resulted in a total of 25 unique model configurations. An optimizer updates the model's weights during training to minimize the loss function.

After evaluation, top 12 models were identified for additional fine-tuning to assess the extent of improvement. During this process, we strategically unfroze only the last two convolutional layers of the pre-trained networks due to the limited dataset and used a low learning rate of 0.00001. A low learning rate prevented abrupt changes to the pre-trained weights, allowing the model to gradually adjust its weights while preserving valuable learned knowledge. Additionally, it enabled the model to carefully explore all local minima and refine its representations based on the target task. The experimental results showed significant improvement after fine-tuning.

All experiments were conducted using TensorFlow 2.0 and the Keras framework, both of which are open-source software libraries for machine learning tasks. jupyter notebook with an Anaconda environment was used for coding and execution. All experiments were performed on a robust hardware setup consisting of an Intel Xeon Gold 5218R processor accompanied by RAM

capacity of 128 GB. In addition to that, a NVIDIA Quadro RTX 4000 GPU with 7.5 compute units is utilised for accelerating computation through the utilisation of CUDA (Compute Unified Device Architecture) technology. For implementation of CUDA, we employed the CUDA toolkit in combination with the cuDNN (CUDA Deep Neural Network) library.

The pre-trained networks that we utilized include VGG19, ResNet50, Xception, Inception-ResNet V2, and DenseNet.

a. VGG19

The design of VGG is simple and straightforward when compared to other deep learning architectures. It comprises convolutional layers, along with pooling layers and fully connected layers [51]. In contrast to ResNet and Xception, VGG has large number of parameters; therefore, training requires significant computational resources.

b. ResNet50

The power of a deep neural network can be increased by adding more layers to the network. But beyond some point, increasing the depth of a network does not provide any benefits.

As the network becomes deeper, it becomes more challenging to train it due to issues like the vanishing gradient problem. To address the problem of vanishing gradient, ResNet (Residual Network) was introduced by [52], utilises skip connections, or residual connections. The skip connections allow the network to pass the input or intermediate representation directly to a deeper layer (instead of the next layer) in the network by skipping one or more layers.

c. Inception-ResNet V2

Instead of single-size filter, the Inception-ResNet V2 model uses multiple filters of different sizes along with skip connections in its architecture. The multiple filters capture information at different scales and learn both local and global features [53].

d. Xception

The Xception, or Extreme Inception, developed by François Chollet is an extension of the Inception modules. It uses depth-wise separable convolutions instead of standard convolutional layers to further reduce the number of parameters and computational complexity. This architecture consists of a linear sequence of depth-wise separable convolutional layers with skip connections [54].

e. DenseNet

DenseNet employs the concept of dense blocks, where all layers within each dense block are connected in such a way that each layer receives input from all the previous layers, promoting

1
2
3
4 strong feature reuse. This design ensures that the feature map dimensions remain constant within
5 each dense block. These dense blocks are connected with transition layers, which reduce the spatial
6 dimensions and the number of feature maps. Notably, DenseNet has fewer parameters compared
7 to ResNet, as it utilises concatenation to combine features from the previous layers, whereas
8 ResNet uses addition [55].
9
10

11 3.3.3 Hybrid Approach:

12

13 A hybrid approach involves the integration of multiple techniques; in our case, we combined
14 machine learning and deep learning techniques. In this approach, pre-trained models were used to
15 extract features, and instead of passing these features through a global average pooling layer and
16 then a final dense layer, we fed the features to classical machine learning classifiers. A hybrid
17 model can be customised by integrating various key components from different techniques.
18
19

20 4. Results and discussion

21

22 To assess the efficiency of proposed approaches, we calculated precision, recall and F1 score.
23 Additionally, we analysed the trends, which were obtained by the loss and accuracy curves of
24 individual deep learning models, to understand the performance and learning dynamics throughout
25 the training process.
26
27

28 In the classical machine learning approach, RFC achieved higher performance with accuracy of
29 57% while SVM achieved accuracy of 29%. The confusion matrix for the classification of bullous
30 pemphigoid (BP), pemphigus vulgaris (PV), and pemphigus foliaceus (PF) obtained using the RFC
31 is shown in Fig. 6. In the matrix, BP is represented by 0, PV by 1, and PF by 2.
32
33



55 Fig. 6 Confusion matrix obtained by random forest classifier

56 In the hybrid approach, Xception integrated SVM (Xception + SVM) showed the best performance
57 with an overall accuracy of 81%, precision of 80%, recall of 82%, and an F1 score of 81%. Figure 7
58
59

shows the corresponding confusion matrix of the best hybrid model. The performance of all hybrid models for the classification of bullous pemphigoid (BP), pemphigus vulgaris (PV), and pemphigus foliaceus (PF) is summarized in Table 1.

Table 1 Precision, Recall, F1-Score, and overall average accuracy of all hybrid models

Hybrid Model/ Overall Accuracy	Disease classes / Average	Precision	Recall	F1 score
ResNet50+SVM Accuracy: 0.78	BP	0.72	0.72	0.72
	PF	0.86	0.93	0.89
	PV	0.8	0.75	0.77
	Average	0.79	0.80	0.80
ResNet50+ RFC Accuracy: 0.67	BP	0.60	0.72	0.66
	PF	0.70	0.78	0.74
	PV	0.75	0.56	0.64
	Average	0.68	0.69	0.68
VGG19+SVM Accuracy: 0.68	BP	0.57	0.66	0.61
	PF	0.88	0.72	0.79
	PV	0.69	0.67	0.68
	Average	0.72	0.68	0.70
VGG19+ RFC Accuracy: 0.62	BP	0.54	0.64	0.58
	PF	0.58	0.72	0.64
	PV	0.80	0.55	0.65
	Average	0.64	0.64	0.63
Xception+ SVM Accuracy: 0.81	BP	0.82	0.74	0.78
	PF	0.71	0.90	0.80
	PV	0.88	0.82	0.85
	Average	0.80	0.82	0.81
Xception+ RFC Accuracy: 0.71	BP	0.66	0.80	0.72
	PF	0.68	0.66	0.67
	PV	0.81	0.67	0.73
	Average	0.72	0.71	0.71
Inception- ResNetv2+SVM Accuracy: 0.74	BP	0.70	0.74	0.72
	PF	0.76	0.78	0.77
	PV	0.77	0.72	0.75
	Average	0.74	0.75	0.74
	BP	0.59	0.80	0.68

Inception-ResNetv2+ RFC Accuracy: 0.71	PF	0.90	0.81	0.85
	PV	0.76	0.56	0.65
	Average	0.75	0.73	0.73
DenseNet-121+SVM Accuracy: 0.77	BP	0.79	0.74	0.76
	PF	0.67	0.81	0.73
	PV	0.83	0.77	0.80
	Average	0.76	0.77	0.77
DenseNet-121+RFC Accuracy: 0.68	BP	0.68	0.78	0.73
	PF	0.62	0.66	0.64
	PV	0.71	0.60	0.65
	Average	0.67	0.68	0.67



Fig. 7 Confusion matrix obtained by the Xception + SVM

In our deep learning methodology, we initially applied a large epoch value of 300 to all 25 models to observe convergence, performance trends, and variations during the training process.

The loss and accuracy graphs of the pre-trained models provided valuable insights. Specifically, Xception and Inception-ResNetv2 networks showed signs of overfitting, as evidenced by the significant gap between training and validation accuracy. Loss and accuracy curves associated with these models are provided in the supplementary material. In contrast, all models trained with the Adadelta optimizer exhibited no progress in learning, corresponding figures are provided in the supplementary material. However, the remaining models effectively learned up to a certain threshold and finally converged without any further improvement in the results. We observed that

the majority of models, except those trained with the Adamax Optimizer, achieved convergence before reaching 150 epochs. Table 2 summarizes the performance metrics obtained by all pre-trained models.

Table 2 Precision, Recall, F1-Score, and overall average accuracy for the classification of bullous pemphigoid (BP), pemphigus vulgaris (PV), and pemphigus foliaceus (PF) obtained from all pre-trained models.

Model + Optimizer / Average Accuracy	Disease classes	Precision	Recall	F1-Score
ResNet-50 + Adam Accuracy: 0.8803	BP	0.81	0.86	0.83
	PF	0.89	1.0	94
	PV	0.94	0.82	0.88
	Average	0.88	0.89	0.88
ResNet50+ RMSprop Accuracy: 0.8592	BP	0.75	90	82
	PF	1.0	0.96	0.98
	PV	0.89	0.75	0.82
	Average	0.88	0.87	0.87
ResNet50 + Adamax Accuracy: 0.8803	BP	0.80	0.88	0.84
	PF	0.96	0.96	0.96
	PV	0.90	0.82	0.86
	Average	0.89	0.89	0.89
ResNet50 + Nadam Accuracy: 0.8521	BP	0.81	0.76	0.78
	PF	0.94	0.96	0.95
	PV	0.83	0.86	0.84
	Average	0.86	0.86	0.86
VGG19 + Adam Accuracy: 0.8873	BP	0.79	0.92	0.85
	PF	0.96	0.90	0.93
	PV	0.94	0.84	0.89
	Average	0.90	0.89	0.89
VGG19 + RMSprop Accuracy: 0.9014	BP	83	90	86
	PF	1.0	1.0	1.0
	PV	0.90	0.84	0.87
	Average	0.91	0.91	0.91

1	VGG19 + Adamax	BP	0.70	0.78	0.74
2	Accuracy: 0.7958	PF	0.93	0.87	0.90
3		PV	0.81	0.75	0.78
4		Average	0.81	0.80	0.81
5					
6	VGG19 + Nadam	BP	0.80	0.90	0.85
7	Accuracy: 0.8873	PF	1.0	0.93	0.96
8		PV	0.90	0.84	0.87
9		Average	0.90	0.89	0.89
10					
11	Xception + Adam	BP	0.71	0.78	0.74
12	Accuracy: 0.79	PF	0.80	0.87	0.84
13		PV	0.86	0.74	0.79
14		Average	0.79	0.80	0.79
15					
16	Xception + RMSprop	BP	0.73	0.74	0.73
17	Accuracy: 0.7817	PF	0.76	0.90	0.83
18		PV	0.84	0.74	0.78
19		Average	0.78	0.79	0.78
20					
21	Xception+ Adamax	BP	0.76	0.74	0.75
22	Accuracy: 0.7958	PF	0.76	0.90	0.83
23		PV	0.84	0.77	0.81
24		Average	0.79	0.81	0.79
25					
26	Xception+ Nadam	BP	0.81	0.78	0.80
27	Accuracy: 0.8451	PF	0.82	0.96	0.88
28		PV	0.88	0.82	0.85
29		Average	0.84	0.86	0.84
30					
31	Inception-ResNetv2 + Adam	BP	0.80	0.82	0.81
32	Accuracy: 0.8239	PF	0.80	0.96	0.87
33		PV	0.86	0.74	0.79
34		Average	0.82	0.84	0.82
35					
36	Inception-ResNetv2 + RMSprop	BP	0.85	0.78	0.81
37	Accuracy: 0.8239	PF	0.78	0.96	0.86
38		PV	0.83	0.77	0.80
39		Average	0.82	0.84	0.82
40					

Inception-ResNetv2 + Adamax	BP	80	82	81
Accuracy: 0.8169	PF	0.77	0.93	0.84
	PV	0.86	0.74	0.79
	Average	0.81	0.83	0.82
Inception-ResNetv2 + Nadam	BP	0.80	0.82	0.81
Accuracy: 0.8239	PF	0.80	0.96	0.87
	PV	0.86	0.74	0.79
	Average	0.82	0.84	0.82
DenseNet-121 + Adam	BP	0.78	0.88	0.83
Accuracy: 0.8662	PF	0.91	0.96	0.94
	PV	0.92	0.79	0.85
	Average	0.87	0.88	0.87
DenseNet-121 + RMSprop	BP	0.76	0.94	0.84
Accuracy: 0.8662	PF	1.0	0.81	0.90
	PV	0.92	0.82	0.87
	Average	0.89	0.86	0.87
DenseNet-121 + Adamax	BP	0.73	0.90	0.80
Accuracy: 0.8451	PF	0.96	0.84	0.90
	PV	0.92	0.79	0.85
	Average	0.84	0.84	0.85
DenseNet-121 + Nadam	BP	0.75	0.84	0.79
Accuracy: 0.8451	PF	0.96	0.81	0.88
	PV	0.87	0.86	0.86
	Average	0.86	0.84	0.85

Among these 25 models, ResNet50, VGG19, and DenseNet-121, showed promising results, were selected for further refinement through the fine-tuning process. Figures 8a, 8b, and 8c illustrate the effective convergence of ResNet50, VGG19, and DenseNet-121 without overfitting.

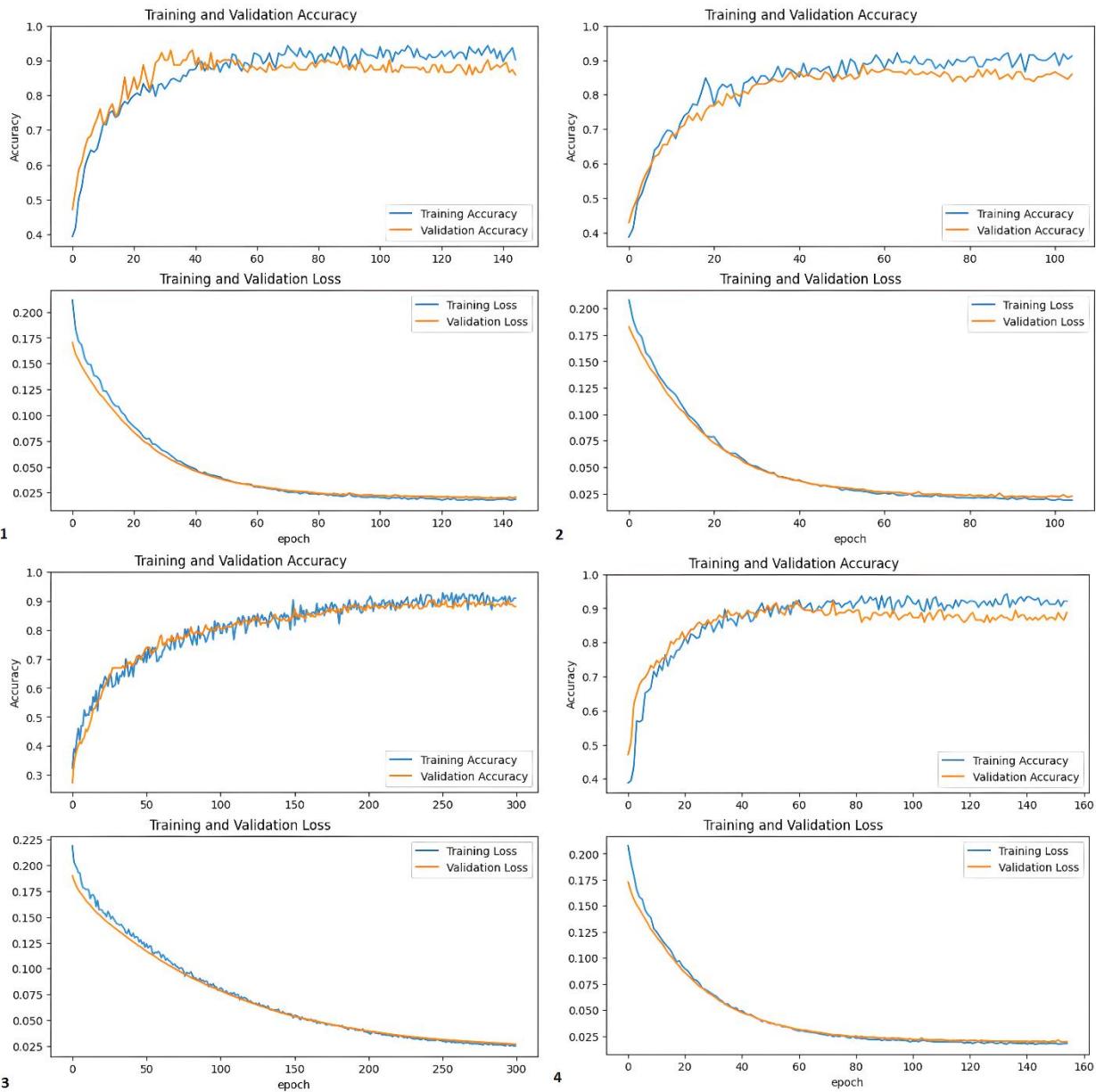


Fig. 8a: Performance curves for ResNet50 using (1) Adam, (2) RMSprop, (3) Adamax, and (4) Nadam optimizers.

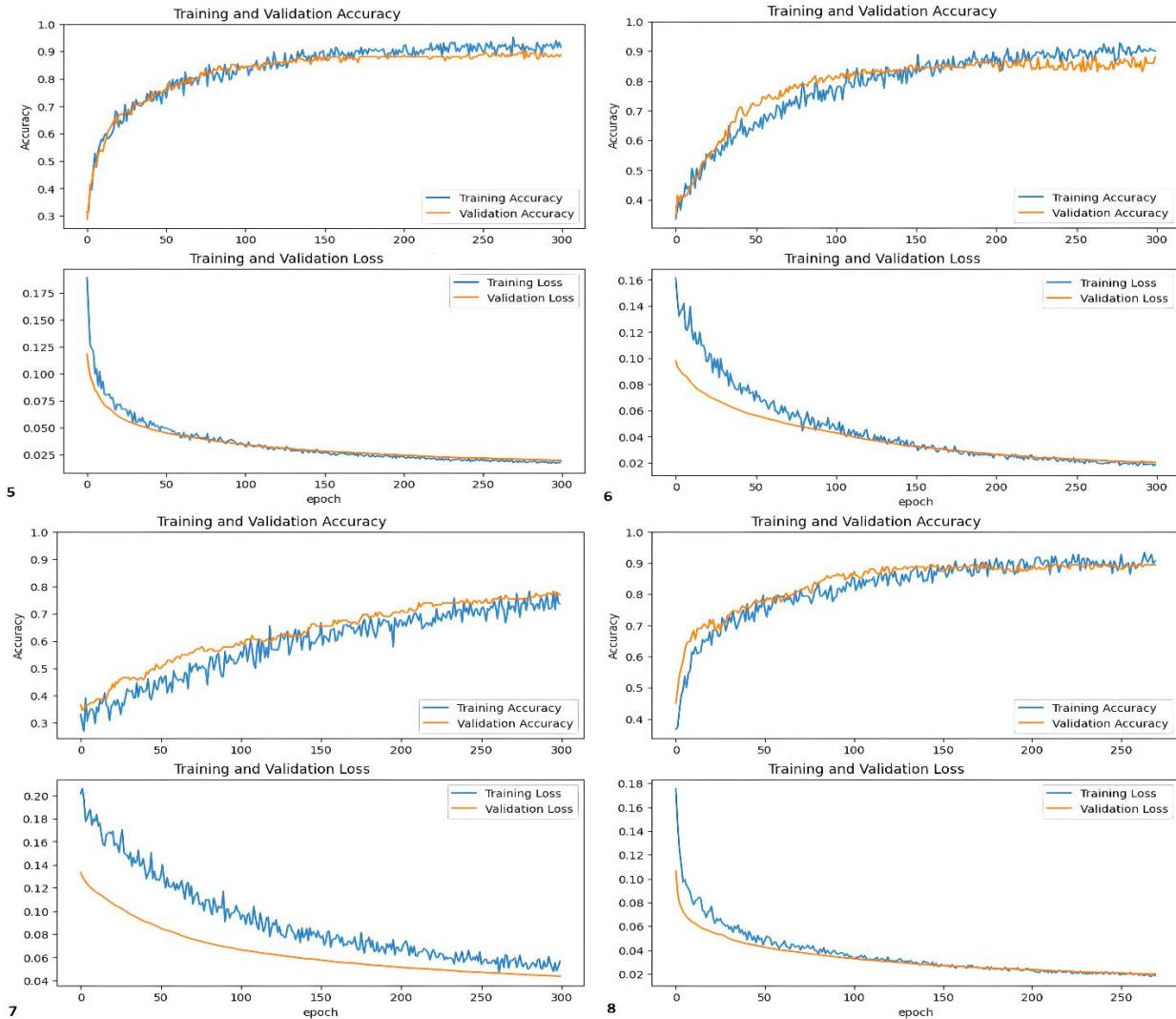


Fig. 8b: Performance of VGG19 with (5) Adam, (6) RMSprop, (7) Adamax, and (8) Nadam optimizers.

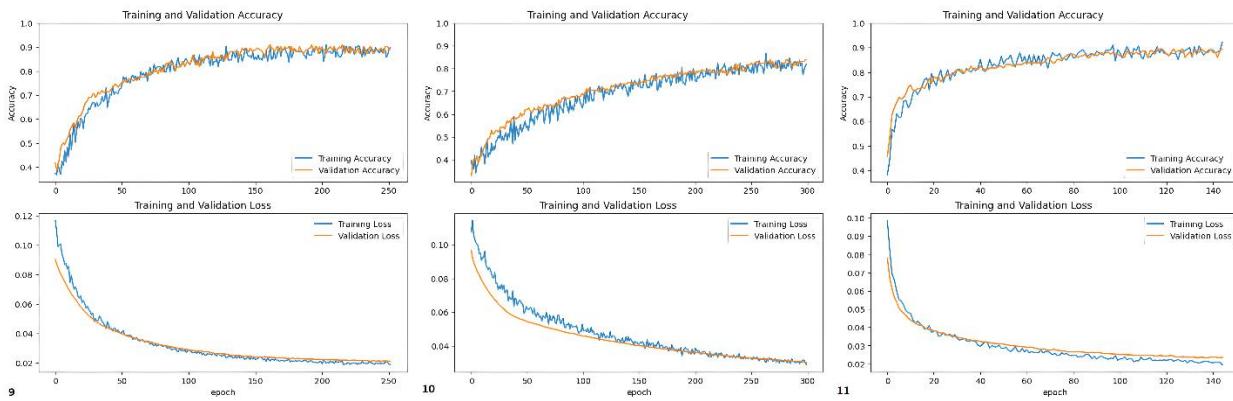


Fig. 8c: Performance curves for DenseNet-121 using (9) Adam, (10) RMSprop, and (11) Adamax optimizers

For second experiment, all selected models were set to non-trainable, and the number of epochs at which each model converged and showed no further improvement was determined by analyzing the accuracy graphs from first experiment. Additionally, an early stopping regularization strategy with a patience value of 5 was used to prevent unnecessary training.

After training these models, we changed the state of the last two convolutional layers from frozen to unfrozen and resumed training from the epoch where the models had previously completed its training. Table 3 summarizes the Precision, Recall, F1-Score, and overall average accuracy of the models before and after the fine-tuning process.

Table 3 Performance of the models before and after the fine-tuning process

Model + Optimizer Overall accuracy: a. Before Fine-tuning b. After Fine-tuning	Diseases Classes	Before Fine-tuning			After Fine-tuning		
		Precision	Recall	F1	Precision	Recall	F1
ResNet50 + Adam a. Accuracy: 0.87 b. Accuracy: 0.97	BP	0.78	0.92	0.84	0.96	0.96	0.96
	PF	0.97	1.0	0.98	1.0	1.0	1.0
	PV	0.91	0.75	0.83	0.96	0.96	0.96
	Average	0.89	0.89	0.88	0.97	0.97	0.97
ResNet50 + RMSprop a. Accuracy: 0.85 b. Accuracy: 0.89	BP	0.79	0.82	0.80	0.82	0.90	0.85
	PF	0.91	1.0	0.95	0.91	1.0	0.95
	PV	0.88	0.81	0.84	0.96	0.82	0.88
	Average	0.86	0.87	0.87	0.89	0.90	0.90
ResNet-50 + Adamax a. Accuracy: 0.89 b. Accuracy: 0.90	BP	0.83	0.88	0.85	0.86	0.86	0.86
	PF	0.94	1.0	0.97	0.94	1.0	0.97
	PV	0.92	0.84	0.88	0.91	0.87	0.89
	Average	0.90	0.90	0.90	0.90	0.91	0.90
ResNet-50 + Nadam a. Accuracy: 0.90 b. Accuracy: 0.97	BP	0.81	0.94	0.87	0.96	0.96	0.96
	PF	1.0	0.93	0.96	1.0	1.0	1.0
	PV	94	0.84	0.89	0.96	0.96	0.96
	Average	0.91	0.90	0.91	0.97	0.97	0.97
VGG19 + Adam a. Accuracy: 0.85 b. Accuracy: 0.98	BP	0.79	0.82	0.80	0.98	0.98	0.98
	PF	0.86	0.93	0.89	1.0	1.0	1.0
	PV	0.92	0.84	0.88	0.98	0.98	0.98
	Average	0.85	0.86	0.86	0.98	0.98	0.98
VGG19 + RMSprop a. Accuracy: 0.83 b. Accuracy: 0.94	BP	0.73	0.88	0.80	0.92	0.92	0.92
	PF	0.90	0.90	0.90	0.97	1.0	0.98
	PV	0.91	0.75	0.83	0.94	0.93	0.93
	Average	0.85	0.85	0.84	0.95	0.95	0.94
VGG19 + AdaMax a. Accuracy: 0.76 b. Accuracy: 0.88	BP	0.66	0.74	0.70	0.79	0.90	0.84
	PF	0.79	0.81	0.80	0.91	0.93	0.92
	PV	0.84	0.74	0.78	0.96	0.82	0.88
	Average	0.81	0.76	0.76	0.88	0.88	0.88

VGG + Nadam	BP	0.74	0.84	0.78	0.92	0.94	0.93
a. Accuracy: 0.83	PF	0.93	0.93	0.93	1.0	1.0	1.0
b. Accuracy: 0.95	PV	0.88	0.77	0.82	0.94	0.93	0.93
	Average	0.85	0.85	0.85	0.95	0.95	0.95
DenseNet-121 + Adam	BP	0.65	0.80	0.71	0.62	0.82	0.71
a. Accuracy: 0.77	PF	0.93	0.87	0.90	0.93	0.84	0.88
b. Accuracy: 0.76	PV	0.83	0.68	0.75	0.84	0.65	0.73
	Average	0.80	0.79	0.79	0.80	0.77	0.77
DenseNet-121+ RMSprop	BP	0.75	0.82	0.78	0.82	0.80	0.81
a. Accuracy: 0.8380	PF	0.88	0.96	0.92	0.89	1.0	0.94
b. Accuracy: 0.8662	PV	0.90	0.77	0.83	0.89	0.84	0.86
	Average	0.84	0.85	0.84	0.86	0.88	0.87
DenseNet + Adamax	BP	0.74	0.86	0.80	0.76	0.74	0.75
a. Accuracy: 0.8451	PF	0.90	0.90	0.90	0.86	0.93	0.89
b. Accuracy: 0.8239	PV	0.92	0.79	0.85	0.85	0.82	0.94
	Average	0.85	0.85	0.85	0.82	0.83	0.83
DenseNet + Nadam	BP	0.82	0.84	0.83	0.85	0.90	0.87
a. Accuracy: 0.88	PF	0.91	1.0	0.95	0.94	1.0	0.97
b. Accuracy: 0.90	PV	0.90	0.84	0.87	0.94	0.86	0.90
	Average	0.88	0.89	0.88	0.91	0.92	0.91

The results demonstrated that the performance of the models was improved through the fine-tuning process. Particularly, the ResNet50 with Adam and Nadam optimizers showed significant improvements in accuracy from 87% to 97% and 90% to 97%, respectively. Similarly, VGG19 achieved accuracy improvements of 85% to 98% (with Adam), 83% to 94% (with RMSprop), and 83% to 95% (with Nadam). The performance of the remaining models also showed improvement, but these improvements were not as significant as the ones observed in the ResNet50 and VGG19 models.

Figures 9a and 9b display the loss and accuracy curves before and after fine-tuning for the ResNet50 model using Adam and Nadam optimizers. In contrast, Figures 10a, 10b, and 10c show the corresponding plots for the VGG19 model with Adam, RMSProp, and Nadam optimizers. Figures 11 and 12 depict ROC curves generated by the ResNet50 and VGG19, respectively, using different optimizers.

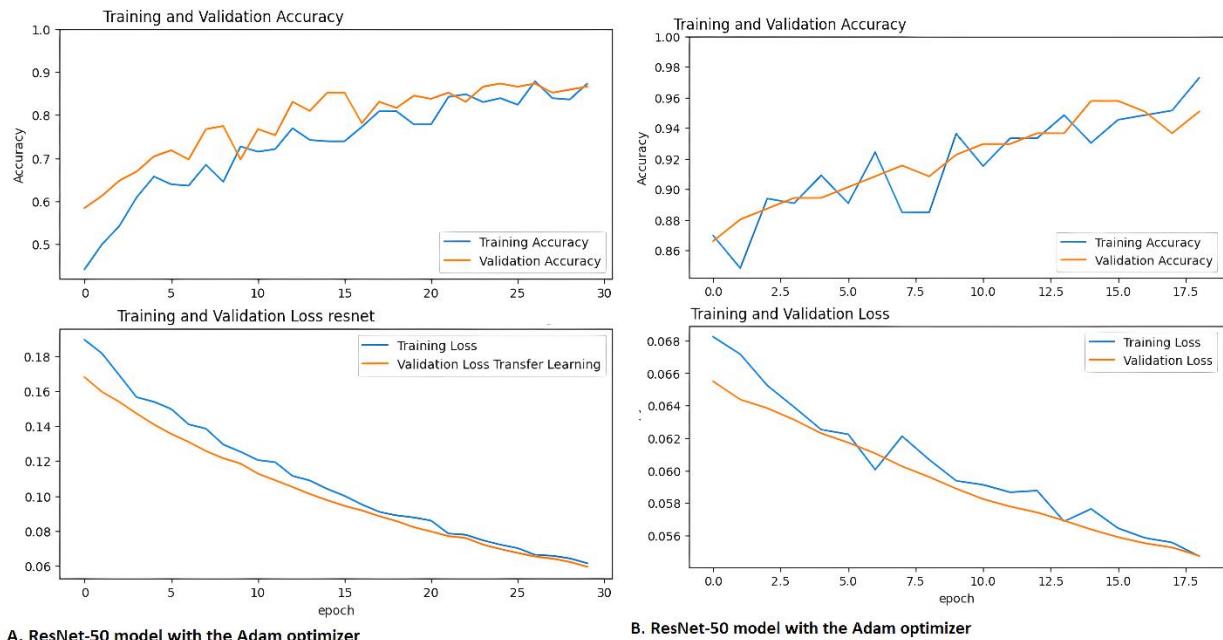


Fig. 9a Loss and accuracy curves for the ResNet50 model using the Adam optimizer: (A) before fine-tuning and (B) after fine-tuning.

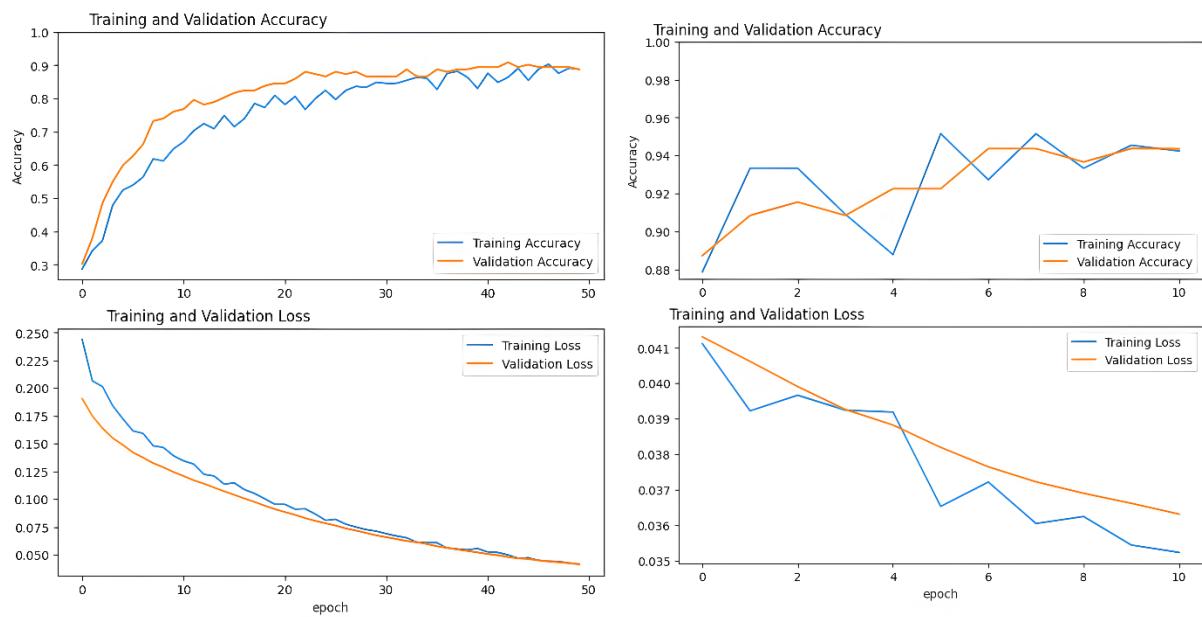
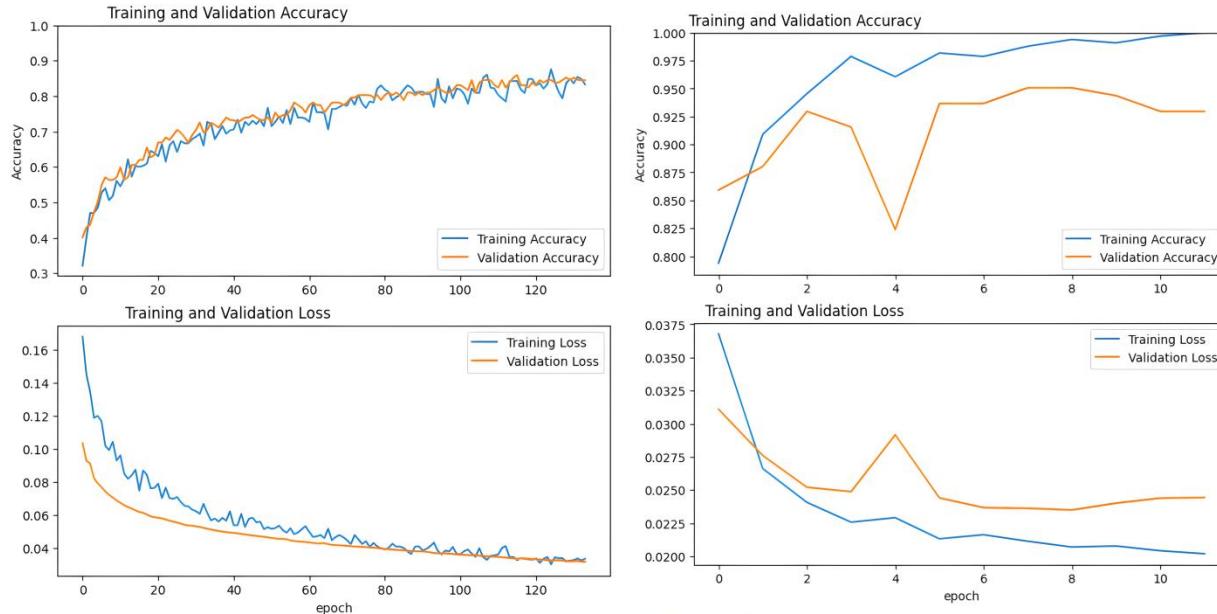


Fig. 9b Loss and accuracy curves for the ResNet50 model using the Nadam optimizer: (A) before fine-tuning and (B) after fine-tuning



25 Fig.10a Loss and accuracy curves (A) before and (B) after fine-tuning using Adam
26



55 Fig.10b Loss and accuracy curves (A) before and (B) after fine-tuning using RMSprop
56

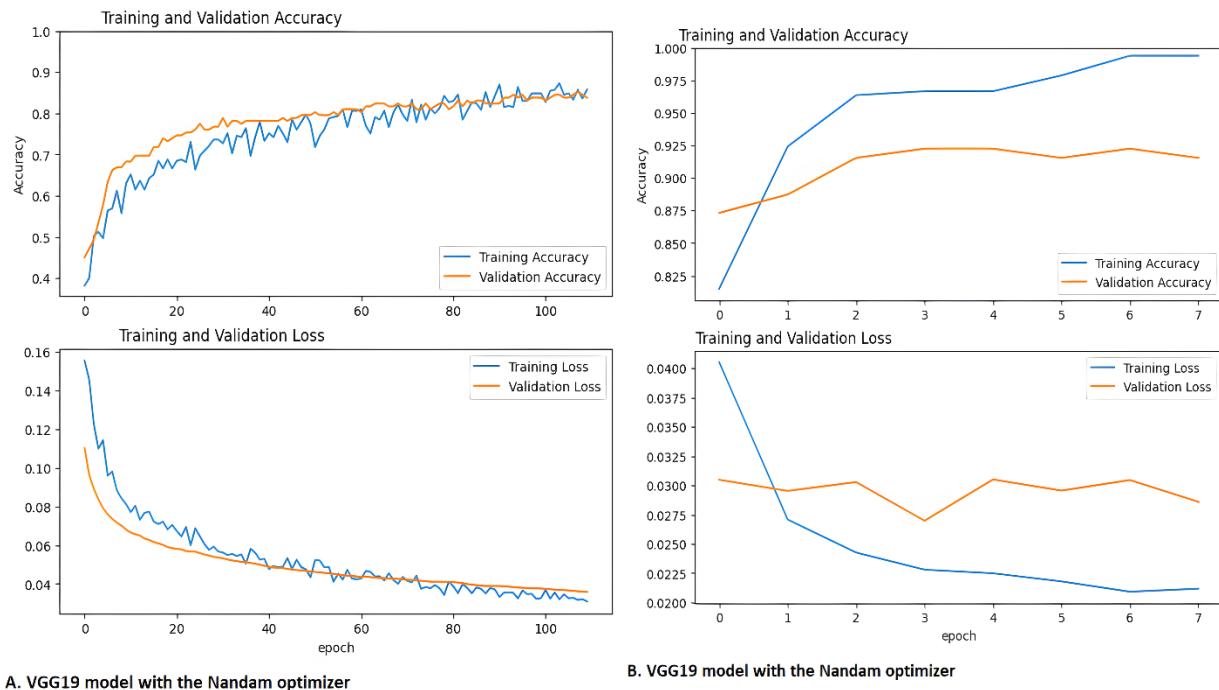


Fig. 10c Loss and accuracy curves (A) before and (B) after fine-tuning Nadam

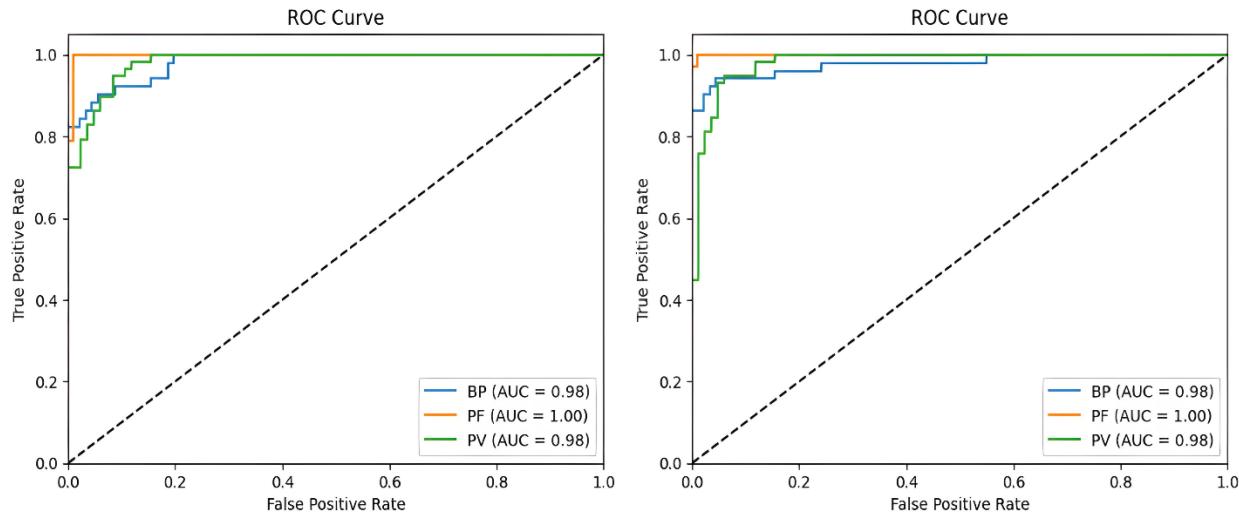


Fig. 11 ROC curves of ResNet50 with Adam and Nadam

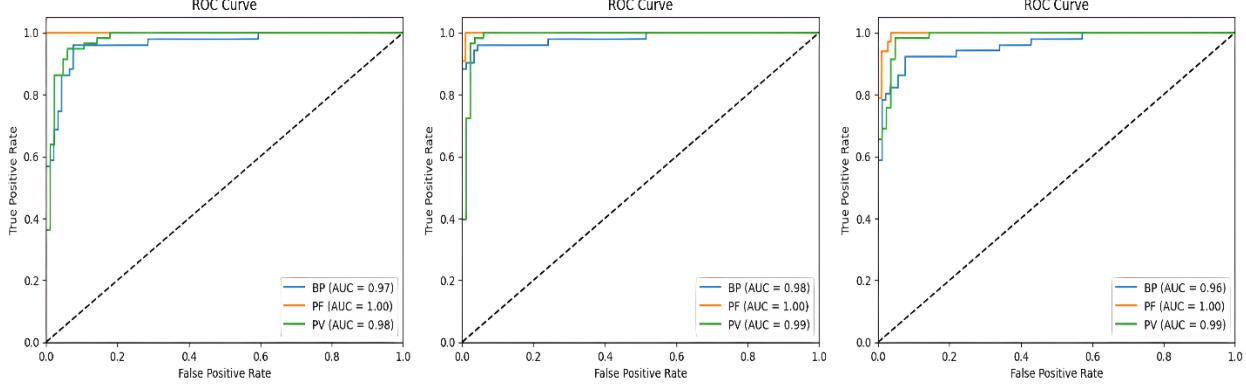


Fig. 12 ROC curves of the VGG19 with Adam, RMSprop, and Nadam

This study evaluates three approaches for classifying autoimmune blistering skin diseases (AIBD): classical machine learning, deep learning, and hybrid models. The observed performance differences primarily arise from variations in feature extraction techniques and learning capabilities.

Classical machine learning models, such as random forests and support vector machines (SVMs), rely heavily on manual feature extraction, which presents significant challenges in the medical domain due to the complex and nuanced patterns inherent in this field. The effectiveness of these models is often constrained by their dependence on handcrafted features, as manually capturing the intricate details required for accurate classification is highly challenging. Traditional techniques like Canny, Scharr, and Gabor filters have been employed for feature extraction; however, these methods may fall short in representing the full complexity of medical data, potentially omitting critical features essential for diagnosis. Furthermore, classical machine learning models require substantial domain expertise and tailored algorithms to identify optimal features, which limits their scalability and adaptability to diverse medical datasets.

In contrast, deep learning models, particularly convolutional neural networks (CNNs), excel in automated feature extraction by learning hierarchical patterns directly from raw image data. CNNs can identify subtle variations and intricate structures that are crucial for medical diagnosis. Fine-tuning pre-trained CNNs has been shown to significantly enhance classification performance, while techniques such as data augmentation and focal loss help mitigate overfitting and improve generalization.

The hybrid approach, which combines CNN-based feature extraction with classical classifiers, aims to balance interpretability and performance. While this method leverages the strengths of both paradigms, it often underperforms compared to standalone deep learning models due to the inherent limitations of classical classifiers in handling high-dimensional, automatically extracted features.

Our findings highlight the effectiveness of pre-trained deep learning models, particularly with transfer learning and fine-tuning, in outperforming classical and hybrid approaches for AIBD classification. These models provide extensive opportunities for further exploration, including optimizing hyperparameters, experimenting with randomly initialized network architectures, and refining fine-tuning strategies to enhance performance.

5. Conclusion:

Artificial intelligence techniques can assist in diagnosing the early stages of autoimmune blistering skin diseases. We collected a dataset comprising 100 images. The objective of this study was to provide the foundation for creating an AI-based model capable of being applied in real clinical scenarios. Through this work, we identified the most appropriate approach and best practices for the classification of AIBD.

We utilised classical machine learning, deep learning, and hybrid approaches in our methodology. Classical machine learning required domain expertise for manual feature engineering using appropriate algorithms; however, it required much smaller amounts of data as compared to deep learning techniques. In the classical approach, we used multiple filters to extract edge and texture features, which were then used for training support vector machine and random forest classifier. The results showed that random forest classifier achieved the best performance, with an accuracy of 57%.

In the deep learning approach, we used pre-trained CNN models. Transfer learning approaches are proven to be useful when dealing with limited dataset sizes. The pre-trained models were tested using different optimizers to assess their performance. After assessment using different hyperparameters, the top-performing models were identified for further fine-tuning on the target dataset. The experimental results demonstrated a significant improvement, with accuracy ranging from 84% to 98% through fine-tuning.

We also explored the hybrid approach, wherein we utilized the pre-trained models for feature extraction and random forest classifier and support vector machine for the classification task. In hybrid methodology, Xception in conjunction with support vector machine showed the highest accuracy of 81%.

The study demonstrates that pre-trained deep learning models, especially with transfer learning and fine-tuning, surpass classical and hybrid approaches for AIBD classification, with scope for further enhancement through hyperparameter optimization and fine-tuning strategies.

In the future, we will implement the best-identified configurations and explore advanced deep learning techniques, such as transformer-based models and segmentation pipelines alongside classification, on a large dataset. Currently, we are collecting images from real clinical settings to build a comprehensive and authenticated dataset with varying levels of taxonomy, ensuring coverage of diverse features relevant to the Indian population. Ethics approval for data collection

1
2
3
4 has been granted. These advancements will contribute to more reliable and scalable solutions for
5 diagnosing autoimmune blistering skin diseases.
6
7
8
9

10 **Data availability** The data (clinical images pertain to the Department of Dermatology,
11 Venereology & Leprology, PGIMER, Chandigarh, India) that support the findings of this study
12 are not openly available due to reasons of privacy for research participants and are accessible
13 from the authors upon reasonable request and with permission from PGIMER, Chandigarh, India.
14 The data are located in controlled access data storage at PGIMER, Chandigarh. However, the
15 data, clinical images, which we collected from online resources, are openly available on DermNet
16 at <https://www.dermnetnz.org>.
17
18

19 **Declarations**

20 **Conflict of Interest:** The authors declare that they have no conflict of interest.
21
22

23 **References**

- 24
25
26
27
28
29
30 1. O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, Riordan
31 D, Walsh J (2020) Deep learning vs. traditional computer vision. In: Advances in Computer
32 Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1 2020 (pp. 128-
33 144). Springer International Publishing.
34
35 2. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel
36 MA, Al-Amidie M, Farhan L (2021) Review of deep learning: Concepts, CNN architectures,
37 challenges, applications, future directions. Journal of Big Data 8:1-74.
38
39 3. Ferreira MD, Corrêa DC, Nonato LG, de Mello RF (2018) Designing architectures of
40 convolutional neural networks to solve practical problems. Expert Systems with Applications
41 94:205-217.
42
43 4. Kieffer B, Babaie M, Kalra S, Tizhoosh HR (2017) Convolutional neural networks for
44 histopathology image classification: Training vs. using pre-trained networks. In: 2017 seventh
45 international conference on image processing theory, tools and applications (IPTA) (pp. 1-6).
46 IEEE.
47
48 5. Kaur T, Gandhi TK (2019) Automated brain image classification based on VGG-16 and transfer
49 learning. In: 2019 International Conference on Information Technology (ICIT) (pp. 94-98). IEEE.
50
51 6. Shaha M, Pawar M (2018) Transfer learning for image classification. In: 2018 second
52 international conference on electronics, communication and aerospace technology (ICECA) (pp.
53 656-660). IEEE.
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 7. Prajapati SA, Nagaraj R, Mitra S (2017) Classification of dental diseases using CNN and transfer
5 learning. In: 2017 5th International Symposium on Computational and Business Intelligence
6 (ISCBI) (pp. 70-74). IEEE.
7
8 8. Mahbod A, Schaefer G, Wang C, Ecker R, Ellinge I (2019) Skin lesion classification using
9 hybrid deep neural networks. In: ICASSP 2019-2019 IEEE International Conference on Acoustics,
10 Speech and Signal Processing (ICASSP) (pp. 1229-1233). IEEE.
11
12 9. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical
13 image database. In: 2009 IEEE conference on computer vision and pattern recognition (pp. 248-
14 255). IEEE.
15
16 10. Hussain M, Bird JJ, Faria DR (2019) A study on cnn transfer learning for image classification.
17 In: Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK
18 Workshop on Computational Intelligence, September 5-7, 2018, Nottingham, UK (pp. 191-202).
19 Springer International Publishing.
20
21 11. Talo M, Baloglu UB, Yıldırım Ö, Acharya UR (2019) Application of deep transfer learning
22 for automated brain abnormality classification using MR images. Cognitive Systems Research
23 54:176-188.
24
25 12. Cheng PM, Malhi HS (2017) Transfer learning with convolutional neural networks for
26 classification of abdominal ultrasound images. Journal of digital imaging 30:234-243.
27
28 13. Dermnet nz. [Online]. Available: <https://www.dermnetnz.org>
29
30 14. Sobel I, Feldman G (1968) A 3x3 isotropic gradient operator for image processing. a talk at
31 the Stanford Artificial Project in:271-272.
32
33 15. ALEnezi NSA (2019) A method of skin disease detection using image processing and machine
34 learning. Procedia Computer Science 163:85-92.
35
36 16. Mohammed KK, Afify HM, Hassanien AE (2020) ARTIFICIAL INTELLIGENT SYSTEM
37 FOR SKIN DISEASES CLASSIFICATION. Biomedical Engineering: Applications, Basis and
38 Communications 32(05):2050036.
39
40 17. Dhivyaa C, Sangeetha K, Balamurugan M, Amaran S, Vetriselvi T, Johnpaul P (2020) Skin
41 lesion classification using decision trees and random forest algorithms. Journal of Ambient
42 Intelligence and Humanized Computing:1-13.
43
44 18. Senan EM, Jadhav ME, Kadam A (2021) Classification of PH2 images for early detection of
45 skin diseases. In: 2021 6th International Conference for Convergence in Technology (I2CT) (pp.
46 1-7). IEEE.
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 19. Kassem MA, Hosny KM, Damaševičius R, Eltoukhy MM (2021) Machine learning and deep
5 learning methods for skin lesion classification and diagnosis: a systematic review. *Diagnostics*
6 11(8):1390.
7
8 9 20. Han D, Liu Q, Fan W (2018) A new image classification method using CNN transfer learning
10 and web data augmentation. *Expert Systems with Applications* 95:43-56.
11
12 13 21. Gavrilov D, Melerzanov AV, Shchelkunov N, Zakirov E (2019) Use of neural network-based
14 deep learning techniques for the diagnostics of skin diseases. *Biomedical Engineering* 52:348-352.
15
16 17 22. Pangti R, Mathur J, Chouhan V, et al. (2021) A machine learning-based, decision support,
18 mobile phone application for diagnosis of common dermatological diseases. *Journal of the*
19 *European Academy of Dermatology and Venereology* 35(2):536-545.
20 21 23 24 25 26 27 28 29 24. Guha SR, Rafizul Haque SM (2020) Performance comparison of machine learning-based
29 classification of skin diseases from skin lesion images. In: *International Conference on*
30 *Communication, Computing and Electronics Systems: Proceedings of ICCCES 2019* (pp. 15-25).
31 Springer Singapore.
32
33 25. Janoria H, Minj J, Patre P (2020) Classification of skin disease from skin images using transfer
34 learning technique. In: *2020 4th International Conference on Electronics, Communication and*
35 *Aerospace Technology (ICECA)* (pp. 888-895). IEEE.
36
37 26. Nair PS, Berihu TA, Kumar V (2020) An image-based gangrene disease classification.
38 *International Journal of Electrical & Computer Engineering* 10(6).
39
40 27. Purwar S, Tripathi R, Ranjan R, Saxena R (2021) Classification of thalassemia patients using
41 a fusion of deep image and clinical features. In: *2021 11th International Conference on Cloud*
42 *Computing, Data Science & Engineering (Confluence)* (pp. 410-415). IEEE.
43
44 28. Palivel L H, Athanesious J J, Deepika V, Vignesh M (2021) Segmentation and Classification
45 of Skin Lesions from Dermoscopic Images.
46
47 29. Islam MM, Islam MZ, Asraf A, Al-Rakhami MS, Ding W, Sodhro AH (2022) Diagnosis of
48 COVID-19 from X-rays using combined CNN-RNN architecture with transfer learning.
49 *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 2(4):100088.
50
51 30. Jimenez-del-Toro O, Otálora S, Andersson M, Eurén K, Hedlund M, Rousson M, Müller H,
52 Atzori M (2017) Analysis of histopathology images: From traditional machine learning to deep
53 learning. In: *Biomedical Texture Analysis* (pp. 281-314). Academic Press.
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 31. Singh M, Singh M, De D, Handa S, Mahajan R, Chatterjee D (2023) Towards Diagnosis of
5 Autoimmune Blistering Skin Diseases Using Deep Neural Network. Archives of Computational
6 Methods in Engineering 30(6):3529-57.
7
8 32. Manuelyan K, Dragolov M, Drenovska K, Shahid M, Vassileva S (2024) Artificial Intelligence
9 in Autoimmune Bullous Dermatoses. Clinics in Dermatology
10
11 33. Schielein MC, Christl J, Sitaru S, Pilz AC, Kaczmarczyk R, Biedermann T, Lasser T, Zink A
12 (2023) Outlier detection in dermatology: Performance of different convolutional neural networks
13 for binary classification of inflammatory skin diseases. Journal of the European Academy of
14 Dermatology and Venereology 37(5):1071-9.
15
16
17 34. Ahmed M, Islam SB, Alif AU, Islam M, Saima SM (2023) AI empowered diagnosis of
18 pemphigus: a machine learning approach for automated skin lesion detection. Informatyka,
19 Automatyka, Pomiary w Gospodarce i Ochronie Środowiska.
20
21 35. Dubey S, Cyril CP. Detection of Pemphigus using Machine Learning (2023) International
22 Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and
23 Computational Intelligence (RAEEUCCI).IEEE (pp. 1-6).
24
25 36. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA (2020)
26 Albumentations: fast and flexible image augmentations. Information 11(2):125.
27
28 37. Pizer SM, Amburn EP, Austin JD, et al. (1987) Adaptive histogram equalization and its
29 variations. Computer vision, graphics, and image processing 39(3):355-368.
30
31 38. Li H-A, Fan J, Zhang J, et al. (2021) Facial image segmentation based on Gabor filter.
32 Mathematical Problems in Engineering 2021:1-7.
33
34 39. Zhao M, Qiu W, Wen T, Liao T, Huang J (2021) Feature extraction based on Gabor filter and
35 Support Vector Machine classifier in defect analysis of Thermoelectric Cooler Component.
36 Computers & Electrical Engineering 92:107188.
37
38 40. Canny J (1986) A computational approach to edge detection. IEEE Transactions on pattern
39 analysis and machine intelligence(6):679-698.
40
41 41. Maini R, Aggarwal H (2009) Study and comparison of various image edge detection
42 techniques. International journal of image processing (IJIP) 3(1):1-11.
43
44 42. Moldovanu S, Moraru L, Biswas A (2016) Edge-based structural similarity analysis in brain
45 MR images. Journal of Medical Imaging and Health Informatics 6(2):539-546.
46
47 43. Sian C, Jiye W, Ru Z, Lizhi Z (2020) Cattle identification using muzzle print images based on
48 feature fusion. In: IOP Conference Series: Materials Science and Engineering, vol 853 (p. 012051).
49 IOP Publishing.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6 44. Basu M (2002) Gaussian-based edge-detection methods-a survey. IEEE Transactions on
7 Systems, Man, and Cybernetics, Part C (Applications and Reviews) 32(3):252-260.
8
9 45. Koschan A, Abidi M (2001) A comparison of median filter techniques for noise removal in
10 color images. In: Proc 7th German workshop on color image processing (pp. 69-79). Institute of
11 Computer Science Erlangen, Germany.
12
13 46. Azar AT, Elshazly HI, Hassanien AE, Elkorany AM (2014) A random forest classifier for
14 lymph diseases. Computer methods and programs in biomedicine 113(2):465-473.
15
16 47. Caie PD, Dimitriou N, Arandjelović O (2021) Precision medicine in digital pathology via
17 image analysis and machine learning Artificial intelligence and deep learning in pathology.
18 Elsevier, p 149-173.
19
20 48. Yu H, Kim S (2012) SVM Tutorial-Classification, Regression and Ranking. Handbook of
21 Natural computing 1:479-506.
22
23 49. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016)
24 Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset
25 characteristics and transfer learning. IEEE transactions on medical imaging 35(5):1285-1298.
26
27 50. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016)
28 Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE
29 transactions on medical imaging 35(5):1299-1312.
30
31 51. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image
32 recognition. arXiv preprint arXiv:1409.1556.
33
34 52. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In:
35 Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
36
37 53. Szegedy C, Ioffe S, Vanhoucke V, Alemi A (2017) Inception-v4, inception-resnet and the
38 impact of residual connections on learning. In: Proceedings of the AAAI conference on artificial
39 intelligence, vol 31, no. 1.
40
41 54. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In:
42 Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).
43
44 55. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional
45 networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (pp.
46 4700-4708).
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



Click here to access/download
Supplementary Material
Supplementary Material R2.docx