A Systematic Literature Review on Machine Learning Techniques for Skin Disease Classification

Fadilah Karamun Nisaa Nadiyah[1], **Nayla Nur Alifah**[2], **Sri Nurdiati**[3], **Elis Khatizah**[4], **Mohamad Khoirun Najib**[5,*]
[1,2,3,4,5]*Applied Mathematics, School of Data Science, Mathematics and Informatics, IPB University, Bogor 16680, Indonesia*
E-mail : fadilahkaramun.nisaa@apps.ipb.ac.id[1], *naylalifah@apps.ipb.ac.id*[2], *nurdiati@apps.ipb.ac.id*[3], *elis_khatizah@apps.ipb.ac.id*[4], *mkhoirun@apps.ipb.ac.id*[5,*]
**Corresponding author*
Received 2 May 2025; Revised 10 May 2025; Accepted 13 May 2025

Abstract - Skin diseases pose health challenges that necessitate precise diagnosis for evaluation, ultimately influencing treatment choices. Medical imaging plays a vital role in the diagnostic procedure. Machine learning technology can aid in the classification of skin diseases through image data, achieving notable accuracy in diagnosis. This research aims to explore the machine learning algorithms that can be employed to create systems for classifying skin diseases based on images. The methodology used is a Systematic Literature Review (SLR), which serves to provide an extensive overview of how machine learning is applied in the classification of skin diseases. The literature search approach utilized the Boolean method, specifically applied to the Scopus database. The chosen articles underwent screening based on established inclusion and exclusion criteria. The findings reveal that the Convolutional Neural Network (CNN) is the most commonly used machine learning algorithm, which has demonstrated the highest classification accuracy.

Keywords - Skin Disease, Machine Learning, Classification, CNN.

1. INTRODUCTION

Skin diseases are represent significant global health challenge. That is complicated by their intricate nature and the extensive time required for accurate diagnosis. Imaging plays a crucial role, serving as a foundational element prior to any surgical or treatment decisions. It enables the establishment of preliminary knowledge and facilitates accurate diagnoses [1]. Consequently, medical imaging has evolved into an essential tool that initiates the treatment process for various diseases, encompassing stages from detection to evaluation and ultimately leading to treatment decisions. In particular, skin diseases represent a medical domain where imaging significantly contributes to the detection, diagnosis, and management of conditions [2].

Computer-aided automatic identification of skin diseases from images can minimize human error and speed up the detection process, thereby assisting clinicians in making diagnoses more efficiently and facilitating timely patient treatment. Typically, there are two main methodological frameworks used. The first is the traditional method, which relies on manually defined features such as color and texture for the identification and detection of skin diseases. However, the selection of relevant features is time-consuming and crucial, as it directly affects classification accuracy. The second framework is an evolutionary approach that incorporates artificial intelligence (AI) and deep learning techniques. This approach enables automatic and effective feature learning of skin disease characteristics by utilizing established image segmentation algorithms, which categorize images based on pixel intensity, edges, and regions. [3-6].

On the other side, machine learning and image processing techniques can help achieve high accuracy in skin diagnosing at the initial stage. Images processing plays an effective role in diagnosis the skin diseases. Machine learning algorithms can used for the classification task [1].

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

Traditional machine learning methods, such as decision trees, random forests, and support vector machines (SVM), have been widely applied to classify skin diseases due to their proficiency with structured data [1, 7, 8, 9, 10]. They were predominantly used before 2010 and remained popular from 2010 to 2017 [11]. The year 2017 marked a significant advancement in classification image processing with the introduction of the Vision Transformer (ViT) architecture [12]. Following this, enhancements like the Separable Vision Transformer (SVT), which merges Vision Transformers with depthwise-separable convolutions [13], as well as SkinDistilViT and SkinSwinViT, have been launched to enhance classification accuracy in dermatological imaging. [14,15].

In addition to traditional machine learning methods and Vision Transformer (ViT) approaches, Convolutional Neural Networks (CNNs) have been extensively used. CNNs are a type of machine learning algorithm that are particularly effective at processing visual data. The convolutional filters used in CNNs are essential for automatically extracting relevant features from images. Although image classification poses significant challenges, the use of deep CNN architectures, supported by powerful graphics processing units, has significantly improved the speed and accuracy of this task [16, 17].

Various architectures of convolutional neural networks (CNNs), such as AlexNet, InceptionV3, ResNet, and MobileNet, have been thoroughly researched for skin disease classification. These models are effective in extracting relevant features from images and performing complex pattern recognition tasks [7, 16, 18]. Besides the models' architectures, image processing techniques are critical for improving diagnostic accuracy [1]. Common techniques include resizing, normalization, contrast enhancement, and noise reduction. Additionally, many research studies utilize data augmentation strategies like color space transformations, rotations, flips, and scaling to boost data variability and decrease the likelihood of overfitting, especially in datasets with a limited number of samples [1,8,19].

The scope of skin conditions addressed in these studies is diverse, encompassing both common dermatological disorders such as acne, eczema, vitiligo, and more serious conditions like melanoma, carcinoma, and Monkeypox [18,7]. For instance, [18] a study carried out in India examined conditions including eczema, psoriasis, vitiligo, melanoma, carcinoma, and blue nevus. In contrast, [7] represented a broader international collaboration (Saudi Arabia, Egypt, Australia, South Korea, Malaysia, USA) classifiying diseases like Monkeypox, Chickenpox, and Measles.

As research in this area continues to grow, it is essential to map out the current research landscape to identify common methodologies, evaluate the effectiveness of models, and uncover existing gaps. This paper presents a Systematic Literature Review (SLR) following PRISMA guidelines to assess the present status of research on skin disease classification using machine learning methods. The review investigates several key research questions, such as which machine learning algorithms are used, the level of data augmentation employed, the specific skin diseases under study, the leading countries contributing to this research area, and the performance metrics commonly used to evaluate models. The findings from this review are expected to guide future research efforts and help develop more efficient and accessible diagnostic tools for skin diseases.

2. RESEARCH METHOD

The research methodology was based on a Systematic Literature Review (SLR), adhering to the guidelines set forth by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). PRISMA provides both a checklist and a flowchart that outline four key stages: identification, screening, eligibility, and inclusion. The aim of the PRISMA flowchart is to improve the quality of systematic reviews and meta-analyses [20]. According to [21], six fundamental components are involved in creating an SLR, which include evaluating the literature phases, describing the methodology, formulating the research question, and identifying potential biases during a comprehensive literature review. To explore the stages of literature assessment,

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

[22] the process involves defining the topic, articulating the research question, identifying relevant keywords, searching the electronic paper repositories, reviewing and assessing publications, collecting data, cleaning the data, testing the publications, constructing and revising summary tables, drafting the methodology, and ultimately writing the final paper. This research utilizes descriptive analysis in conjunction with descriptive statistics to extract data, followed by analysis and synthesis to answer the research questions. A comprehensive overview of this process is illustrated in Figure 1 [23].

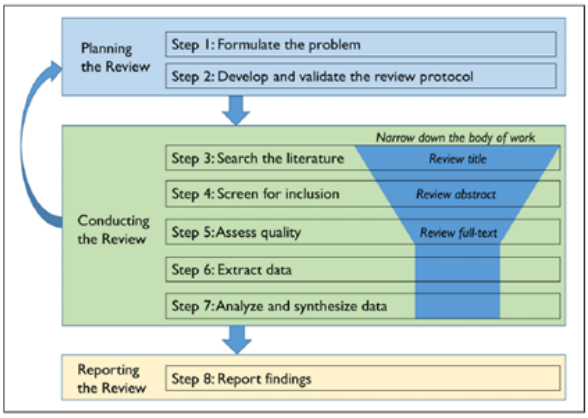

Figure 1. The Systematic Literature Review (SLR) Process

2.1. Reaseach Strategy

The search implemented a search methodology for articles using the Boolean Technique, specifically designed to extract essential search terms from the article collection. The specific search terms applied were 'skin disease AND machine learning,' as shown in Figure 2. The source for the database utilized in the article search was sourced from Scopus. The articles obtained through this search method conformed to the inclusion and exclusion criteria described in Table 1. The screening phase involved reviewing titles, abstracts, and keywords to identify studies relevant to the research focus. Articles that advanced beyond this preliminary stage underwent a comprehensive examination of the full text to ensure they aligned with the study's objectives. This approach was carried out methodically and transparently, thereby reducing bias and improving the credibility of the literature base utilized for analysis.

Table 1. Inclusion and Exclusion Criteria

No.	Inclusion Criteria	Excision Criteria
1	Must be articles published in 2014 – 2025.	Articles are written in another language.
2	The article should include the development and implementation of the Machine learning for classification and detection skin diseases.	Machine learning does not focus on classification and detection of skin diseases.
3	Must be written in English	Articles do not use images data.
4	QA Value>80%, QA questions refer to Table 3.	

2.2. Research Questions

The suggestions detailed in this study, as referenced by [24], focus on assessing methodologies aimed at reducing bias and improving the trustworthiness of article evaluations in this Systematic Literature Review (SLR). The articles acquired through the search strategy are evaluated according to their titles and abstracts to respond to the research question outlined in Table 2.

Table 2. Research Question

No.	Research Question
RQ1	What machine learning algorithms are applied?
RQ2	Are they applied data augmentation in image processing?
RQ3	What types of skin disease problems are addressed in the reviewed studies?
RQ4	Which countries have developed the classification of skin disease using machine learning?
RQ5	What evaluation metrics are used in the studies?

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

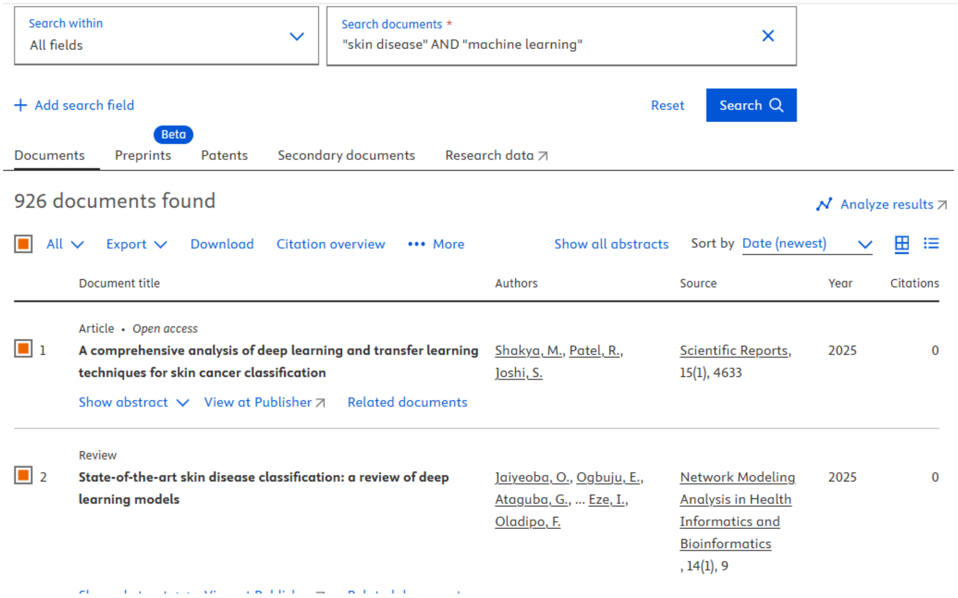

Figure 2. The search terms used in the article collection

2.3. Quality Assurance

An important element to consider, in addition to the criteria for inclusion and exclusion the assessment of quality. A quality assessment checklist consisting of six criteria was developed and used by [25] to create a systematic method for assessing the quality of the selected research articles for further analysis (N=12). The quality assessment checklist is illustrated in Table 3. Each criterion on the checklist was evaluated using a three-point scale, where 'Yes' was given 1 point, 'No' received 0 points, and 'Partially' was awarded 0.5 points. Consequently, each study could receive a score ranging from 0% to 100%, with a higher total score indicating a more satisfactory response to the research questions. Articles will be considered for inclusion in the list if their quality assessment score exceeds 80%.

Table 3. Quality Assessment

No.	Quality Assessment
QA1	Is the objective of the study clearly stated?
QA2	Is the information presented clearly and concisely?
QA3	Does the study provide a sufficient explanation of its methodology?
QA4	Do the findings contribute to the understanding of skin disease classification using image data?
QA5	Are the image data sources transparently disclosed?
QA6	Are the conclusions logically derived, clearly identified and consistent with the overall flow of the paper?

3. RESULTS AND DISCUSSION

The search conducted using the keywords outlined in the research strategy yielded 926 results, comprising articles, conference presentations, reviews, letters, and book chapters published from 2014 to 2025, all focused on skin diseases and machine learning. After applying the predetermined inclusion and exclusion criteria, the dataset was refined to consist solely of those identified as articles. These chosen scientific articles were sourced from a single database, Scopus. In the course of this process, two articles were flagged as duplicates. The remaining articles were assessed based on their titles and abstracts following the criteria outlined in Table 1.

Following the initial review of titles and abstracts, 156 articles were scrutinized regarding the research question stated in Table 2. From this assessment, 11 articles were selected for quality evaluation, as presented in Table 3, adhering to the evaluation methodology previously discussed. The outcomes from the analysis of these 11 articles are summarized in Table 4. Three articles

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

were omitted due to a percentage score lower than 80%, resulting in the inclusion of only 8 papers in this study. The specifics of the article selection process are illustrated in Figure 3.

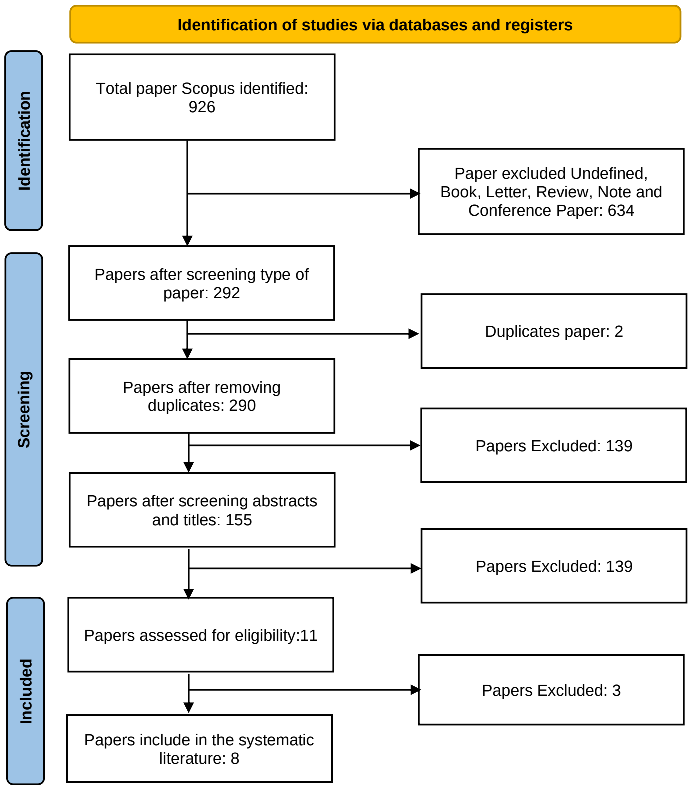

Figure 3. The Systematic Literature Review (SLR) diagram

3.1. Quality Assurance Results

Utilizing the questions in Table 4 along with the evaluation criteria mentioned earlier, the assessor evaluated the 11 articles included in this research. The highest possible percentage score is 100%. A number of articles received scores lower than 80%, signifying that they did not fulfill the predetermined quality assurance standards. As a result, this study excludes any articles with a quality assurance score under 80%. The assessment results for the 11 articles are shown in Table 4 where QA1-QA6 indicate Quality Assurance 1-6, A1-A2 refer to assessors 1-2, and S1-S11 correspond to the sources of articles 1-11.

Table 4. Quality Assurance Results

Paper No.	**QA1**		**QA2**		**QA3**		**QA4**		**QA5**		**QA6**		**Pct**
	A1	A2	A1	A2	A1	A2	A1	A2	A1	A2	A1	A2	**(%)**
S1	1	1	0.5	1	1	1	1	1	1	1	1	1	91.7
S2	1	1	1	1	1	1	1	1	1	1	1	1	100
S3	1	1	1	1	1	1	1	1	1	1	1	1	100
S4	1	1	1	1	1	1	1	1	0.5	0.5	1	1	83.3
S5	1	1	1	1	1	1	0	0	0	0	1	1	66.7
S6	1	1	1	1	1	1	1	1	0.5	0.5	1	1	83.3
S7	1	1	1	1	1	1	1	1	1	1	1	1	100
S8	1	1	1	1	1	0.5	0	0	1	1	1	1	75
S9	1	1	1	1	1	1	1	1	1	1	1	1	100
S10	0.5	1	1	1	1	1	0	0	1	1	1	1	75
S11	1	1	1	1	1	1	1	1	1	1	1	1	100

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

3.2. Descriptive Analysis

Research on the classification and detection of skin diseases through machine learning has seen a remarkable rise over the past decade. As shown in Figure 4, there has been a significant rise in the number of scholarly articles published each year. The time frame from 2023 to 2024 reflects a remarkable upward trajectory, with an average annual growth rate of 66.42%. In 2024 alone, a total of 265 articles were released, with 60 of those published in journals categorized as Q1. Figure 5 illustrates that the primary emphasis of studies utilizing skin disease data is on classification (58.43%), followed by detection (28.76%), segmentation (8.76%), and monitoring (4.04%).

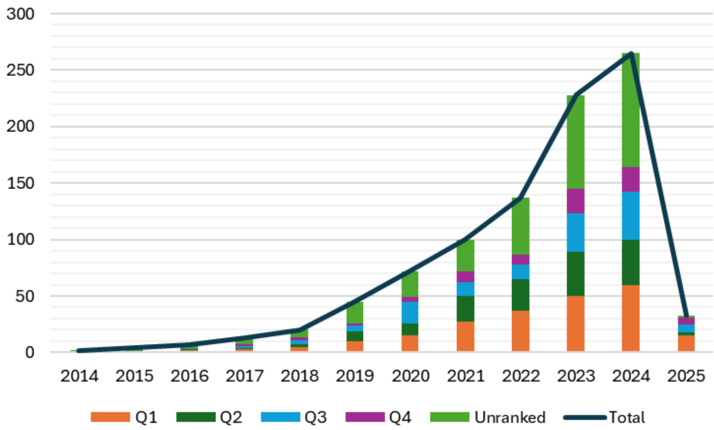

Figure 4. Publication Each Year

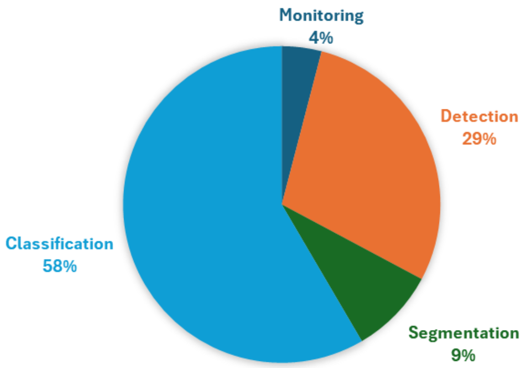

Figure 5. Application Skin Diseases

The classification of skin disease data mainly utilizes Convolutional Neural Networks (CNN) along with traditional models like Support Vector Machines (SVM) and Random Forest. Classification techniques are often paired with detection, leveraging either symptom-focused or image-focused data. Segmentation techniques are primarily employed to pinpoint affected areas on the skin, thereby improving the accuracy of subsequent detection and classification methods. Numerous research efforts also concentrate on monitoring, especially through the evaluation of health applications related to dermatology.

Most research publications concerning skin diseases encompass conditions such as melanoma, psoriasis, eczema, acne, and other similar ailments, as illustrated in Figure 6. Furthermore, various research have progressed in developing models for the classification and detection of skin cancers and tumors.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

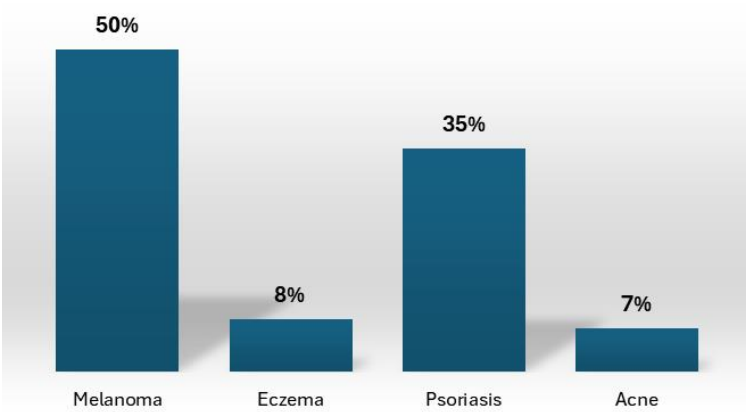

Figure 6. Application Skin Diseases

Based on Figure 7, the top countries engaged in researching machine learning models for skin diseases are India (28.49%), China (15.11%), the USA (13.24%), Spain (3.42%), and the UK (3.25%). Citation analysis highlights five nations as the major hubs of research in this field: Austria (16.58%), the USA (14.83%), India (13.98%), China (11.40%), and Korea (5.97%). This prominence reflects both the quantity and quality of their publications, which are often backed by access to advanced datasets, collaborations with dermatological institutions, and the development of cutting-edge machine learning frameworks.

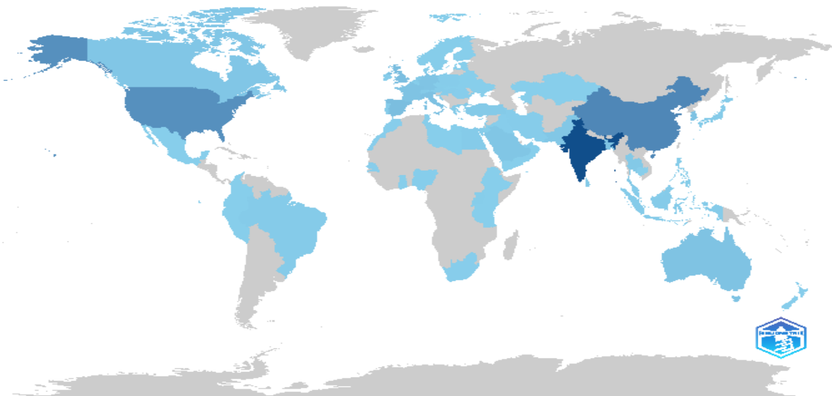

Figure 7. Application Skin Diseases

3.3. Answer to Research Questions

The analysis presented in Table 5 which summarizes the articles reviewed, underscores the significance of machine learning algorithms in the classification and detection of skin diseases. The most commonly employed method in these studies is Convolutional Neural Networks (CNN), often compared with traditional machine learning approaches such as Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), Gaussian Naïve Bayes (NB), and Linear Discriminant Analysis (LDA). Research conducted by [7,9] specifically assessed the efficacy of the CNN architecture in comparison to traditional machine learning methods, evaluating outcomes based on metrics such as accuracy, precision, sensitivity, specificity, and F1-score. In the study by [5], the authors employed Negative Predictive Value (NPV) and Receiver Operating Characteristic (ROC) Curve Analysis to assess performance models, in conjunction with statistical significance tests including ANOVA and Wilcoxon Tests. Moreover, studies [16,19,26] compared different CNN architectures, which included AlexNet, VGG19, GoogLeNet, ResNet-50, Inception V3, and

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

MobileNetV2. Interestingly, [9] reported that the InceptionV3 model reached an accuracy of 95.14%, while [19] noted accuracy findings ranging from 87.25% ± 2.24% to 86.63% ± 5.78%.

The utilization of image datasets for classifying skin diseases typically involves various preprocessing steps, with data augmentation being a significant aspect. Data augmentation is used to balance class distributions or to increase the dataset's diversity by generating more variations of existing images. The augmented data are use as the training set to support the model's enhanced learning. These variations result from modifications to particular features of the images, enabling the model to more accurately recognize disease patterns, even in different imaging scenarios. Overfitting is also avoided by data augmentation [9,29,30]. Numerous studies, such as [7,9,16,18,26,27] have implemented data augmentation techniques. Augmentation techniques applied encompass resizing, flipping, random scaling, rotation, reflection, translation, brightness adjustments, color space transformation and shear transformations. Research referenced in [16] indicates that data augmentation can significantly influence the accuracy of skin disease classification. Additionally, this approach proves particularly advantageous in scenarios where the available datasets are limited. For instance, [7], which refers to [31], demonstrated that data augmentation was essential in enhancing a small dataset related to Monkeypox, thereby improving the model's performance.

In contrast, studies that did not implement data augmentation concentrated on preprocessing techniques aimed to enhance image quality and ensure consistency within the dataset, as shown in [1,8,19]. Specifically, traditional machine learning techniques often depend on preprocessing procedures to prepare the data [8]. The emphasis of preprocessing is on improving the cleanliness and uniformity of the data rather than increasing its diversity. This issue arises because, in classification tasks involving multiple disease categories, the likelihood of misclassifications increases when images are affected by external influences. As noted in [19], data augmentation was not employed; instead, preprocessing was prioritized due to the sufficiency of the existing dataset, which was considered large and representative enough for each category, thereby ensuring adequate model performance.

A review of the literature shows that many skin diseases are often associated with skin lesions, which include melanocytic nevus, basal cell carcinoma, psoriasis, seborrheic keratosis, acne, cherry angioma, melanoma, squamous cell carcinoma, and various skin conditions, both malignant and benign, such as eczema, monkeypox, chickenpox, and measles. A significant number of studies utilized the HAM10000 dataset, which categorizes skin conditions into seven distinct groups: Actinic Keratoses, Basal Cell Carcinoma, Benign Keratosis-like Lesions, Dermatofibroma, Melanoma, Melanocytic Nevi, and Vascular Lesions.

In addition, several studies made use of datasets from dermatological image banks along with clinical patient data, customized to align with the specific research aims and the prevalent skin conditions in various regions. his variation in data sources not only emphasizes the diversity of research objectives but also highlights the significance of local dermatological concerns. For instance, [18] conducted in India focused on chronic skin conditions such as eczema, psoriasis, vitiligo, melanoma, carcinoma, and blue nevus. In contrast, [7] involved a collaboration across several countries, including Saudi Arabia, Egypt, Australia, South Korea, Malaysia, and the United States, with an emphasis on infectious skin diseases like monkeypox, chickenpox, and measles.

The amalgamation of the reviewed literature indicates that a range of evaluation metrics is commonly used to assess the performance of CNN architectures, with most studies measuring effectiveness through metrics like accuracy, precision, sensitivity, specificity, and F-measure. Specifically, specificity is described as the true negative rate, calculated based on the correct identification of negatives. The formula for specificity is provided in the subsequent equation:

$$\text{Specificity} = TN/(TN+FP)$$

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

Secondly, the sensitivity refers to the rate of true positive which is predicated as correctly positive. The sensitivity is illustrated in the following equation:

$$\text{Sensitivity} = TP/(TP+FN)$$

Thirdly, the precision refers to the consistency of the results. The precision is illustrated in the following equation:

$$\text{Precision} = TP/(TP+FP)$$

Fourthly, the accuracy refers to the overall correctness of the proposed classifier. The accuracy is illustrated in the following equation:

$$\text{Accuracy} = (TP+TN)/(TP+FN+TN+FP)$$

Finally, the F-measure refers to the measure of tests accuracy. The F-measure is illustrated in the following equation:

$$\text{F-measure} = (2\times\text{precision}\times\text{sensitivity})/(\text{precision}+\text{sensitivity})$$

Table 5. Summary of articles

Source	Aims	Findings
[9] (S1)	to categorize skin lesions by employing machine learning algorithms and Convolutional Neural Networks (CNN) on dermoscopic images sourced from the HAM10000 dataset, with the objective of enhancing the accuracy of skin cancer detection through advanced image analysis techniques and computational classification methods. The study utilized various algorithms including Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), Gaussian Naïve Bayes (NB), Linear Discriminant Analysis (LDA), and Convolutional Neural Networks (CNN), while also incorporating the Inception V3 model.	• This work utilized the HAM10000 dataset, which comprises 10015 images of skin lesions, each with a resolution of 600 x 450 pixels. The skin lesions are divided into seven categories: Actinic Keratoses, Basal Cell Carcinoma, Benign Keratosis-like Lesions, Dermatofibroma, Melanoma, Melanocytic Nevi, and Vascular Lesions. • Data augmentation techniques were employed, including horizontal flipping and various random transformations (such as rotations and shearing). • The models' performance was evaluated using metrics such as accuracy, precision, recall, and F1-score. • The InceptionV3 model achieved an accuracy of 95.14%, while the CNN model attained an accuracy of 95.18%. • The proposed CNN model demonstrates significantly less computational overhead in total training time compared to the InceptionV3 model, attributed to its greater number of layers.
[19] (S2)	develop a decision support system that combines human expertise with artificial intelligence to enhance diagnostic choices, overcome the shortcomings of computer-aided diagnosis systems by merging clinical knowledge with machine learning techniques, and boost decision-making, increase diagnostic precision, and augment human intelligence in dermatological evaluations.	• This utilized a dataset of skin diseases including melanocytic nevus, BCC, psoriasis, and seborrheic keratosis sourced from the dermatology department at Peking Union Medical College Hospital. The clinical database comprises over 28,000 dermoscopic images analyzed with MoleMax HD 1.0 dermoscopic devices. • Data augmentation was not employed; only preprocessing measures were applied (such as hair reduction, dermoscopic gel reduction, and brightness adjustment). • This study implemented a deep learning model for classification, specifically a CNN with the Inception V3 architecture, achieving an accuracy of 87.25% ± 2.24% on Dataset A (which included 1067 images and was imbalanced), and an accuracy of 86.63% ± 5.78% on Dataset B (comprising 528 images and balanced across categories). • Metrics for Dataset B included Precision, Recall, and F1, with the highest values recorded for melanocytic nevus, BCC, and psoriasis, and the lowest for seborrheic keratosis. • To evaluate and analyze misclassifications, a semantic error analysis was conducted, categorizing errors into four groups: Characteristics of Disease, Multiple Diseases, Interference Factors, and Limitations of Algorithm Accuracy.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

Source	Aims	Findings
[1] (S3)	developing a fast and effective model for classifying and diagnosing skin disorders through image processing methods and machine learning algorithms (SVM, Random Forest, and K-Nearest Neighbors) to facilitate early detection and enhance diagnostic precision.	• This study utilized a collection of 377 images depicting skin diseases, gathered from Dermnet NZ and Atlas Dermatologico, which included 80 images of Acne, 37 images of Cherry Angioma, 80 images of Melanoma, and 180 images of Psoriasis. • SVM achieved the highest performance with an accuracy of 90.7%, Precision of 91%, Recall of 90.8%, and an F1-score of 90.8%. In contrast, the Random Forest algorithm recorded an accuracy of 84.2%, while the K-Nearest Neighbors (K-NN) showed an accuracy of 67.1%. • No data augmentation techniques were implemented; only preprocessing methods were applied (including resizing images, using a median filter for noise reduction, converting color images to grayscale, and normalizing pixel values). The feature extraction process in this research employed Gabor filters, Entropy, and the Sobel edge detector.
[8] (S4)	develop an automated system for the early identification and categorization of human skin tumors utilizing image analysis, and enhance the classification performance by integrating deep learning-derived feature extraction with an upgraded machine learning classifier (Enhanced SVM).	• The dataset comprises 15,000 high-resolution images (2598×1944 pixels) featuring various types of skin tumors, including melanoma, basal cell carcinoma, and squamous cell carcinoma, which encompass both malignant and benign cases. • Only preprocessing techniques, such as enhancement and noise reduction, were applied; no data augmentation was utilized. • This research evaluated performance using metrics such as accuracy, specificity, sensitivity, precision, F-measure, and recognition rate. • The use of an enhanced classifier (ISVM) with optimal feature selection led to achieving high accuracy (95%) and recognition rate (93%).
[10] (S5)	developing a predictive model for skin diseases through machine learning methods, focusing on binary classification and multi-model ensemble strategies, while identifying the top-performing classifier for dermatological data by utilizing Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Naive Bayes (NB), and Support Vector Machine (SVM).	• Used a dermatology dataset sourced from the UCI repository, which contains 366 instances and 34 attributes. • The Extra Trees (ET) ensemble classifier reached the highest accuracy of 99% in predicting skin diseases, while the Logistic Regression (LR) classifier, after hyperparameter tuning and data standardization, achieved an accuracy of 97.6%. • This study was assessed through ten-fold cross-validation and accuracy metrics.
[18] (S6)	developing effective skin disease detection system employs a dynamic graph cut algorithm (DGCA) for segmenting skin lesions based on intensity, color, and texture, and utilizes a Naïve Bayes classifier for categorizing the diseases into three groups: melanoma, keratosis, and benign lesions.	• The system was performed using the International Skin Imaging Collaboration (ISIC) 2017 Challenge Dataset. • Data augmentation techniques were utilized, including contrast enhancement, correlation, energy calculation, homogeneity, entropy, color space transformation, and rotation. • The models used for comparison included Fully Convolutional Network (FCN), SegNet, and rule-based diagnostic systems based on the ABCD criteria. • The dynamic graph cut method successfully segmented lesions by adjusting to various color and texture differences. • The performance of the matrix classifier was assessed using sensitivity, specificity, and accuracy, yielding results of 94.3% for benign cases, 91.2% for melanoma, and 92.9% for keratosis.
[16] (S7)	exploring the elements that influence the precision of multiclass skin disease classification through deep learning techniques, utilizing the HAM10000 dataset, which includes basal cell carcinoma, benign keratosis, melanoma, and melanocytic nevi. It assesses the performance of various CNN architectures, such as AlexNet and	• The augmentation methods employed include random scaling, rotation, reflection, and shear transformations, which are assessed in comparison to a control scenario without any augmentation. • Data augmentation techniques do not consistently enhance accuracy and should be applied with caution, as the choice of mini-batch size can affect classification performance. • Models based on ResNet50 demonstrate superior performance compared to those based on AlexNet.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

Source	Aims	Findings
	ResNet50, by analyzing their results using confusion matrices and accuracy metrics for the classification of skin diseases.	• Utilizing transfer learning greatly enhances accuracy in contrast to training a model from the ground up. • Decreasing the depth of the ResNet50 network can maintain or potentially boost accuracy.
[27] (S8)	the new method for automatic skin lesion segmentation is referred to as Morphological Geodesic Active Contour (MGAC), which integrates geodesic active contours with mathematical morphology to achieve effective and stable segmentation. It was applied to the PH² (Pedro Hispano Hospital) dermoscopic image dataset.	• The augmentation techniques implemented include resizing and transforming the color space. • The MGAC approach shows strong performance across all similarity metrics analyzed, such as the Jaccard Index (86.16%), Dice coefficient (92.09%), and Matthews correlation coefficient (87.52%), while also achieving impressive results in sensitivity (91.72%), specificity (97.99%), accuracy (94.59%), and F-measure (93.82%). • This method operates quickly and does not require a training phase, making it suitable for real-time applications. • MGAC demonstrates resilience to noise in dermoscopic images and successfully segments skin lesions, highlighting its potential for assisting in medical diagnoses.
[26] (S9)	to develop a reliable system for identifying skin cancer through deep learning methods, classify seven distinct types of skin lesions, and evaluate the effectiveness of different Convolutional Neural Network (CNN) architectures based on accuracy.	• The dataset utilized dermoscopic images from HAM10000, consisting of 7 categories: melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, and vascular lesions, totaling 7,000 images that were balanced through augmentation. • The augmentation methods implemented included rotation, scaling, reflection, translation, brightness adjustments, horizontal and vertical shifts, and zooming. • The convolutional neural network (CNN) architectures employed in the study were MobileNetV2, ResNet50, and VGG16. • The highest performance was recorded with MobileNetV2 (using PyTorch), achieving an accuracy of 80.79%, with the best class-wise accuracy being 93% for Vascular Lesions, while the lowest class-wise accuracy was 59% for melanoma. • The primary challenges encountered in this research included data imbalance and a limited number of melanoma samples.
[28] (S10)	to analyze psoriasis vulgaris and chronic eczema of the hands and/or feet, a medical smartphone monitoring application was utilized, along with Automated Machine Learning (AutoML) to forecast the progression of symptoms (such as itching, pain, and DLQI). The process of feature extraction comprised therapy modifications, age, BMI, and the level of disease activity.	• The research did not incorporate image-based data analysis. Even though patients uploaded skin images through the app, the image data were omitted from the AutoML evaluation due to significant variability. • The machine learning models were exclusively trained using clinical and questionnaire data (pain and itch scores, DLQI, HADS). • No data augmentation was utilized; only preprocessing steps such as data merging, feature extraction, and feature creation were conducted. • The AutoML models demonstrated strong performance, with LightGBM and Random Forest being the leading models. LightGBM excelled at predicting itching development with a log loss of 0.9167, while Random Forest was used to model pain development with a log loss of 1.0976, and DLQI development was also modeled using Random Forest, yielding a log loss of 1.3974. • Performance was assessed using several metrics, including F1-score, recall, precision, logarithmic (log) loss, area under the curve (AUC), accuracy, and the fraction of variance explained (FVE) for binomial outcomes.
[7] (S11)	develop a deep learning classification framework aimed at detecting monkeypox from skin images. Introduce a novel optimization approach, referred to as BERSFS (Al-Biruni Earth Radius Optimization-Based Stochastic Fractal	• The dataset utilized is from Kaggle (Monkeypox Skin Image Dataset 2022), comprising four categories: Monkeypox (279 images), Chickenpox (107 images), Measles (91 images), and Normal (327 images).

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

Source	Aims	Findings
	Search), to optimize CNN hyperparameters for enhanced classification accuracy. Evaluate the proposed technique against an array of conventional machine learning models, deep learning structures, and other optimization-driven CNN models to assess its efficacy.	• Data augmentation techniques were applied to normalize the dataset, including rotation, flipping, adjustments to brightness, and color jitter/noise. • The proposed BERSFS-CNN model demonstrated enhanced classification performance over basic models (SVM, k-NN, Decision Tree), deep learning architectures (AlexNet, VGG19, GoogLeNet, ResNet-50), and optimization-based models (BER-CNN, SFS-CNN, PSO-CNN, GWO-CNN, and WOA-CNN). • Performance was assessed using metrics such as accuracy, precision, sensitivity, specificity, F1-score, NPV, ROC Curve Analysis, along with ANOVA & Wilcoxon tests to determine statistical significance. • BERSFS-CNN achieved an impressive accuracy of 98.83%, with notable results in sensitivity (85.71%), specificity (99.21%), and F1-score (80.54%).

4. CONCLUSION

In conclusion, this article aims to conduct a systematic review classification of skin disease using machine learning. The main goal of this review was to gain insight into classification of skin disease using machine learning and to find the best approach to implement them. Generally, skin disease classification is conducted using convolutional neural network (CNN) models or traditional machine learning techniques like support vector machines (SVM) and random forests (RF). Most researches have compared CNNs with traditional machine learning methods or analyzed various CNN architectures, such as AlexNet, MobileNet, and Inception V3, based on metrics like accuracy, precision, sensitivity, specificity, and F1-score. The majority of the research focused on skin diseases like melanoma, psoriasis, eczema, acne, and other related disorders. Commonly used augmentation methods include resizing, flipping, random scaling, rotation, reflection, translation, brightness modifications, color space alterations, and shear transformations. The leading countries in research on developing machine learning models for skin diseases are India, China, the United States, Spain, and the United Kingdom.

REFERENCES

[1] AlDera SA, Othman MT Ben. A Model for Classification and Diagnosis of Skin Disease using Machine Learning and Image Processing Techniques. International Journal of Advanced Computer Science and Applications. 2022; 13(5): 252–259. doi:10.14569/IJACSA.2022.0130531

[2] Bajaj L, Kumar H, Hasija Y. Automated System for Prediction of Skin Disease using Image Processing and Machine Learning. International Journal of Computer Applications. 2018; 180(19): 9–12. doi:10.5120/ijca2018916428

[3] Amarathunga AALC, Ellawala EPWC, Abeysekara GN, Amalraj CRJ. 2015. Expert System For Diagnosis Of Skin Diseases. *Int. J. Sci. Technol. Res.* 4(01):174–178. https://ijstr.org/final-print/jan2015/Expert-System-For-Diagnosis-Of-Skin-Diseases.pdf

[4] Chatterjee S, Dey D, Munshi S, Gorai S. 2019. Extraction of features from cross correlation in space and frequency domains for classification of skin lesions. *Biomed. Signal Process. Control*. 53:101581. doi:10.1016/j.bspc.2019.101581

[5] Wei LS, Gan Q, Ji T. 2018. Skin Disease Recognition Method Based on Image Color and Texture Features. *Comput. Math. Methods Med.* 2018. doi:10.1155/2018/8145713

[6] Zaqout I. 2019. Diagnosis of Skin Lesions Based on Dermoscopic Images Using Image Processing Techniques. *Pattern Recognit. - Sel. Methods Appl.*:1–18. https://cdn.intechopen.com/pdfs/68127.pdf

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

[7] Khafaga DS, Ibrahim A, El-Kenawy ESM, Abdelhamid AA, Karim FK, Mirjalili S, Khodadadi N, Lim WH, Eid MM, Ghoneim ME. An Al-Biruni Earth Radius Optimization-Based Deep Convolutional Neural Network for Classifying Monkeypox Disease. Diagnostics. 2022; 12(11). doi:10.3390/diagnostics12112892 [3] Fan J, Kim J, Jung I, Lee Y. A study on multiple factors affecting the accuracy of multiclass skin disease classification. Applied Sciences. 2021; 11(17). doi:10.3390/app11177929

[8] Xiong X, Guo X, Wang Y. Modeling of Human Skin by the Use of Deep Learning. Complexity. 2021. doi:10.1155/2021/5531585

[9] Shetty B, Fernandes R, Rodrigues AP, Chengoden R, Bhattacharya S, Lakshmanna K. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. Scientific Reports. 2022; 12(1): 1–11. doi:10.1038/s41598-022-22644-9

[10] Chaurasia V, Pal S. Machine learning algorithms using binary classification and multi model ensemble techniques for skin diseases prediction. International Journal of Biomedical Engineering and Technology. 2020; 34(1): 57–74. doi:10.1504/IJBET.2020.110361

[11] Aladhadh S, Alsanea M, Aloraini M, Khan T, Habib S, Islam M. 2022. An Effective Skin Cancer Classification Mechanism via Medical Vision Transformer. *Sensors*. 22(11). doi:10.3390/s22114008

[12] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, *et al.* 2021. an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale. *ICLR 2021 - 9th Int. Conf. Learn. Represent.* doi: 10.48550/arXiv.2010.11929

[13] Abbas Q, Daadaa Y, Rashid U, Ibrahim MEA. 2023. Assist-Dermo: A Lightweight Separable Vision Transformer Model for Multiclass Skin Lesion Classification. *Diagnostics*. 13(15). doi:10.3390/diagnostics13152531

[14] Lungu-Stan VC, Cercel DC, Pop F. 2023. SkinDistilViT: Lightweight Vision Transformer for Skin Lesion Classification. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 14254 LNCS:268–280. doi:10.1007/978-3-031-44207-0_23

[15] Tang K, Su J, Chen R, Huang R, Dai M, Li Y. 2024. SkinSwinViT: A Lightweight Transformer-Based Method for Multiclass Skin Lesion Classification with Enhanced Generalization Capabilities. *Appl. Sci.* 14(10). doi:10.3390/app14104005

[16] Fan J, Kim J, Jung I, Lee Y. A study on multiple factors affecting the accuracy of multiclass skin disease classification. Applied Sciences. 2021; 11(17). doi:10.3390/app11177929

[17] Nurdiati S, Najib MK, Bukhari F, Ardhana MR, Rahmah S, Blante TP. 2022. Perbandingan AlexNet dan VGG untuk Pengenalan Ekspresi Wajah pada Dataset Kelas Komputasi Lanjut. *Techno.Com*. 21(3):500–510. doi:10.33633/tc.v21i3.6373

[18] Balaji VR, Suganthi ST, Rajadevi R, Krishna Kumar V, Saravana Balaji B, Pandiyan S. Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier. Measurement: Journal of the International Measurement Confederation. 2020; 163: 107922. doi:10.1016/j.measurement.2020.107922

[19] Zhang X, Wang S, Liu J, Tao C. Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. BMC Medical Informatics and Decision Making. 2018; 18(Suppl 2). doi:10.1186/s12911-018-0631-9

[20] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Journal of Clinical Epidemiology. 2009; 62(10): 1006–1012. doi:10.1016/j.jclinepi.2009.06.005

[21] Albuquerque V, Dias MS, Bacao F. Machine learning approaches to bike-sharing systems: A systematic literature review. ISPRS International Journal of Geo-Information. 2021; 10(2). doi:10.3390/ijgi10020062

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696

[22] Pickering C, Byrne J. The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. Higher Education Research & Development. 2014; 33(3): 534–548. doi:10.1080/07294360.2013.841651
[23] Xiao Y, Watson M. Guidance on Conducting a Systematic Literature Review. Journal of Planning Education and Research. 2019; 39(1): 93–112. doi:10.1177/0739456X17723971
[24] Wee B Van, Banister D. How to Write a Literature Review Paper? Transport Reviews. 2016; 36(2): 278–288. doi:10.1080/01441647.2015.1065456
[25] Al-Emran M, Mezhuyev V, Kamaludin A, Shaalan K. The impact of knowledge management processes on information systems: A systematic review. International Journal of Information Management. 2018; 43(October 2017): 173–187. doi:10.1016/j.ijinfomgt.2018.08.001
[26] Sönmez AF, Çakar S, Selamet F, Kotan M, Delibaşoğlu I, Çit G. Deep Classification Dermoscopic Images for Skin Lesions. Sakarya University Journal of Computer and Information Sciences. 2023; 6(2): 114–122.
[27] Ximenes Vasconcelos FF, Medeiros AG, Peixoto SA, Rebouças Filho PP. Automatic skin lesions segmentation based on a new morphological approach via geodesic active contour. Cognitive Systems Research. 2019; 55: 44–59. doi:10.35377/saucis...1314638
[28] Bibi I, Schaffert D, Blauth M, Lull C, Von Ahnen JA, Gross G, Weigandt WA, Knitza J, Kuhn S, Benecke J, et al. Automated machine learning analysis of patients with chronic skin disease using a medical smartphone app: Retrospective study. Journal of Medical Internet Research. 2023; 25(1). doi:10.2196/50886
[29] Tschandl P, Rosendahl C, Kittler H. 2018. Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data*. 5:1–9.doi:10.1038/sdata.2018.161
[30] Mikołajczyk A, Grochowski M. 2018. Data augmentation for improving deep learning in image classification problem. *2018 Int. Interdiscip. PhD Work.*:117–122.doi:10.1109/IIPHDW.2018.8388338.
[31] Ali SN, Ahmed MT, Paul J, Jahan T, Sani SMS, Noor N, Hasan T. 2022. Monkeypox Skin Lesion Detection Using Deep Learning Models: A Feasibility Study. :2–5. doi:10.48550/arXiv.2207.03342

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i2.12696