Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Preprocessed Vision Transformers and Classical Classifiers in Diagnosing Skin Diseases

Bilal ŞENOL[1*], Uğur DEMİROĞLU[2]

[1] Aksaray University, Software Engineering Department, bilal.senol@aksaray.edu.tr, Orcid No: 0000-0002-3734-8807
[2] Kahramanmaraş İstiklal University, Software Engineering Department, ugurdemiroglu@istiklal.edu.tr, Orcid No: 0000-0002-0000-8411

ARTICLE INFO

Article history:

Received 1 November 2024
Received in revised form 21 January 2025
Accepted 4 March 2025
Available online 26 March 2025

Keywords:

Skin Diseases, Preprocessing, Adaptive Histogram Equalization, Classification, Vision Transformers

Doi: 10.24012/dumf.1577835

* Corresponding author

ABSTRACT

Vision Transformers (ViTs) represent the cutting edge of deep learning technology in medicine. Due to their large number of parameters, ViTs require extensive datasets for effective learning, which has become feasible with the digitization of healthcare. In contrast, classical classifiers typically operate with smaller datasets and fewer input features. High-resolution images, such as those obtained from dermoscopy, confocal microscopy, reflectance confocal microscopy, and Raman spectroscopy, are commonly used for diagnosing skin diseases in clinical practice. ViTs show significant promise in this area due to their unique design, which eliminates the need for convolutional operations used in convolutional neural networks (CNNs). This study introduces an artificial intelligence method for classifying skin diseases into five categories: normal, melanoma, arsenic-related conditions, psoriasis, and eczema. Images from the dataset were firstly preprocessed using the Adaptive Histogram Equalization (AHE) technique to enhance contrast and reveal critical details. ViTs were then employed to extract complex visual features from these images. These features were subsequently combined with traditional machine learning classifiers, resulting in highly accurate skin disease diagnoses. Additionally, comparative experiments were performed on dermatoscopic images from another dataset, reinforcing the versatility and effectiveness of this approach. The findings highlight the potential of integrating ViTs with classical classifiers to improve the accuracy and reliability of medical image classification tasks.

Introduction

Skin diseases represent a significant public health issue due to their high prevalence and the detrimental effects they have on the quality of life of affected individuals. The skin and its appendages play a role in various disease processes, often displaying signs that are observable from afar. Dermatological disorders contribute to physical morbidity and are associated with anxiety, depression, and psychological disturbances. Hyperactive sebaceous glands contribute to conditions like acne and seborrheic dermatitis, characterized by the overproduction of sebum. In conditions such as ichthyosis, reduced sebum production frequently results in dry, flaky scales, contributing to a rough skin appearance. 'Skin diseases' is a broad term encompassing various abnormal conditions impacting the dermal and epidermal layers of the skin. The skin can indicate underlying diseases resulting from alterations in biochemical compounds. Skin diseases may be classified as primary, where the skin is directly affected, or secondary, where the skin exhibits significant lesions as a result of changes in another system or organ. These may arise from the irregular functioning of sweat and oil glands or from exposure to environmental pollutants, resulting in the development of anticipated skin diseases or conditions. This manuscript presents a thorough analysis of the causes, symptoms, and treatments of common skin diseases. Skin diseases represent a prevalent category of disorders, prompting millions of patients to seek medical attention. Psychological and social factors, including feelings of embarrassment and the concealment of a disease, primarily motivate individuals to consult a practitioner for skin disorders. Untreated skin diseases pose various significant health risks that may impede early diagnosis and appropriate treatment [1-4].

Skin diseases are becoming increasingly prevalent across all races, ages, and specific occupations in contemporary society. Skin diseases appear to be increasingly prevalent. Skin diseases impact a wide range of individuals globally, manifesting in different tissue and organ systems, including hair and nails. Inflammatory skin rashes can be categorized based on their etiology, which may involve microorganisms or not. They may present with or without serum or plasma protein exudation and can be classified as primary or secondary effects of the lesions, exhibiting either acute or chronic symptoms. Acne, eczema, psoriasis, rosacea, and skin cancer represent prevalent dermatological conditions in contemporary society. The evolving understanding of skin diseases, which is a significant concern, has led to increased efforts in their detection and treatment.

Significant changes in classification are occurring, informed by recent advancements in basic research and clinical support. While these skin diseases do not pose significant harm to humans, they can, in certain instances, disrupt the body's various internal regulatory systems. The significance of quality of life and psychodermatology research is crucial for comprehending such behaviors. Multiple factors contribute to the development of skin diseases, particularly the incidence of infections. The majority is attributed to contagious factors. Nonetheless, certain non-communicable factors may primarily arise from genetic influences, pharmaceuticals, dietary choices, and similar elements. These factors are crucial for preventing the initiation and cessation of a specific treatment. Prior to advancing, it is essential to comprehend the characteristics of the disease itself [5-9].

It is now widely recognized that dermatological conditions impact individuals' quality of life. The precise diagnosis of dermatological conditions has consistently posed a difficulty for healthcare professionals. Historically, the sole method for diagnosing skin problems was by visual inspection. Dermatologists began to acquire the expertise necessary to diagnose skin disorders using factors beyond morphological features, including systemic alterations and medical history. These procedures in clinical practice are hindered by uncertainty, protracted timelines, and the necessity for collaboration with the clinical laboratory. The advancement of technology has enabled the utilization of non-invasive patient assessment instruments across multiple medical fields. Due to advancements in artificial intelligence (AI) and computational capabilities, dermatology practitioners have developed tools to assist in the investigation of skin problems [10-13].

Combining AI and machine learning (ML) methods carries the potential for analyses, decision rules, and disease predictions based on big dermatological data. The ML methods enable the classification of skin lesions as accurately as dermatologists. Consequently, the current improvements have moved the long-used rule-based expert systems towards applied computer health diagnostics and a meaningful patient care supporting methodology [14]. It is quite well known that in medicine the error rate in diagnosis changes with diseases, diagnostic aims, and physicians. The further role of AI in medicine is inevitable and will continuously grow, portending an increasing role in dermatology. The everyday diagnostic methodology of skin diseases is performed by either primary care physicians, dermatologic residents, or field-specialized dermatologists. The decreasing number of dermatologists ignites difficulties in people's accessibility to dermatological diagnostic facilities. An additional problem within diagnosis is the inter- and intraobserver variability. Several solutions have been proposed over the years in order to decrease this variability, and the improvement of the methods has started to grow due to the availability of big datasets [15, 16]. The renewed level of interest in the importance of AI in the field of dermatological diagnostics has been observed over the last five years. All the above-mentioned arguments formed the basis for this review, proving the state of the art regarding computer-assisted dermatological diagnostics, and calling for an in-depth review of recent applications in this field. For the above-mentioned reasons, the objectives of this review are to perform an in-depth analysis and comparison of the potential area of deep learning, especially convolutional neural networks, and ML algorithms. Moreover, the final aim of this review is to propose a future perspective for AI enhancement in dermatology [17, 18]. The literature reveals a progressive evolution in the application of ML and deep learning (DL) methodologies to enhance diagnostic accuracy and accessibility in dermatological practice. Kawahara and Hamarneh provide an essential foundation by highlighting the critical role of accurate skin disease diagnosis in determining effective treatment strategies [19]. Building on this groundwork, [20] assess the robustness of deep learning methods within clinical workflows. Chan et al. expand the discussion by providing a comprehensive overview of current ML applications in dermatology [21]. An innovative interactive deep learning system aimed at differential diagnosis of malignant skin lesions is introduced in [22]. Chowdary et al. offer a comprehensive survey of various ML and DL techniques for diagnosing dermatological diseases in [23] and [24] further emphasizes the role of automated analysis in dermatology, detailing the use of convolutional neural networks (CNNs) for skin cancer classification. The study in [25] discuss the transformative potential of AI in dermatology, particularly in addressing the global shortage of dermatologists. Mode specifically, [26] provides a comprehensive overview of the application of ViTs in the automated segmentation and classification of skin lesions, particularly in the context of dermoscopy images. The authors underscore the significance of precise lesion segmentation as a critical precursor to accurate skin cancer diagnosis, highlighting the transformative impact that advanced deep learning and machine learning models have had on this field. The study in [27] highlights the advantages of ViTs over traditional CNNs, particularly their ability to capture long-range dependencies within data, which is crucial for accurately interpreting complex medical images. This capability is particularly relevant in the context of skin disease detection, where the identification of subtle patterns can significantly impact diagnosis and treatment. the comprehensive survey in [28] emphasizes the promising results achieved by ViTs in medical image segmentation, particularly in histopathology. The authors discuss various innovative architectures, such as DHUnet and MaxViT-UNet, which leverage ViTs for improved feature aggregation and segmentation accuracy. The study in [29] investigates the incorporation of ViTs for enhanced skin lesion diagnosis. It emphasizes the benefits of ViTs, especially their capacity to capture long-range relationships and intricate patterns in dermoscopic pictures. The article highlights the versatility of ViTs in managing intricate datasets and illustrates their capability to enhance diagnostic precision while minimizing interobserver variability. The results support the implementation of ViTs in practical dermatological processes to improve diagnosis accuracy. In addition, [30] presents Assist-Dermo, a streamlined ViTs model intended

for multiclass skin lesion categorization. The model substantially decreases computing complexity by the utilization of separable attention techniques, maintaining accuracy. Assist-Dermo specializes in the classification of various skin lesions, such as melanoma, basal cell carcinoma, and benign diseases. The study highlights the model's efficacy in resource-limited conditions, rendering it appropriate for implementation in distant or under-resourced clinical contexts. Experimental findings illustrate the model's resilience and its capacity to democratize access to AI-driven diagnostic instruments for skin health. The ViTs has been used in pre-diagnosis of skin diseases in [31]. The study indicates potential in utilizing Visual Transformers for skin disease detection, enhancing accuracy, computing efficiency, and scalability, which may greatly aid clinical environments and practical medical diagnostics. Skin cancer classification study using medical ViTs can be found in [32]. The paper validates the effectiveness of medical ViTs in improving skin cancer classification, presenting a robust alternative to conventional methods and demonstrating great promise for future medical applications. There can be found more valuable studies in the literature dealing with the issue.

ViTs have several advantages compared to conventional CNNs, especially in the realm of medical picture processing. In contrast to CNNs, which utilize convolutional layers to identify local patterns, ViTs utilize self-attention processes that are proficient at collecting global dependencies over an entire image. This skill is essential for the analysis of intricate medical pictures, particularly in skin disease categorization, where nuanced patterns and extensive contextual linkages greatly affect diagnostic precision. Furthermore, ViTs remove the inductive biases seen in CNNs, including localization and translational invariance, enabling superior generalization across varied datasets. Their patch-based processing enables the management of high-resolution photos without significant preprocessing. The experimental findings of this work indicate that ViTs, when integrated with traditional classifiers, improve classification performance and display resilience to fluctuations in picture quality and data distribution. These characteristics establish ViTs as a revolutionary instrument in enhancing dermatological diagnostics, beyond the constraints of CNN-based methods, and facilitating the development of more accurate and dependable automated diagnostic systems.

This paper presents an AI approach for diagnosing various skin diseases. The dataset which includes the images categorize the images into 5 clusters which are normal, melanoma, arsenic, psoriasis and eczema. In the study, skin images were first preprocessed using the Adaptive Histogram Equalization (AHE) technique, enhancing image contrast to reveal critical details. Following this preprocessing, features were extracted from the images using ViTs, known for their ability to capture intricate visual information. These extracted features were then classified using traditional ML classifiers, enabling accurate diagnosis of the skin conditions under investigation. The results highlight the effectiveness of combining ViTs with classical classifiers in medical image classification tasks.

In further sections, the dataset of the images is briefly presented and the structure of ViTs is given. An illustrative example is considered to show the effectiveness of the method and the results are clearly discussed.

Background

As given in the previous section, the dataset has skin images categorized into 5 groups. The dataset used in this study is the *Skin Diseases Dataset* downloaded from Kaggle [33]. Sample images from the dataset can be seen in Figure 1.

Figure 1. Sample images of skin diseases from the dataset.

The dataset size is 3.02 GB and consists of images categorized into five subfolders: Normal, Melanoma, Arsenic, Psoriasis, and Eczema. It contains a total of 7,356 skin disease images, distributed as follows: 1,815 Normal images, 1,575 Melanoma images, 741 Arsenic images, 1,724 Psoriasis images, and 1,501 Eczema images. Each image has a resolution of 72 DPI, 24-bit depth, and dimensions predominantly around 3000 x 4000 pixels, saved in JPG format. This dataset is publicly licensed, freely accessible, and widely used in fields such as medicine,

cancer research, and computer vision, offering uninterrupted access and download.

The class descriptions are as follows:

- **Normal:** Images of healthy, unaffected skin.
- **Melanoma:** Images depicting melanoma, a serious type of skin cancer.
- **Arsenic:** Skin lesions resulting from exposure to arsenic, a known risk factor for various skin diseases.
- **Psoriasis:** Chronic autoimmune condition that causes rapid skin cell accumulation, leading to scaling and inflammation.
- **Eczema:** Also known as atopic dermatitis, characterized by itchy and inflamed skin.

The dataset can be applied in several areas:

- **Classification:** Training models to classify different skin conditions from images.
- **Computer Vision Tasks:** Used for segmentation, object detection, and other image analysis tasks.
- **Medical Research:** Developing AI models to assist dermatologists in diagnosing skin conditions.

After a brief description about the dataset, the ViTs method can be presented. The exploration of ViTs in image classification has garnered significant attention in recent years, reflecting a shift from traditional CNNs towards architectures that leverage self-attention mechanisms. The literature reveals a rich tapestry of insights into how ViTs operate, their advantages over CNNs, and the challenges that remain in optimizing their performance for various applications. General ViTs architecture is illustrated in Figure 2 [34].

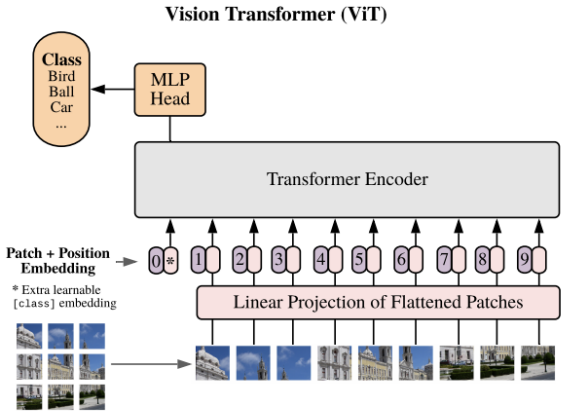

Figure 2. General Architecture of the ViTs.

In 2020, Han et al. provided a foundational understanding of ViTs, emphasizing the role of the self-attention layer in facilitating global interactions between image patches. They introduced several innovative approaches aimed at enhancing the self-attention mechanism, including DeepViT's cross-head communication and KVT's k-NN attention, which prioritize locality and reduce noise in attention calculations. Their work underscored the importance of network architecture, hinting at the potential for new designs that could further bridge the performance gap between ViTs and established CNNs [35].

After this foundation, Naseer et al. delved into the properties of learned representations in ViTs, particularly in safety-critical domains. Their comparative analysis with CNNs illuminated the robustness and generalization capabilities of ViTs, revealing that the self-attention mechanism allows for effective long-range interaction modeling. This study highlighted the adaptability of ViTs to various nuisances in data, showcasing their potential for real-world applications where reliability is paramount [36]. Cao et al. further explored the training dynamics of ViTs, demonstrating that these architectures could achieve state-of-the-art performance even with limited datasets. Their findings indicated that the representations learned from small datasets could enhance performance on larger datasets, suggesting a promising avenue for leveraging ViTs in scenarios where data is scarce [37]. In the same year, Parvaiz et al. examined the application of ViTs in medical computer vision, particularly for diagnosing diabetic retinopathy. They emphasized the advantages of ViTs over CNNs in terms of accuracy, driven by the attention mechanism's ability to assess global context. This work pointed to the transformative potential of ViTs in healthcare, while also acknowledging the challenges that need to be addressed for broader adoption [38]. Jelassi et al. contributed to the understanding of ViTs by demonstrating that these models can learn spatially localized patterns effectively. Their introduction of a positional attention mechanism provided insights into how ViTs can maintain spatial structure while generalizing across different datasets, reinforcing the notion that ViTs are not merely a replacement for CNNs but rather a complementary approach with unique advantages [39]. Nguyen et al. further expanded on the operational differences between ViTs and CNNs, investigating the robustness of ViTs against various perturbations. Their visualization techniques offered a deeper understanding of how ViTs process information, revealing clustering behaviors in feature embeddings that could inform future architectural improvements [40]. General block diagram of the method in this study is given in Figure 3.

ViTs has 3 models, which can be listed as follows:

- **base-16-imagenet-384** — A base-sized model with 86.8 million parameters and a patch size of 16. The network is fine-tuned using the ImageNet 2012 dataset at a resolution of 384x384.
- **small-16-imagenet-384** — A small-sized model with 22.1 million parameters and a patch size of 16. The network is fine-tuned using the ImageNet 2012 dataset at a resolution of 384x384.
- **tiny-16-imagenet-384** — A tiny-sized model with 5.7 million parameters and a patch size of 16. The network is fine-tuned using the ImageNet 2012 dataset at a resolution of 384x384.

In this study, the **base-16-imagenet-384** model is implemented.

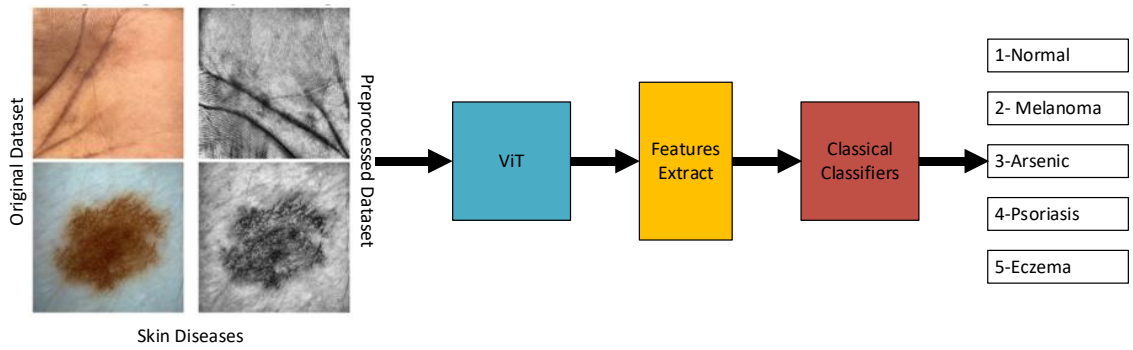

Figure 3. General block diagram of the method in this paper.

Classification Study

This study allocated 80% of the dataset for training and retained the remaining 20% exclusively for testing, guaranteeing that no test data were utilized during training. Initially, diagnostic outcomes were produced via the ViTs network applied directly to the original dataset. Subsequently, other preprocessing measures were implemented to improve the model's performance. The preprocessing approach commenced with the transformation of the dataset's color structure, translating the RGB images into the Lab color space, a model that differentiates luminance from chromaticity. Only the initial channel, which denotes brightness, was taken from this change. The channel was further treated with Contrast-Limited Adaptive Histogram Equalization (CLAHE), an adaptive histogram technique aimed at improving local contrast more efficiently. Enhancing the visual quality of digital images by increasing their contrast has been an attractive research topic in the field of image processing. The histogram equalization method is employed to improve the contrast of the image by altering the histogram of the accumulated pixel intensity values. However, the traditional approach of histogram equalization amplifies the noise and diminishes the quality of the image through loss of useful information such as detailed features. To resolve these problems and to preserve the pertinent information of the image, CLAHE was proposed. The basic principle of CLAHE lies in employing adaptive histogram equalization locally to minimize the contrast limit effect in an efficient manner [41]. The improvement in image processing techniques, along with the wide and growing use of medical imaging applications, has necessitated the introduction of new CLAHE implementation techniques, which are the major focus of this paper. To improve the performance of the detection system, better image contrast is essential in image processing applications. In most of the research works, the enhancement of images has been carried out using histogram equalization, which provides uniform gray values for an image. The most popular technique used to improve both the visual quality of the digital image and the diagnostic characteristics of the image is the algorithm contrast-limited adaptive histogram equalization. The limitations of the histogram equalization method are solved by using CLAHE for applications such as night photography, medical imaging, satellite images, fingerprint recognition, and iris recognition [42].

Following preprocessing, two datasets were established: one with the original images and another with the preprocessed images. Both datasets were used for training to evaluate the impact of preprocessing on classification performance, allowing a comparison between the ViT network's performance on raw versus enhanced data. Figure 4 shows samples of the original and the preprocessed images.

The scanned images in the dataset were resized to a uniform dimension of 384x384 pixels with three color channels, ensuring consistency across all samples. These images were then scaled and normalized, maintaining their color properties, and subsequently used in both training and testing phases. The training parameters are listed in Table 1. The network training utilized Stochastic Gradient Descent with Momentum (SGDM) as the optimizer, paired with a stochastic solver to enhance convergence speed and stability. To accelerate the training process, parallel computing was employed on a GPU with 16 concurrent workers. Key training parameters included an initial learning rate set to 1e-4, shuffling at each epoch to prevent overfitting, and an execution environment configured for GPU processing.

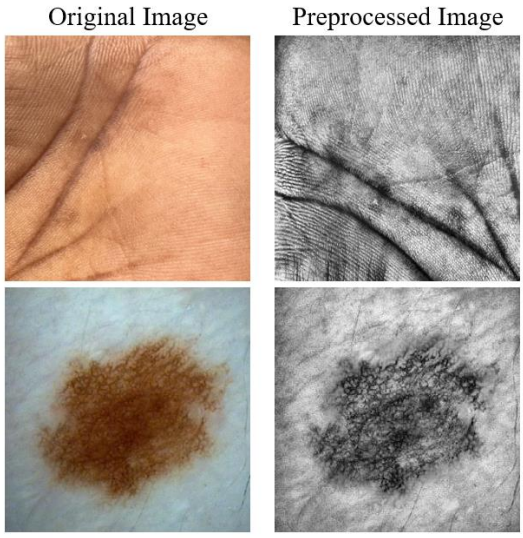

Figure 4. Original and Preprocessed Images

Table 1. The training parameters.

MiniBatchSize		16
MaxEpochs		5
Iterations PerEpoch	Original Dataset:	367
	Preprocessed Dataset:	735
Validation Frequency	Original Dataset:	92
	Preprocessed Dataset:	184

After the training, accuracy for the original dataset was achieved at 0.7383, with the training process completed in 7 hours, 35 minutes, and 58 seconds. Similarly, the training accuracy for the preprocessed dataset improved to 0.7631, though the training duration was longer, taking 10 hours, 26 minutes, and 51 seconds to complete. The confusion matrix obtained with the original dataset is given in Figure 5.

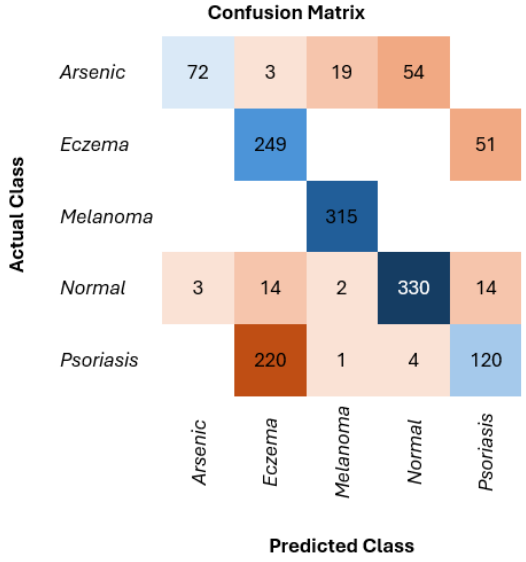

Figure 5. Confusion Matrix obtained with the original dataset.

The confusion matrix analysis reveals the classification results for 1,471 test samples, distributed across five categories: 148 Arsenic, 300 Eczema, 315 Melanoma, 363 Normal, and 345 Psoriasis images. The model achieved varied accuracy across these categories, demonstrating areas of strength and some weaknesses.

In the Arsenic category, 72 images were classified correctly, while 76 were misclassified, indicating room for improvement in distinguishing this condition. For Eczema, the model performed well, with 249 images correctly classified and only 51 misclassified. Notably, all 315 Melanoma images were correctly identified, showcasing the model's strength in Melanoma detection. For Normal skin images, 330 were accurately classified, with 33 images incorrectly labeled. The Psoriasis category, however, presented the greatest challenge, with only 120 images accurately predicted and a substantial 225 misclassified. This pattern highlights the need for further refinement, particularly for Arsenic and Psoriasis, to enhance the model's diagnostic accuracy in these categories. The confusion matrix thus provides valuable insights into performance strengths, as well as areas for targeted improvements. The confusion matrix obtained by the preprocessed dataset is given in Figure 6.

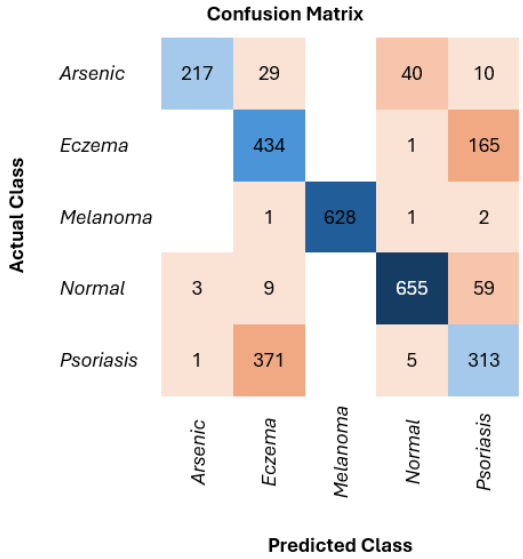

Figure 6. Confusion Matrix obtained with the preprocessed dataset.

The confusion matrix analysis reveals the classification results for a total of 2,942 test samples, which include 296 Arsenic, 600 Eczema, 630 Melanoma, 726 Normal, and 690 Psoriasis images. The model's performance varied across these categories, highlighting both strengths and areas needing improvement.

For the Arsenic category, 217 images were accurately classified, with 79 misclassifications, indicating a moderate level of performance that could be enhanced. The Eczema category showed that 434 images were correctly predicted, but there were 166 misclassifications, suggesting further refinement is necessary for this condition as well. The model demonstrated exceptional accuracy in the Melanoma category, with 626 images correctly identified and only 4 misclassified, indicating a robust capability in detecting this type of skin disease. In the Normal category, 655 images were accurately classified, while 71 were misclassified, showcasing strong performance but also a need to address those misclassifications. In contrast, the Psoriasis category revealed significant challenges, with only 313 images correctly identified and a considerable 377 misclassified. This considerable number of errors indicates that the model's performance in classifying Psoriasis requires substantial improvement. Overall, the breakdown of results highlights the model's strengths in identifying Melanoma and Normal images, while emphasizing the need for enhancements in the classification of Arsenic, Eczema, and particularly Psoriasis. Table 2 summarizes the key metrics observed during the training process, including the progress of training iterations, the duration of each iteration, mini-batch performance, test performance, and errors. This table provides a comprehensive overview of how the model's performance evolved over time, showcasing the effectiveness of the training process. The training and error graphics obtained with the preprocessed dataset is given in Figure 7.

Table 2. Iteration process of the preprocessed dataset.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation Loss
1	1	00:01:29	12.50%	21.52%	48.207	51.771
1	184	00:32:50	93.75%	67.91%	0.5019	11.021
1	368	01:04:02	68.75%	70.12%	0.5567	0.9623
1	552	01:35:09	75.00%	71.35%	0.5287	0.8154
2	736	02:06:14	81.25%	72.64%	0.4622	0.9427
2	920	02:37:35	81.25%	71.86%	0.4056	0.9531
2	1104	03:09:03	100.00%	70.84%	0.1247	12.708
2	1288	03:40:13	75.00%	73.28%	0.5681	0.9984
3	1472	04:11:27	93.75%	75.32%	0.1451	0.7364
3	1656	04:42:55	81.25%	74.34%	0.4851	0.7988
3	1840	05:14:22	81.25%	72.94%	0.3430	0.8805
3	2024	05:45:53	87.50%	73.69%	0.3709	0.7215
4	2208	06:17:06	87.50%	75.66%	0.2403	0.7593
4	2392	06:48:13	75.00%	75.19%	0.4032	0.7778
4	2576	07:19:23	87.50%	75.56%	0.1639	0.9141
4	2760	07:50:37	93.75%	75.49%	0.2257	0.8536
5	2944	08:21:57	93.75%	76.07%	0.1682	0.7871
5	3128	08:53:01	93.75%	76.55%	0.2448	0.7558
5	3312	09:24:12	87.50%	77.02%	0.3339	0.7229
5	3496	09:55:22	81.25%	74.37%	0.3821	10.234
5	3675	10:25:42	81.25%	76.31%	0.2276	0.7182

As illustrated in Figure 7, the test accuracy achieved during the training process using the preprocessed data with the ViTs network reached 76.31%. In comparison, the test accuracy obtained from the training process with the unprocessed data was 73.83%. This indicates that the preprocessing step contributed to an enhancement of 2.48% in overall performance. The features from the dataset were extracted from the ViT network prior to the classification layer, and these features were subsequently classified using a range of traditional classifiers, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Neural Networks, Ensemble methods, Discriminant Analysis, Efficient Logistic Regression, among others. The results of these classification efforts are summarized in Table 3. Upon analysis of the classification results, it is noteworthy that the accuracy increased significantly to 90.6%, reflecting an impressive 18.04% improvement compared to the baseline training accuracy.

This substantial increase underscores the effectiveness of using the ViT network for feature extraction and the subsequent classification process, demonstrating the potential of this approach in enhancing diagnostic performance for skin disease classification.

Figure 8 shows the confusion matrix for the Weighted KNN model which has the highest classification accuracy according to Table 3.

Upon analyzing the confusion matrix for the classical classifier that achieved the highest accuracy, we observe that the dataset comprises a total of 14,716 training and test samples. This dataset includes 1,482 Arsenic images, 3,002 Eczema images, 3,152 Melanoma images, 3,630 Normal images, and 3,448 Psoriasis images.

The classification results for each category are as follows:

- **Arsenic images:** Out of 1,482 images, 1,396 were correctly classified, with 86 misclassifications, indicating strong performance in this category.
- **Eczema images:** Of the 3,002 images, 2,261 were accurately classified, while 741 were misclassified, revealing significant challenges in correctly identifying this condition.
- **Melanoma images:** All 3,139 images were correctly identified, although 13 were misclassified, demonstrating exceptional accuracy in detecting this type of skin disease.
- **Normal images:** From the 3,630 images, 3,513 were correctly predicted, with 117 misclassifications, showing robust performance but highlighting some areas for improvement.
- **Psoriasis images:** Among the 3,448 images, 2,598 were accurately classified, but 850 were misclassified, indicating that this category presents substantial difficulties for the classifier.

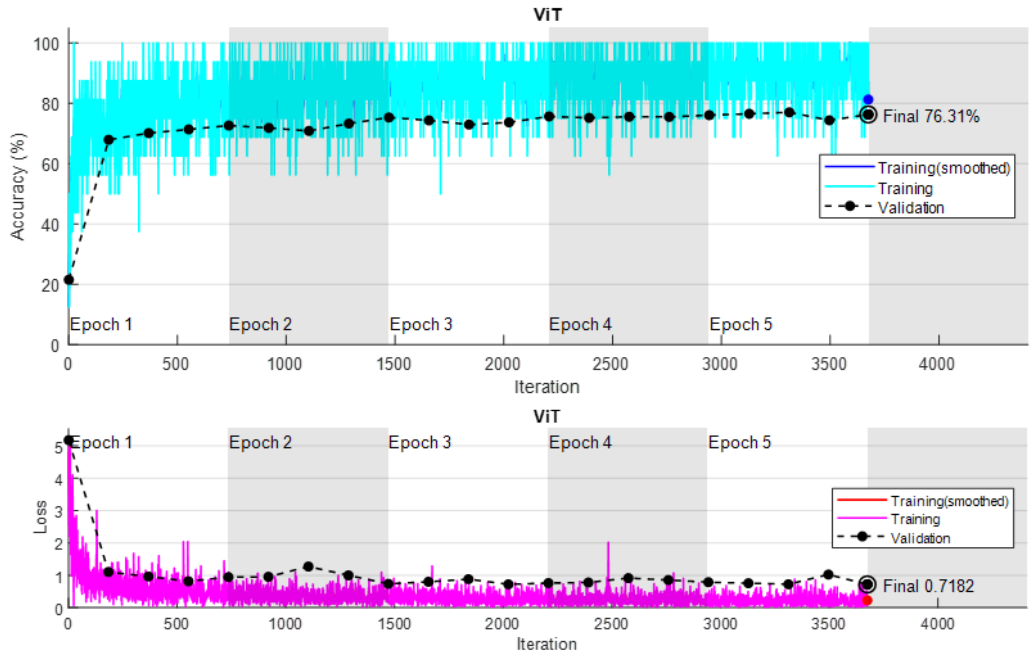

Figure 7. Training and error parameters of the preprocessed dataset.

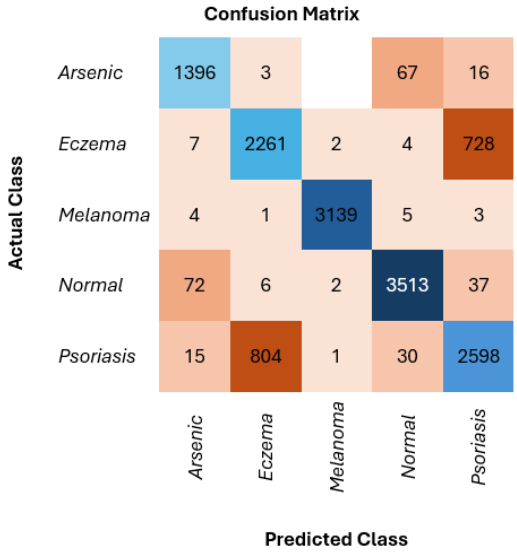

Figure 8. Confusion Matrix obtained with the Weighted KNN model

This analysis illustrates the classifier's strong overall performance, particularly in identifying Melanoma images, where it achieved perfect accuracy. However, it also highlights challenges in accurately classifying Eczema and Psoriasis images, suggesting the need for further refinement in these areas to improve diagnostic accuracy across all skin disease categories. Finally, Figure 9 shows the ROC curve for the weighted KNN model.

After preprocessing the dataset, the Psoriasis class, with 3,448 samples, has the highest number of examples, following the Normal class, which contains 3,002 samples. When examining the confusion matrix for both the original dataset and the preprocessed dataset, it is observed that the Psoriasis class is most frequently misclassified as the Eczema class. A closer comparison of the images from these two classes with the other three classes in the dataset reveals that the visual features of the Psoriasis and Eczema classes are very similar to each other. In contrast, their differences from the other classes are much more distinct. This similarity explains why the Psoriasis class is most often misclassified as the Eczema class. It also highlights that these misclassifications are not primarily due to the number of images in the classes but rather due to the high degree of visual similarity between certain classes.

Table 3. Classification results of the best 15 classifiers.

Model	Sub Model	Accuracy
KNN	**Weighted KNN**	**87.72%**
SVM	Weighted KNN	87.68%
SVM	Quadratic SVM	87.67%
Neural Network	Bilayered Neural Network	87.64%
Neural Network	Medium Neural Network	87.63%
Ensemble	Bagged Trees	87.56%
Neural Network	Narrow Neural Network	87.56%
Neural Network	Trilayered Neural Network	87.47%
KNN	Coarse KNN	87.40%
KNN	Medium KNN	87.20%
KNN	Cosine KNN	87.20%
SVM	Coarse Gaussian SVM	87.16%
KNN	Cubic KNN	87.11%
Discriminant	Quadratic Discriminant	87.05%
Efficient Logistic Regression	Efficient Logistic Regression	87.01%

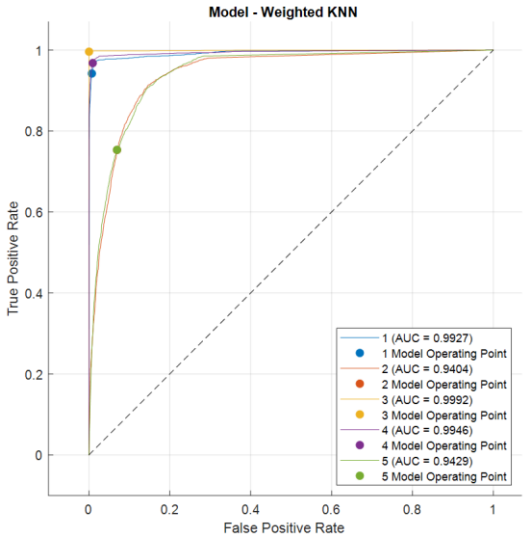

Figure 9. ROC curve for the weighted KNN model.

Hyperparameter tuning is a crucial element for attaining optimal performance using ViTs models, a topic that was not well addressed in the initial discourse. This work carefully changed many hyperparameters to optimize the ViT architecture for skin disease categorization. The learning rate was set at AA, a figure selected to optimize the balance between convergence speed and stability during training. The batch size was configured to 16, enhancing memory efficiency while preserving the model's generalization capability. A maximum of five epochs were utilized for training, guided by early termination conditions to prevent overfitting. The Adam optimizer, together with weight decay regularization, was utilized to improve optimization efficiency and reduce the likelihood of overfitting to limited datasets. The patch size, an essential hyperparameter of the ViT, was established at 16 to maintain an optimal equilibrium between global context and local feature representation in the pictures. The quantity of attention heads and transformer layers was optimized to enhance model performance while maintaining tolerable computing complexity. To mitigate the fluctuation in input photos, data augmentation techniques, including random cropping and flipping, were employed during training to enhance the model's resilience. Hyperparameter tuning was conducted iteratively, with validation accuracy as the principal metric, with grid search and human modifications directing the selection process. Through the optimization of these hyperparameters, the ViT model attained improved feature extraction skills, illustrating its appropriateness for intricate medical picture classification tasks. Subsequent research will investigate sophisticated optimization techniques, including Bayesian optimization, to enhance model efficiency.

The findings of this study demonstrate the efficacy of combining ViTs with classical classifiers for accurate skin disease diagnosis. To provide context and highlight the significance of these results, a comparison with existing studies in the literature was conducted. Table 4 summarizes key metrics, such as accuracy, methods used, datasets employed, and clinical applicability, for several state-of-the-art approaches in skin disease classification. This table highlights the comparative advantages of the proposed framework, particularly its ability to achieve high accuracy with reduced computational requirements. The combination of ViTs for feature extraction and classical classifiers for final diagnosis proves to be a balanced approach, suitable for deployment in resource-constrained clinical settings.

Table 4. Comparison with some of the existing studies

Study	Methodology	Dataset Size and Classes	Accuracy	Clinical Applicability
This Study	ViT + Classical Classifiers	7,356 images, 5 classes	90.6%	High (ViT for global context, low compute cost of classifiers)
DEEPSCAN [29]	ViT-Based Diagnostics	Dermoscopic Dataset, Binary (Benign/Malignant)	92.4%	High (Focus on malignant lesion detection)
Assist-Dermo [30]	Lightweight Separable ViT	Multiclass Dataset, 3 classes	88.7%	Moderate (Efficient but less generalizable)
Srinivasu et al. [12]	MobileNetV2 + LSTM	ISIC 2018, Binary	87.2%	Low (Convolutional focus, limited generalization)
Inthiyaz et al. [7]	CNN-Based Deep Learning	Custom Dataset, 7 classes	85.4%	Moderate (High compute requirements)
Chan et al. [21]	Ensemble of CNNs	ISIC 2017, Multiclass	86.9%	Low (High complexity and compute-intensive)

Parvaiz et al. [38]	ViT + Pretrained ResNet	Custom Dataset, 5 classes	89.1%	High (Focus on medical image generalization)
Jelassi et al. [39]	ViT with Positional Attention	Multiclass Dataset, 4 classes	91.0%	Moderate (ViT adaptation for small datasets)
Das et al. [14]	CNN + SVM Hybrid	Custom Dataset, Binary	84.3%	Low (Limited scalability to multiclass tasks)
Khan et al. [26]	Vision Transformers for Skin Cancer	Dermoscopic Dataset, Multiclass	93.1%	High (Focused on segmentation + classification synergy)

Conclusions

ViTs represent a cutting-edge deep learning technology with significant applications in the medical field, particularly in skin disease diagnosis. Due to their architecture, ViTs require a large number of parameters and, consequently, a substantial dataset for effective learning. The rapid digitization of healthcare has now made this feasible. In contrast, classical classifiers operate with relatively fewer input data, making them suited for cases with limited datasets. In clinical settings, high-resolution imaging techniques such as dermoscopy, confocal microscopy, reflectance confocal microscopy, and Raman spectroscopy are frequently employed for diagnosing skin diseases. ViTs are especially promising in clinical practice as they do not rely on convolutional operations, offering an advantage over traditional convolutional neural networks (CNNs). In this study, the experiments were based on classifying the images in the dataset which were categorized into five classes. We conducted these experiments on preprocessed images from two datasets, one of which consisted of dermatoscopic images. The primary dataset used in this study contained skin images classified into five categories: normal, melanoma, arsenic, psoriasis, and eczema. To enhance image quality, Adaptive Histogram Equalization (AHE) was applied in preprocessing, which improved contrast and highlighted essential details. Following this, ViTs were utilized to extract intricate features from the images, which were then input into traditional machine learning classifiers to aid in the accurate diagnosis of skin conditions. The results underscore the effectiveness of combining ViTs with classical classifiers for medical image classification tasks, suggesting that this approach can significantly enhance diagnostic accuracy in skin disease detection.

This study's results, showcasing the effectiveness of integrating ViTs with traditional classifiers for skin condition identification, possess considerable promise for practical use. The exceptional accuracy attained, especially for situations such as melanoma, highlights the model's dependability in identifying significant skin disorders that necessitate immediate action. ViTs' capacity to identify complex visual characteristics, including small lesions, improves diagnostic accuracy, rendering them appropriate for use in clinical settings where misinterpretation may lead to serious repercussions. The suggested architecture may be integrated into teledermatology systems, facilitating remote diagnosis in under-resourced or rural regions. The lightweight characteristics of classical classifiers allow for a reduction in the processing demands of utilizing ViT-extracted features, rendering the system suitable for edge devices such as smartphones and portable diagnostic instruments. This accessibility might mitigate the global deficit of dermatologists and enhance early detection rates, particularly for high-risk disorders. Moreover, the system's modular architecture facilitates ongoing upgrades, including fresh datasets to enhance generalization progressively. Integrating with electronic health record systems might yield a full diagnostic procedure, merging image analysis with patient history and clinical documentation. To guarantee effective adoption, thorough validation with varied, real-world datasets is required. Creating intuitive interfaces and offering training for healthcare providers would augment its effectiveness. The suggested approach signifies progress in incorporating AI-driven solutions into clinical practice, providing a scalable, efficient, and precise instrument for improving dermatological treatment.

References

[1] M. A. Richard, C. Paul, T. Nijsten, P. Gisondi, C. Salavastru, C. Taieb, ... & EADV Burden of Skin Diseases Project Team. “Prevalence of most common skin diseases in Europe: a population-based study,” *Journal of the European Academy of Dermatology and Venereology*, vol. 36, no. 7, pp. 1088-1096, 2022.

[2] E. T. Anwar, N. Gupta, O. Porwal, A. Sharma, R. Malviya, A. Singh & N. K. Fuloria, “Skin diseases and their treatment strategies in sub-saharan african regions,” *Infectious Disorders-Drug Targets Disorders)*, vol. 22, no. 2, pp. 41-54, 2022.

[3] J. A. Rossow, F. Queiroz-Telles, D. H. Caceres, K. D. Beer, “A One Health Approach to Combatting Sporothrix brasiliensis: Narrative Review of an Emerging Zoonotic Fungal Pathogen in South America,” *Journal of Fungi*, vol. 6, no. 4, p. 247, 2020.

[4] D. M. Elston, “Occupational skin disease among health care workers during the coronavirus (COVID-19)

epidemic," *Journal of the American Academy of Dermatology*, vol. 82, no. 5, p. 1085, 2020.

[5] V. R. Balaji, S. T. Suganthi, R. Rajadevi, V. K. Kumar, "Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier," *Measurement*, vol. 163, p. 107922, 2020.

[6] Y. Liu, A. Jain, C. Eng, D. H. Way, K. Lee, P. Bui, et al., "A deep learning system for differential diagnosis of skin diseases," *Nature medicine*, vol. 26, no. 6, pp. 900-908, 2020.

[7] S. Inthiyaz, B. R. Altahan, S. H. Ahammad, V. Rajesh, R. R. Kalangi, L. K. Smirani and A. N. Z. Rashed, "Skin disease detection using deep learning," *Advances in Engineering Software*, vol. 175, p. 103361, 2023.

[8] N. Hameed, A. M. Shabut, and M. K. Ghosh, "Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques," *Expert Systems with Applications*, vol. 141, p. 112961, 2020.

[9] N. Melnyk, I. Vlasova, W. Skowrońska, and A. Bazylko, "Current knowledge on interactions of plant materials traditionally used in skin diseases in Poland and Ukraine with human skin microbiota," *International Journal of Molecular Sciences*, vol. 23, no. 17, p. 9644, 2022.

[10] H. Li, Y. Pan, J. Zhao, and L. Zhang, "Skin disease diagnosis with deep learning: A review," *Neurocomputing*, vol. 464, pp. 364-393, 2021.

[11] SS Han, I Park, SE Chang, W Lim, MS Kim, "Intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin," *Dermatology*, vol. 140, no. 9, pp. 1753-1761, 2020.

[12] P.N. Srinivasu, J.G. SivaSai, M.F. Ijaz, A.K. Bhoi, and W. Kim, "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, p. 2852, 2021.

[13] N. Ghaffar Nia, E. Kaplanoglu, and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," *Discover Artificial Intelligence*, vol. 3, no. 1, p. 5, 2023.

[14] K. Das, C. J. Cockerell, A. Patil, and P. Pietkiewicz, "Machine learning and its application in skin cancer," *International Journal of Environmental Research and Public Health*, vol. 18, no. 24, p. 13409, 2021.

[15] J. Jayashree, S. S. V. Nukala, and J. Vijayashree, "The rise of AI in the field of healthcare," *Cognitive Machine Intelligence*, pp. 221-244, 2024.

[16] A. Mubeen and U. N. Dulhare, "Metaheuristic Algorithms for the Classification and Prediction of Skin Lesions: A Comprehensive Review," *Machine Learning and Metaheuristics: Methods and Analysis*, pp. 107-137, 2023.

[17] A. Gomolin, E. Netchiporouk, and R. Gniadecki, "Artificial intelligence applications in dermatology: where do we stand?," *Frontiers in medicine*, vol. 7, p. 100, 2020.

[18] T. B. Jutzi, E. I. Krieghoff-Henning, and T. Holland-Letz, "Artificial intelligence in skin cancer diagnostics: the patients' perspective," *Frontiers in medicine*, vol. 7, p. 233, 2020.

[19] J. Kawahara and G. Hamarneh, "Visual Diagnosis of Dermatological Disorders: Human and Machine Performance," *arXiv preprint arXiv:1906.01256*, 2019.

[20] S. Mishra, S. Chaudhury, H. Imaizumi, and T. Yamasaki, "Assessing Robustness of Deep learning Methods in Dermatological Workflow," *arXiv preprint arXiv:2001.05878*, 2020.

[21] S. Chan, V. Reddy, B. Myers, Q. Thibodeaux et al., "Machine Learning in Dermatology: Current Applications, Opportunities, and Limitations," *Dermatology and therapy*, vol. 10, pp. 365-386. 2020.

[22] D. Sonntag, F. Nunnari, and H. J. Profitlich, "The Skincare project, an interactive deep learning system for differential diagnosis of malignant skin lesions. Technical Report," *arXiv preprint arXiv:2005.09448*, 2020.

[23] G. J. Chowdary, "Machine Learning and Deep Learning Methods for Building Intelligent Systems in Medicine and Drug Discovery: A Comprehensive Survey," *arXiv preprint arXiv:2107.14037*, 2021.

[24] M. Qays Hatem, "Skin lesion classification system using a K-nearest neighbor algorithm," *Visual Computing for Industry, Biomedicine, and Art*, vol. 5, no. 1, p. 7, 2022.

[25] Z. Li, K. Christoph Koban, T. Ludwig Schenck, R. Enzo Giunta et al., "Artificial Intelligence in Dermatology Image Analysis: Current Developments and Future Trends," *Journal of clinical medicine*, vol. 11, no. 22, p. 6826, 2022.

[26] S. Khan, H. Ali, and Z. Shah, "Identifying the role of vision transformer for skin cancer—A scoping review," *Frontiers in Artificial Intelligence*, vol. 6, 1202990, 2023.

[27] T. Lai, "Interpretable Medical Imagery Diagnosis with Self-Attentive Transformers: A Review of Explainable AI for Health Care," *BioMedInformatics*, vol. 4, no. 1, pp. 113-126. 2023.

[28] A. Khan, Z. Rauf, A. Rehman Khan, S. Rathore et al., "A Recent Survey of Vision Transformers for Medical Image Segmentation," *arXiv preprint arXiv:2312.00634*, 2023.

[29] V. Ravi, T. J. Alahmadi, T. Stephan, P. Singh and M. Diwakar, "DEEPSCAN: Integrating Vision Transformers for Advanced Skin Lesion Diagnostics," *The Open Dermatology Journal*, vol. 18, no. 1, 2024.

[30] Q. Abbas, Y. Daadaa, U. Rashid and M. E. Ibrahim, "Assist-dermo: A lightweight separable vision transformer model for multiclass skin lesion classification," *Diagnostics*, vol. 13, no. 15, p. 2531, 2023.

[31] E. G. Espinosa, J. S. R. Castilla and F. G. Lamont, F. G. "Skin Disease Pre-diagnosis with Novel Visual Transformers," *In Workshop on Engineering Applications (pp. 103-113)*. Cham: Springer Nature Switzerland, 2024.

[32] S. Aladhadh, M. Alsanea, M. Aloraini, T. Khan, S. Habib and M. Islam, "An effective skin cancer

classification mechanism via medical vision transformer," *Sensors*, vol. 22, no11, p. 4008, 2022.

[33] https://www.kaggle.com/datasets/sayedhossainjobayer/skin-diseases-identification, Last Accessed On: 30.10.2024.

[34] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv: 2010.11929*, 2020.

[35] K. Han, Y. Wang, H. Chen, X., Chen, J. Guo, Z. Liu & D. Tao, "A survey on visual transformer," *arXiv preprint arXiv:2012.12556*, 2020.

[36] M. M. Naseer, K. Ranasinghe, S. H. Khan, "Intriguing properties of vision transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 23296-23308, 2021.

[37] Y. H. Cao, H. Yu, and J. Wu, "Training Vision Transformers with Only 2040 Images," *European Conference on Computer Vision (pp. 220-237)*, 2022.

[38] A. Parvaiz, M. Anwaar Khalid, R. Zafar, H. Ameer et al., "Vision Transformers in Medical Computer Vision - A Contemplative Retrospection," *Engineering Applications of Artificial Intelligence*, vol. pp. 122, 2022.

[39] S. Jelassi, M. E. Sander, and Y. Li, "Vision Transformers provably learn spatial structure," *Advances in Neural Information Processing Systems*, vol. 35, pp. 37822-37836, 2022.

[40] V. A. Nguyen, K. Pham Dinh, L. Tung Vuong, T. T. Do et al., "Vision Transformer Visualization: What Neurons Tell and How Neurons Behave?," *arXiv preprint arXiv:2210.07646*, 2022.

[41] S. A. Khan, S. Hussain, and S. Yang, "Contrast enhancement of low-contrast medical images using modified contrast limited adaptive histogram equalization," *Journal of Medical Imaging and Health Informatic*s, vol. 10, no. 8, pp. 1795-1803. 2020.

[42] V. Banupriya and A. Kalaivani, "Improved retinal fundus image quality with hybrid image filter and enhanced contrast limited adaptive histogram equalization," *International Journal of Health Sciences*, vol. I, pp. 9234-9246, 2022.