



Advanced Classification of Poxvirus-Based Skin Diseases Using Deep Learning Techniques

Kaan Arik^{1*}, Mehmet T. Ağdaş², Adem Korkmaz³, Selahattin Koşunalp³, Teodor Iliev⁴

¹ Department of Computer Technologies, Sakarya Applied Sciences University, Sakarya 54540, Türkiye

² Department of Computer Technologies, Munzur University, Tunceli 62600, Türkiye

³ Department of Computer Technologies, Bandırma Onyedi Eylül University, Bandırma 10200, Türkiye

⁴ Department of Telecommunications, University of Ruse, Ruse 7017, Bulgaria

Corresponding Author Email: kaanarik@subu.edu.tr

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420528>

ABSTRACT

Received: 20 October 2024

Revised: 27 March 2025

Accepted: 30 July 2025

Available online: 31 October 2025

Keywords:

poxvirus classification, YOLOv8, deep learning, skin disease detection, viral infections, medical image analysis, artificial intelligence in healthcare

Viral infections, especially those of the poxvirus family, present significant diagnostic challenges due to their similar clinical symptoms. This study proposes an innovative deep learning-based approach to classify six categories of poxvirus-related skin diseases: chickenpox, cowpox, healthy, measles, monkeypox, and smallpox. A dataset of 9,120 augmented images was used to train, validate, and test three advanced deep-learning models—YOLOv8, YOLOv5, and ResNet32. Among the models, YOLOv8 demonstrated superior performance, achieving an accuracy of 99.80%, precision of 99.28%, and recall of 99.14%, significantly outperforming YOLOv5 and ResNet32. The results underscore the potential of YOLOv8 in medical image analysis, providing a robust and efficient tool for the early detection and accurate classification of viral skin diseases. Comparisons with related studies highlight the effectiveness of the proposed approach, making it a state-of-the-art solution for improving diagnostic accuracy in healthcare. Future work will focus on extending the dataset and evaluating the model's applicability in real-time clinical environments.

1. INTRODUCTION

Viruses are one of the leading causes of human diseases. They are formed in different species and family structures and are known to be responsible for various health problems today [1]. Viral infections are particularly prominent during epidemics. In this context, viral infections pose a risk to the lives of people around the world and harm the economic, social, and cultural structures [2]. Some diseases caused by viral infections—common during the COVID-19 pandemic, for example—are hard to diagnose. For instance, during a pandemic, diagnosing patients with sporadic flu cases has been challenging because the symptoms of these diseases are very similar, and they require intervention through a specific process. After the COVID-19 pandemic, which began in 2019 and caused the deaths of 6.53 million people, there has been panic that another virus might spread worldwide [3].

The same situation can be observed in skin diseases caused by viral infections [4-6]. Monkeypox, one of the viral diseases from the Pox virus family, has emerged as a potential new pandemic by reminding itself of April 2022 as a previously known disease. It has been declared endemic by the World Health Organization, but the disease has not reached the levels of the current pandemic numbers. Over the past months, an alarm has been issued regarding monkeypox, highlighting the need for heightened awareness and preparedness in dealing with this viral threat. Again, some diseases from the past pox virus family (Monkeypox, Chickenpox, Smallpox, and

Cowpox) are rarely encountered in developing countries today [7, 8].

Considering the increasing global concern surrounding poxvirus-related skin diseases and the diagnostic difficulties they pose due to overlapping clinical symptoms, there is a critical need for reliable, efficient, and accessible diagnostic methods. In recent years, deep learning-based image analysis techniques have emerged as powerful tools in the field of medical diagnostics, particularly in dermatology, offering significant potential for the automated detection and classification of skin lesions. Building on this technological advancement, the present study aims to develop a deep learning-based classification model capable of distinguishing between six different skin diseases caused by viruses of the Poxviridae family. By integrating artificial intelligence into the diagnostic process, this research seeks to enhance early detection capabilities, support timely medical intervention, and ultimately contribute to the effective management of potential outbreaks.

1.1 Pox virus family

This study aims to enhance the detection of poxvirus-related skin diseases by utilizing artificial intelligence models to analyze clinical images of six distinct conditions associated with the Poxviridae family. In this context, detailed information is provided regarding the diagnostic features, symptomatic profiles, mortality rates, and epidemiological

data of these six diseases. Furthermore, the common clinical and pathological characteristics shared among these viruses—particularly in terms of symptomatology and transmission routes—are discussed to offer a comprehensive understanding of their manifestation and spread.

Viruses responsible for diseases such as Chickenpox, Cowpox, Measles, Smallpox, and Monkeypox belong to the Poxviridae family and are known to cause dermatological infections characterized by distinct skin lesions. These viruses exhibit genetic similarity, follow similar infection pathways, and pose serious threats to public health due to their potential for rapid transmission and historical involvement in large-scale epidemics. They are also defined by their complex life cycles and unique morphological structures.

Research on poxviruses has significantly contributed not only to the understanding of viral pathogenesis and host–virus interactions but also to the development of effective vaccines and public health intervention strategies. Insights gained from studies in this field have laid the groundwork for innovative approaches in managing viral outbreaks and highlight the ongoing need for advanced research in virology to address emerging infectious diseases.

Among the numerous members of this viral family, there exists one such pathogen that has gained significant attention over the past few years. Monkeypox is a double-stranded enveloped DNA virus of the Orthopoxvirus genus and is predominantly found in West and Central Africa. It is quite similar to illnesses like smallpox and measles and is hard to distinguish [9-11]. Monkeypox is primarily spread from wildlife and rodents to humans via several routes: contact with contaminated animals, consumption of undercooked meat, or via contact with body fluids [12-15]. While human-to-human transmission was limited previously, the spread in Africa, Europe, and America over the past few years has raised its global public health concern significantly. The disease presents a range of symptoms including fever, joint pain, lymph node swellings, and skin blister formations. Though the mortality rate ranges between 3% and 6%, the application of early detection methods like PCR tests and AI-aided image recognition has enhanced diagnostic capability [16, 17]. However, access to adequate healthcare remains a significant challenge, particularly in affected regions of Africa.

While monkeypox continues to be a growing concern, it's crucial to understand its relationship to other poxviruses, particularly its more infamous relative. Smallpox is an acute viral disease caused by the variola virus. Known for causing devastating epidemics throughout history, smallpox spreads through respiratory droplets or direct contact with infected items. The virus has an incubation period of 7-17 days, after which patients typically present with fever, malaise, and characteristic pus-filled blisters [18, 19]. Historically, quarantining patients for six weeks was the primary method of controlling its spread, and survivors often bore permanent facial scarring. A major turning point came with Edward Jenner's pioneering work on the smallpox vaccine in the 18th century, which eventually led to the disease's global eradication [20].

In contrast to smallpox, chickenpox represents a generally milder but still significant viral infection. Caused by the varicella virus, it predominantly affects children and spreads through respiratory droplets. With a longer incubation period of 13-22 days, chickenpox usually presents as a mild illness, though complications can occur in some cases [21]. While the introduction of vaccination programs has significantly reduced

case numbers, infections still occur, albeit typically with milder symptoms [22, 23].

Beyond the poxvirus family, other viral diseases share similar transmission patterns and affect comparable demographics. Measles is another highly contagious viral disease, spread through respiratory droplets. It has a short incubation period, and symptoms include high fever, runny nose, and red eyes. It primarily affects children and can cause outbreaks, especially in unvaccinated populations [23-26].

On the other hand, Cowpox, another infectious disease, caused by a virus from the same family as smallpox, primarily affects cattle. Transmitted primarily through insect vectors such as mosquitoes and ticks, it causes nodular swellings and can lead to severe economic losses in livestock populations. First identified in southern Africa in 1929, the disease remains prevalent in various parts of Africa and can spread efficiently between cattle. While there is no direct treatment for cowpox, preventive measures such as vaccination programs and controlling insect vector populations are key to managing the disease [27-29].

In these viral infections, patients commonly develop rashes similar to those seen in chickenpox, smallpox, and monkeypox, typically appearing on the face and genital areas. While these diseases share many symptoms, making them difficult to distinguish clinically, there are some differentiating factors, including regional variation in symptom presentation. The rashes are generally accompanied by systemic symptoms such as fever, arthralgia, chills, and shivering. These systemic manifestations are characteristic of poxvirus infections and related diseases, though their severity and specific presentation may vary.

In this respect, skin rashes are particularly significant for artificial intelligence-assisted diagnosis, as these visual symptoms can be processed and analyzed to classify different diseases. Advanced models can be developed not only through image processing but also by quantifying various factors such as rash size and other symptoms on a numerical scale. However, significant challenges exist in differentiation due to the remarkable similarity of symptoms among these diseases and the variable frequency and distribution of rashes on the body [30-32]. To address these limitations, ongoing refinements in image processing techniques show promise. Nevertheless, it's important to acknowledge that distinguishing between these diseases remains difficult without specialist expertise. Therefore, our deep learning-supported classification system for smallpox and related diseases is designed not to replace but to assist medical experts in their diagnostic process. By providing an additional analytical tool, this system aims to enhance the accuracy of clinical decisions and contribute to advancements in the field of viral disease diagnosis.

2. MATERIAL AND METHOD

For the development of our classification model, the dataset was partitioned using a 70:10:20 ratio for training, validation, and testing sets, respectively, as detailed in Table 1. Figure 1 provides most samples of diverse pox diseases. The picture preprocessing pipeline blanketed annotation in a folder-primarily based total structure, accompanied by means of data augmentation strategies by each horizontal and vertical transformation with a total of 9,914 images. For data, horizontal and vertical flip processes were used as

augmentation techniques, and although those technics were applied to increase dataset size and improve generalization and robustness, their potential effects on classification prediction were also examined. These augmentation methods may introduce fine-tuning of the model at a level that could affect its ability to discriminate between patterns in predictions.

Table 1. Train, test, and validation sets

	Train	Validation	Test
Chickenpox	1068	171	445
Cowpox	1079	173	465
Healthy	1013	162	426
Measles	1053	137	159
Monkeypox	1049	136	154
Smallpox	1089	141	154
Total	6351	921	1803



Figure 1. Sample images from the augmented images were applied with a 50% probability of horizontal flipped and a 50% probability of vertical flipped

To compare models without any differences, all models were trained under the same conditions, including the same dataset split, learning rate, batch size, and number of epochs. The methodological framework hired for this has a look at applied Python-primarily based totally libraries, in particular Pandas and NumPy, for data manipulation and analysis. The implementation was accomplished at the Google Colab platform, leveraging the PyTorch and TorchVision libraries for our model. On the other hand, we used high-performance virtual machines provided by GoogleColabPro+ for data analysis and modeling. Thus, it provides NVIDIA A100 and up to 32 GB RAM, which provides a suitable environment for large-scale data processing and training deep learning models. Therefore, a small change in performance may be due primarily to differences in model structure rather than degradation in the training set.

2.1 Data preprocessing and collecting

The dataset utilized in this study was carefully curated from two publicly available sources: Google Image Search and Kaggle. These platforms offer a wide range of visual resources; however, they also present inherent challenges, including variations in image quality and labeling accuracy. To address these issues, a meticulous selection and preprocessing process was implemented. All images were evaluated for clarity, relevance, and resolution, with efforts made to anonymize any residual identifiers through cropping and resolution adjustments. No personally identifiable information (PII) was associated with the images. Since the dataset comprised only publicly sourced and anonymized images, formal ethical

approval was not required. Nonetheless, recognizing the sensitive nature of medical imagery, we affirm our strong commitment to maintaining high ethical standards and adhering to data privacy principles in the use and dissemination of clinical image data.

- To tackle these limitations, we established a detailed screening process. Our team came up with a comprehensive selection method that went beyond merely collecting images. Each image was thoroughly assessed based on a few key criteria:
- Visual clarity and sharpness to ensure the image is clear,
- Sufficient resolution so we can conduct detailed analysis,
- Display of clear and important skin-related features to ensure relevance in dermatological studies.

Moreover, we were committed to creating a representative and inclusive dataset. This meant intentionally seeking out images that showcased the rich diversity of skin types and appearances. We carefully selected images captured under various lighting conditions and representing different skin tones and textures. By taking these meticulous steps, we aimed to minimize potential biases and create a robust, representative dataset that could provide meaningful insights into dermatological research.

This dataset focuses on the classification of five distinct diseases: Monkeypox, Measles, Chickenpox, Cowpox, and Smallpox, along with samples representing healthy human skin. The creation of the dataset involved a meticulous selection process to ensure that it accurately reflects the diversity of the targeted diseases. The distribution of the dataset and representative image samples are detailed in Table 1.

Data augmentation, the process of generating new and synthetic data by applying certain operations to existing datasets, is an important technique in machine learning [33]. In machine learning applications, it is inevitable to use large datasets to improve model performance. However, working with limited data can lead to overfitting models. This will negatively affect the performance of the model. As a solution to this problem, data augmentation methods are an alternative. The most common data augmentation method is to create transformed copies of images that belong to the same class as the existing images in the training dataset. Commonly used image augmentation methods include translation, rotation, scaling, zooming, and cropping. One of the main goals is to expand the training set by providing alternative, current, and relevant examples that reflect changes the model may encounter in real-world conditions. As shown in Figure 1, rotating an image horizontally and zooming in from different angles are among the options that can be used in this context. This is because images can be taken from different angles, such as left or right.

Conversely, applying a vertical flip to an image may be inappropriate, as it is unlikely that the model will encounter upside-down representations of the subject in real-world scenarios. This underscores the importance of judiciously selecting data augmentation techniques that are tailored to the specific context of the training dataset and informed by domain knowledge. Furthermore, when working with a limited prototype dataset, it can be beneficial to evaluate data augmentation techniques both individually and in combination to assess their impact on model performance. Such an approach allows for a systematic investigation of which techniques yield measurable improvements in the model's effectiveness [34].

Preprocessing was performed by applying each of the following steps to every image: auto-orienting pixel facts, including EXIF-orientation stripping, resizing to 224×224 pixels and stretching. Further, augmentation was made on each source image to produce three versions, with a 50% chance of horizontal flip and a 50% chance of vertical flip. Other augmentation strategies include rotation and brightness adjustment; these will also be taken into consideration for version robustness.

All pores and skin lesion photos have been shown using Google's Reverse Image Search and then cross-referenced with other sources. The ones that were no longer recognizable, of low resolution, or terrible quality went through a two-step screening process for discarding. Unique photos that met the high standards were picked, then cropped to identification at the area of interest and resized to 224×224 pixels while maintaining the aspect ratio. Normalization techniques were followed afterward to ensure that the pixel values were correctly scaled for training the model with the aim of facilitating improved convergence at the time of training.

The dataset is partitioned into three subsets: training, validation, and test sets. The training set, comprising 6,351 images (70% of the total data), is employed for model training and fitting. The validation set includes 921 images (10% of the data), playing a critical role in fine-tuning the model and mitigating overfitting during the training phase. The test set, containing 1,803 images (20% of the data), is reserved for evaluating the model's performance on unseen data, ensuring robust generalization. During the preprocessing phase, 43 data instances were identified as invalid and excluded from the dataset. This balanced distribution enables efficient training, validation, and testing of the model.

Table 2. Key hyperparameters for deep learning model training

Parameter	Value	Description
task	classify	Defines the task as classification.
mode	train	Indicates the mode of operation.
epochs	100	Number of training epochs.
batch	16	Size of the training batch.
image size	128	Input image size for the model.
workers	8	Number of workers for data loading.
pretrained	true	Utilizes pretrained weights.
optimizer	auto	Automatically selects the optimizer.
lr0	0.01	Initial learning rate.
momentum	0.937	Momentum for the optimizer.
Weight decay	0.0005	L2 regularization factor.
Warmup epochs	3.0	Number of warmup epochs for the learning rate.
Label smoothing	0.0	Smoothing parameter for labels.
nbs	64	Effective batch size.
hsv_h	0.015	Hue augmentation range.
hsv_s	0.7	Saturation augmentation range.
hsv_v	0.4	Value augmentation range.
fliplr	0.5	Probability of horizontal flip.
mosaic	1.0	Mosaic augmentation probability.
cfg	null	Configuration file (if any).

As shown in Table 2, hyperparameters selection was guided by preliminary experiments and empirical optimization. Learning rates were selected based on a grid search strategy to balance convergence speed and stability. Batch size was determined by considering both model performance and GPU memory constraints.

The table above presents the major hyperparameters for

developing this deep-learning model to classify images. Firstly, the --task is a parameter that defines the task; in this context, it was just image classification. Secondly, mode is "train", indicating that the model should be in training mode. Thirdly, the model should train up to 100 epochs, as defined with epochs, while efficient training will be done on a batch size of 16. All images are resized to 128×128 pixels normalizing the input size. Grayscale images are loaded parallel through 8 workers, further speeding up the training. More importantly, the model makes use of pre-trained weights. The optimizer is specified as "auto" which means automatic selection. The initial learning rate, lr0 stands at 0.01.

Additionally, the parameters include a momentum value of 0.937 and a weight decay factor of 0.0005 to regulate training dynamics and prevent overfitting. A warmup phase of 3 epochs is incorporated to stabilize the learning process initially. The inclusion of label smoothing (set to 0.0) and augmentation parameters, such as hue, saturation, and value adjustments, further aims to enhance model robustness and generalization capabilities. The probability of horizontal flipping is set at 0.5, and the mosaic augmentation is employed with a probability of 1.0, contributing to the diversity of training samples [35].

2.2 ResNet32

Residual Networks represent architecture, developed to cope with performance degradation along with the increase in depth in deep learning. As shown in Figure 2, the ResNet32 architecture was made to overcome the problem of gradient loss at each step, encountered during deep network training. It has a 32-layer network, and the "skip connections" [36] are directly available between layers. This architecture enables the training of deep networks and alleviates the problem of gradient loss. It focuses on attention with its high accuracy rates, especially in tasks such as image classification and object recognition. Hayward is another advantage of this model having 32 layers; the remaining connections let the model go deeper. The learning process then will be more stable, and effective, and reach high accuracy in complex image processing. Its success also comes to reinforcing the importance of ResNet architecture due to winning deep learning research.

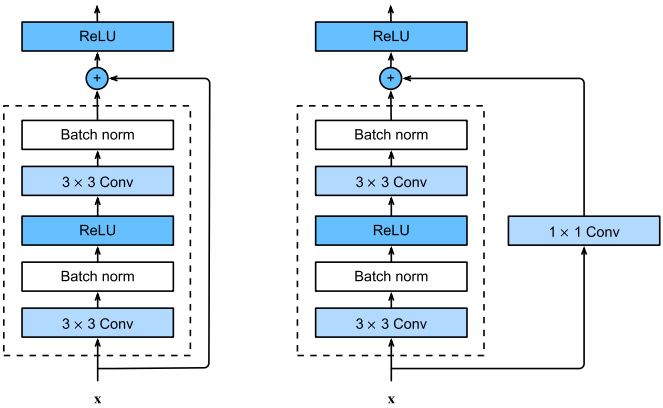


Figure 2. Traditional RESNET32 architecture

2.3 YOLOv5

You Only Look Once (YOLOv5) is a deep-learning architecture designed for real-time object detection. As shown in Figure 3, each forward pass processes objects within images,

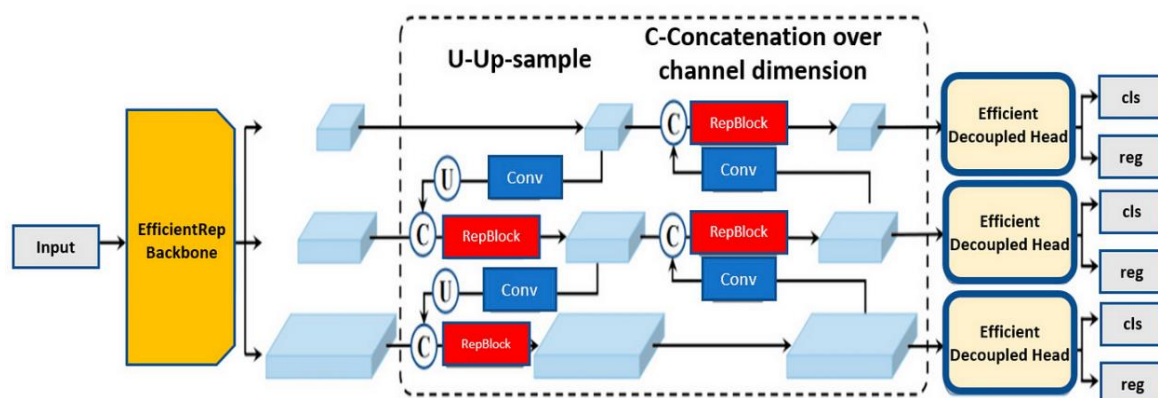
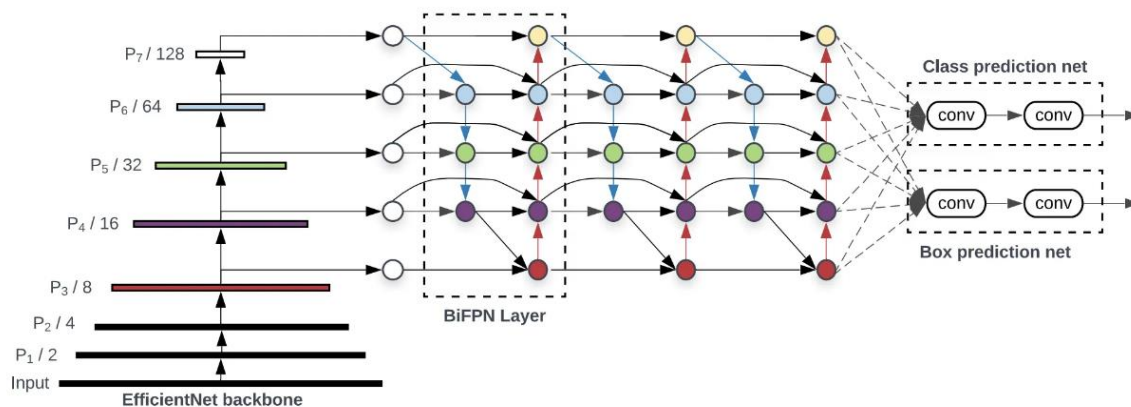
allowing for precise and rapid object detection simultaneously [37]. With the use of convolutional layers along with a certain design of the network, the model is particularly efficient for real-time applications. Because of the multilayered architecture of this model, it is capable of quite effective object detection for different scales [38]. YOLOv5 model is well known among learners performing object detection in complex environments because of its straightforward implementation and great community support. It is, however, more known for precise and accurate object detection as posed by moderate difficulties given by overlapping objects and variations in object size and pose. Due to this quality, it has been used widely across multiple industries such as automation, security, and self-driving cars [39]. Due to this quality, it has been widely used across different industries such as automation, security, and autonomous vehicles. Also, it is user-friendly and has great community support which allows developers to use it for many applications.

The newest member of the YOLO family, YOLOv8, brought several enhancements for object detection, as shown in Figure 4, with important gains in both accuracy and speed compared to the previous ones. Using state-of-the-art layers and modern optimization techniques for better feature extraction, YOLOv8 can detect a wide range of objects very fast [40]. Its upgraded configurations allow it to perform effectively across various applications. According to Sohan et al. [41] this latest version is designed to enhance object detection performance through innovative features. YOLOv8's

2.5 Model performance metrics

The process of choosing the optimal classifier is one of the most important issues in machine/deep learning classification methods model development. It's essential to choose the training dataset for building the model and the test section for testing the model when choosing the best classifier.

True Positive (TP) and True Negative (TN) metrics indicate the number of times the algorithm correctly predicts positive and negative samples, respectively, and are utilized to measure the algorithm's accuracy. The False Positive (FP) value reflects the number of instances where the algorithm incorrectly predicts a negative sample as positive, serving as a measure of the algorithm's precision. Conversely, the False Negative (FN) metric denotes the number of times the algorithm predicts a positive sample as negative, which is used to measure the algorithm's Recall. The confusion matrix is illustrated in Figure 5.



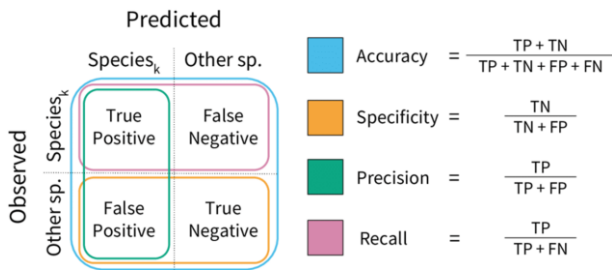


Figure 5. Metrics for classification and confusion matrix

Accuracy is calculated as the ratio of the number of samples correctly classified by the algorithm (TP + TN) to the total number of samples (TP + FP + FN + TN). Confusion matrix graphs were generated on a navy-blue background in the Findings and Results section. In these visualizations, a darker shade of blue indicates a performance value approaching 100%. Conversely, as the color intensity decreases, it can be inferred that the performance also diminishes. For the YOLOv8 model, the confusion matrix was derived from the validation data; however, it is important to note that the results were normalized to a range of 0 to 1, reflecting performance metrics within this interval.

3. RESULTS

The study used a dataset consisting of Monkeypox, Cowpox, Measles, Chickenpox, Smallpox, and Healthy images for classification problems. The dataset consisted of 6351 images in raw form, as shown in Table 1. However, corrupted photos have been eliminated with preprocessing techniques, and the sample of images has been increased to 9120 by filtering image algorithms. During the preprocessing stage, 43 data instances were deemed invalid by the model. This balanced distribution allows for effective model training, validation, and testing. 70% of the dataset was proportioned as training, 10% as validation, and 20% as test data.

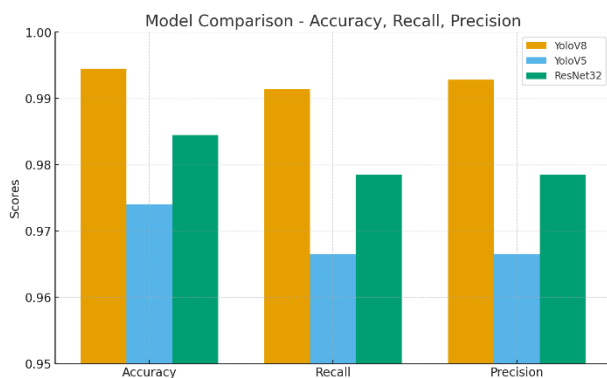


Figure 6. Numerical results of the experimental study

Figure 6 provides a comparative summary of the performance metrics—accuracy, recall, and precision—across the YoloV8, YoloV5, and ResNet32 models. Among the three models, YoloV8 exhibits the highest overall accuracy (99.45%) and precision (99.28%), while also maintaining a strong recall (99.14%). It communicates that YoloV8 has the best balance between true positives and false positives relative to other models. On the other hand, YoloV5 has the poorest performance by all metrics with 97.40%, 96.65%, and 96.65%

as the accuracy, recall, and the precision, respectively. YoloV5 is strong model, yet he is more prone to variance and misclassification of diseases when compared to other models. On one hand, ResNet32 is capable of achieving 98.45% accuracy and 97.85% in recall and precision, which is lower, but still quite competitive. These confirm that YoloV8 still has the best results but ResNet32 performs significantly with active competition versus YoloV8 depending on the use-case scenario.

Figure 7 illustrates the test images used in the experimental results of our model, categorized according to the YoloV8 algorithm, which was identified as the most successful classification model in this study. The images are associated with six categories of viral skin diseases: chickenpox (0), cowpox (1), healthy (2), measles (3), monkeypox (4), and smallpox (5). YoloV8 algorithm was able to correctly classify these features as belonging to their respective categories due to the accurate labeling and separation of unique features associated with each disease. The model accuracy in classifying these test images indicates that the model is able to capture complex visual features, like the typical viral infection rashes. This differentiation capability for diseases with common symptoms indicates the value of YoloV8 for medical image analysis, particularly in the automation of early-stage diagnosis of skin diseases associated with pox viruses.

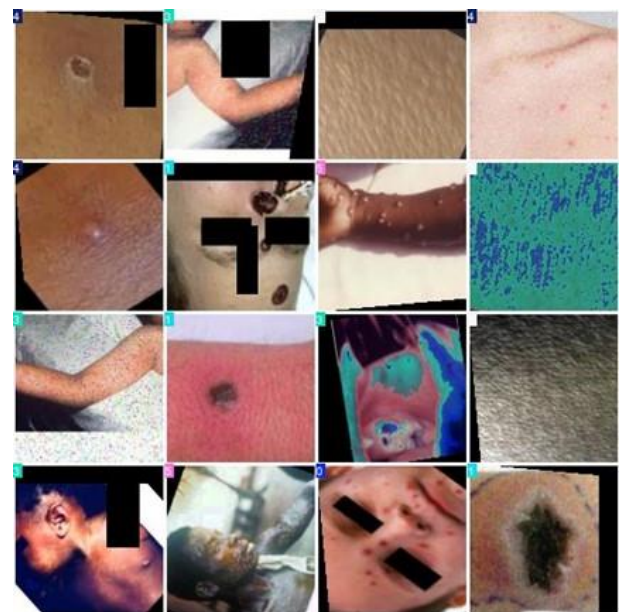


Figure 7. Test Images of experimental results in our model (Chickenpox: 0, Cowpox: 1, Healthy:2, measles:3, monkeypox: 4, smallpox:5)

Figure 8 shows the Confusion Matrix graph obtained as a result of experimental dataset training with YoloV8 (a), YoloV5 (b) and ResNet32 (c) models. According to the graphs, the most successful Chickenpox detection was obtained with the ResNet32 model. Cowpox was detected by YoloV8 model without error. Healthy was detected with the YoloV8 model without error. Measles was detected with the YoloV5 model without error. However, the YOLOv8 model detected Measles with one error, showing that the model was also successful here. YoloV8 and ResNet32 performed similarly in Monkeypox prediction. Unlike YOLOv8, ResNet32 predicted monkeypox with one more correct prediction. In smallpox prediction, the YOLOv8 model predicted perfectly. In the

prediction of this class, the performance of ResNet32 and YOLOv8 models was far behind the prediction of other classes.

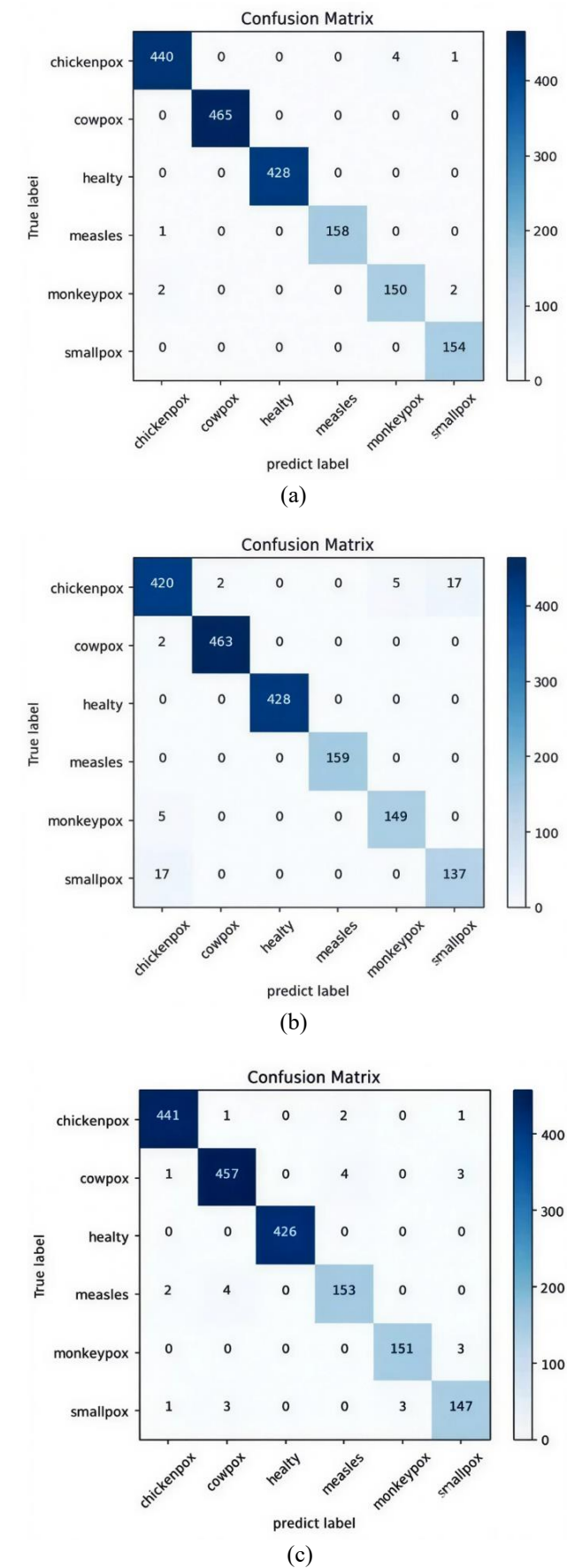


Figure 8. Confusion matrix output of the study's analyses (a- YoloV8, b-YoloV5, c-Resnet32)

The YoloV8 model shows the highest overall accuracy, particularly excelling in the classification of Smallpox, Healthy and Cowpox, where no misclassifications occur. Overall performance remains robust, although minor misclassifications were observed for chickenpox, where a few cases were mistaken for measles and smallpox. There was slight confusion between monkeypox and smallpox, with two cases of monkeypox misclassified as smallpox. Despite these few misclassifications, YoloV8 stands out as the most reliable model for detecting these viral skin diseases. In contrast, the YoloV5 model, while still performing well, shows a higher degree of misclassification, especially between chickenpox and smallpox; 17 cases of chickenpox were misclassified as smallpox. In addition, five cases of monkeypox were misclassified as chickenpox, suggesting some confusion between these classes. However, YoloV5 performs strongly with minimal errors in the cowpox and healthy categories. ResNet32, on the other hand, strikes a balance between YoloV8 and YoloV5 in terms of accuracy. While it shows high precision in classifying chickenpox, only one case was misclassified as chickenpox and another as smallpox. However, ResNet32 has some difficulty distinguishing measles from other diseases and there is some confusion between monkeypox and smallpox, with three cases misclassified in both directions. Overall, YoloV8 emerges as the most suitable model for correct classification, while YoloV5 and ResNet32 show some room for improvement, especially in distinguishing specific disease categories.

The analysis of confusion matrices revealed certain misclassifications, particularly between diseases exhibiting similar dermatological features, such as monkeypox and smallpox. These classification errors can be attributed to overlapping visual symptoms, including pustular rashes and lesion distributions, which often appear nearly identical in two-dimensional clinical images. In some instances, image resolution limitations or inconsistent lighting conditions may also obscure subtle diagnostic cues, leading the model to misclassify. From a clinical standpoint, such errors are significant, as incorrect differentiation between monkeypox and smallpox could delay appropriate public health responses or treatment strategies. Therefore, while the model exhibits high overall accuracy, these findings emphasize the need for incorporating additional clinical metadata, such as lesion progression patterns or patient history, to further reduce ambiguity in future implementations.

Table 3 presents a comparative analysis of the performance metrics from our study alongside related works in the field of viral skin disease classification. Previous studies have explored both binary and multiclass classification problems using various deep-learning architectures. For instance, Nanni et al. [42] achieved 89.42% accuracy using a DenseNet + SVM model for binary classification of cowpox, while Ji and Wu [43] reached 97.7% accuracy for measles using ResNet50 combined with DeepLabv3. Similarly, Ali et al. [44] employed models such as VGG-16, ResNet50, and InceptionV3 for multiclass classification of measles, monkeypox, and chickenpox, with ResNet50 achieving 82.96% accuracy. More recent works, such as Lakshmi and Das [45] and Khan and Ullah [46], focused on the binary classification of monkeypox using advanced architectures like ResNet101 and Inception-Resnet, achieving accuracies of 94.25% and 97%, respectively.

In comparison, our study utilizes a more comprehensive dataset of 9120 images and focuses on the multiclass classification of chickenpox, cowpox, healthy, measles,

monkeypox, and smallpox. The YoloV8 model demonstrated superior performance, achieving an impressive accuracy of 99.80%, significantly outperforming previous models. This highlights the effectiveness of our approach, particularly in handling complex multiclass classification tasks, and positions

YoloV8 as a state-of-the-art model for viral skin disease detection. This comparison underscores the advancements made in this study and the potential of deep learning techniques in improving diagnostic accuracy for pox virus-related skin diseases.

Table 3. Comparison of performance metrics with related works

Class	Authors	Size	Model	Accuracy
CowPox	Nanni et al. [42]	1500	DenseNet +SVM	89.42%
Measles	Ji and Wu [43]	500	ResNet50 +DeepLabv3	97.7%
MonkeyPox	Ali et al. [44]	3196	VGG-16, ResNet50, and InceptionV3	ResNet50: 82.96%
chickenpox				
Measles	Lakshmi and Das [45]	835	VGG16, VGG19, ResNet50, ResNet101, DenseNet201, and AlexNet	ResNet101: 94.25%
Monkeypox				
Monkeypox	Khan and Ullah [46]	558	VGG16, VGG19, ResNet50, Inception and Inception-Resnet	Inception-Resnet: 97%
Chickenpox	Our study	9120	YoloV8, YoloV5 and ResNet32	YoloV8: 99.80%
Cowpox				
Measles				
Monkeypox				
Smallpox				

4. DISCUSSION AND CONCLUSIONS

The findings of this study demonstrate the significant potential of deep learning techniques, particularly the YOLOv8 model, in the accurate classification of viral skin diseases from the poxvirus family. By leveraging a dataset comprising six different categories—chickenpox, cowpox, healthy, measles, monkeypox, and smallpox—the YOLOv8 model achieved an outstanding accuracy of 99.80%, outperforming other models like YOLOv5 and ResNet32. This performance is particularly noteworthy in the context of multiclass classification, a more complex task compared to binary classification explored in previous studies.

The comparative analysis of YOLOv8, YOLOv5, and ResNet32 reveals that YOLOv8 consistently delivers superior results, not only in terms of accuracy but also in precision and recall. This can be attributed to the architectural improvements introduced in YOLOv8, which enhance both detection speed and accuracy through advanced feature extraction and optimized layers. While YOLOv5 and ResNet32 also performed well, the increased misclassification rates between categories like chickenpox and smallpox in YOLOv5, as well as some confusion between monkeypox and smallpox in ResNet32, indicate that these models may require further refinement to match the robust performance of YOLOv8.

In comparing our work with previous studies, the YOLOv8 model's performance represents a significant advancement in the field. For example, Nanni et al. [42] achieved 89.42% accuracy in the binary classification of cowpox using DenseNet + SVM, while Ali et al. [44] reported 82.96% accuracy for multiclass classification using ResNet50. In contrast, our study demonstrates that by utilizing a more comprehensive and augmented dataset, combined with state-of-the-art deep learning architectures, it is possible to achieve substantially higher classification accuracy. This highlights the importance of using larger, more diverse datasets and advanced models like YOLOv8 to improve diagnostic accuracy in medical image analysis.

The application of image augmentation techniques, including horizontal and vertical flips, further enhanced the performance of the models by generating diverse training examples, thereby preventing overfitting and improving the

generalization capability. This was crucial in achieving high classification performance, especially in distinguishing between diseases with similar visual presentations, such as chickenpox and smallpox.

Although hyperparameters such as learning rate and batch size were selected through preliminary tuning and grid search, it is important to acknowledge that these values were not independently optimized for each model architecture. Using a uniform set of hyperparameters across all models ensured fairness in comparison but may have prevented each model from achieving its optimal performance. Future studies should incorporate model-specific hyperparameter optimization strategies to potentially enhance the accuracy and robustness of each deep learning architecture.

Despite the significant improvements observed through data augmentation techniques, including horizontal and vertical flipping, certain limitations must be acknowledged. Some augmentation strategies—such as vertical flipping—may introduce unrealistic visual patterns that are unlikely to occur in real clinical scenarios, potentially impacting the model's generalization capability. Furthermore, while these methods expand the training set and reduce overfitting, they do not contribute to novel pathological features and may fail to reflect the complexity of real-world variability, such as different imaging devices, lighting conditions, or lesion evolution stages. Thus, although data augmentation has proven beneficial in boosting performance, future studies should explore more advanced augmentation methods such as GAN-based synthetic data generation or domain-specific transformations tailored to dermatological contexts to further enhance model robustness and clinical applicability.

Despite the high classification accuracy achieved by the proposed deep learning model, several limitations must be acknowledged, particularly regarding its applicability in real-world clinical settings. Variability in skin tones, lighting conditions, image resolution, and acquisition angles can significantly impact model performance, as these factors are often uncontrollable in practical diagnostic environments. Moreover, the dataset used in this study was derived from publicly available sources, which may not adequately represent the full clinical diversity, including rare or atypical cases. Consequently, the model's generalizability to less

common presentations remains uncertain. To address these challenges, future research should focus on incorporating more diverse, real-world datasets and validating the model within prospective clinical workflows. Additionally, the integration of expert dermatological insights could further enhance diagnostic precision and help mitigate potential biases, reinforcing the model's role as a supportive tool rather than a standalone diagnostic solution.

In conclusion, this study demonstrates that deep learning models, particularly YOLOv8, are highly effective tools for the classification of viral skin diseases from the poxvirus family. The results not only confirm the superiority of YOLOv8 over other models but also underscore the potential of integrating artificial intelligence into diagnostic processes to support medical professionals in early detection and accurate disease classification. Future research could explore further improvements in model architecture and investigate the application of these techniques in real-time clinical settings. Additionally, expanding the dataset to include other types of skin lesions or diseases could further enhance the generalizability and applicability of the models in broader healthcare contexts.

ACKNOWLEDGMENT

This study was supported by the European Union—NextGenerationEU through the National Recovery and Resilience Plan of the Republic of Bulgaria (Grant No.: BG-RRP-2.013-0001).

REFERENCES

- [1] Bhadoria, P., Gupta, G., Agarwal, A. (2021). Viral pandemics in the past two decades: An overview. *Journal of Family Medicine and Primary Care*, 10(8): 2745-2750. https://doi.org/10.4103/jfmpe.jfmpe_2071_20
- [2] Bankar, N.J., Tidake, A.A., Bandre, G.R., Ambad, R., Makade, J.G., Hawale, D.V. (2022). Emerging and re-emerging viral infections: An Indian perspective. *Cureus*, 14(10): 30062. <https://doi.org/10.7759/cureus.30062>
- [3] Mallya, S., Mallya, S. (2022). Emerging and reemerging viral infections in globe with special emphasis in India-A review. *Biomedicine*, 42(6): 1138-1149. <https://doi.org/10.51248/v42i6.2098>
- [4] Park, K.C., Han, W.S. (2002). Viral skin infections: Diagnosis and treatment considerations. *Drugs*, 62(3): 479-490. <https://doi.org/10.2165/00003495-200262030-00005>
- [5] Peeling, R.W., Heymann, D.L., Teo, Y.Y., Garcia, P.J. (2022). Diagnostics for COVID-19: Moving from pandemic response to control. *The Lancet*, 399(10326): 757-768. [https://doi.org/10.1016/S0140-6736\(21\)02346-1](https://doi.org/10.1016/S0140-6736(21)02346-1)
- [6] Younes, N., Al-Sadeq, D.W., AL-Jighefee, H., Younes, S., Al-Jamal, O., Daas, H.I., Yassine, Hadi. M., Nasrallah, G.K. (2020). Challenges in laboratory diagnosis of the novel coronavirus SARS-CoV-2. *Viruses*, 12(6): 582. <https://doi.org/10.3390/v12060582>
- [7] Mitjà, O., Ogoina, D., Titanji, B.K., Galvan, C., Muyembe, J.J., Marks, M., Orkin, C.M. (2023). Monkeypox. *The Lancet*, 401(10370): 60-74. [https://doi.org/10.1016/S0140-6736\(22\)02075-X](https://doi.org/10.1016/S0140-6736(22)02075-X)
- [8] Rasizadeh, R., Shamekh, A., Shiri Aghbash, P., Bannazadeh Baghi, H. (2023). Comparison of human monkeypox, chickenpox and smallpox: A comprehensive review of pathology and dermatological manifestations. *Current Medical Research and Opinion*, 39(5): 751-760. <https://doi.org/10.1080/03007995.2023.2200122>
- [9] Banuet-Martinez, M., Yang, Y., Jafari, B., Kaur, A., Butt, Z.A., Chen, H.H., Yanushkevich, S., Moyles, I.R., Heffernan, J.M., Korosec, C.S. (2023). Monkeypox: A review of epidemiological modelling studies and how modelling has led to mechanistic insight. *Epidemiology and Infection*, 151: e121. <https://doi.org/10.1017/S0950268823000791>
- [10] Letafati, A., Sakhavarz, T. (2023). Monkeypox virus: A review. *Microbial Pathogenesis*, 176: 106027. <https://doi.org/10.1016/j.micpath.2023.106027>
- [11] Shchelkunova, G.A., Shchelkunov, S.N. (2022). Smallpox, monkeypox and other human orthopoxvirus infections. *Viruses*, 15(1): 103. <https://doi.org/10.3390/v15010103>
- [12] Elsheikh, R., Makram, A.M., Vasanthakumaran, T., Tomar, S., Shamim, K., Tran, N.D., Elsheikh, S.S., Van, N.T., Huy, N.T. (2023). Monkeypox: A comprehensive review of a multifaceted virus. *Infectious Medicine*, 2(2): 74-88. <https://doi.org/10.1016/j.imj.2023.04.009>
- [13] Islam, M.M., Dutta, P., Rashid, R., Jaffery, S.S., Islam, A., Farag, E., Zughaier, S.M., Bansal, D., Hassan, M.M. (2023). Pathogenicity and virulence of monkeypox at the human-animal-ecology interface. *Virulence*, 14(1): 2186357. <https://doi.org/10.1080/21505594.2023.2186357>
- [14] Kaler, J., Hussain, A., Flores, G., Kheiri, S., Desrosiers, D. (2022). Monkeypox: A comprehensive review of transmission, pathogenesis, and manifestation. *Cureus*, 14(7): 26531. <https://doi.org/10.7759/cureus.26531>
- [15] Mrudula, A.S.S., Alla, D., Muneesh, S., Sireesha, T., Bhavani, B.D., Vandana, K., Chaturya, G. (2022). Emergence of monkeypox virus: A public health threat. *International Journal of Advances in Medicine*, 9(10): 1078. <https://doi.org/10.18203/2349-3933.ijam20222410>
- [16] Hu, R., Queen, C.M., Zouridakis, G. (2013). Detection of Buruli ulcer disease: Preliminary results with dermoscopic images on smart handheld devices. In 2013 IEEE Point-of-Care Healthcare Technologies (PHT), Bangalore, India, pp. 168-171. <https://doi.org/10.1109/PHT.2013.6461311>
- [17] Yotsu, R.R., Ding, Z., Hamm, J., Blanton, R.E. (2023). Deep learning for AI-based diagnosis of skin-related neglected tropical diseases: A pilot study. *PLOS Neglected Tropical Diseases*, 17(8): e0011230. <https://doi.org/10.1371/journal.pntd.0011230>
- [18] Priyamvada, L., Satheshkumar, P.S. (2021). Variola and monkeypox viruses (Poxviridae). *Encyclopedia of Virology*, 2: 868-874. <https://doi.org/10.1016/B978-0-12-809633-8.21545-8>
- [19] Xiang, Y., White, A. (2022). Monkeypox virus emerges from the shadow of its more infamous cousin: Family biology matters. *Emerging Microbes & Infections*, 11(1): 1768-1777. <https://doi.org/10.1080/22221751.2022.2095309>
- [20] Weidenthaler, H. (2022). Smallpox and Other Orthopoxvirus Diseases. *VacciTUTOR*. <https://id->

- ea.org/wp-content/uploads/2022/07/VacciTUTOR-Chapter-58.pdf.
- [21] Falcón, D.I.M., Báez, D.A.D., Vera, M.D., Gruter, E.L.M.F., Paredes, M.R.N., Campos, F.E.L. (2023). Varicella Zoster Virus (VZV) infection: A comprehensive review of chickenpox. *International Journal of Medical Science and Clinical Research Studies*, 3(10): 2479-2484. <https://doi.org/10.47191/ijmscrs/v3-i10-67>
 - [22] Duncan, D.L. (2019). Chickenpox: Presentation, transmission, complications and prevention. *British Journal of School Nursing*, 14(10): 482-485. <https://doi.org/10.12968/bjsn.2019.14.10.482>
 - [23] Leung, A.K., Hon, K.L., Leong, K.F., Sergi, C.M. (2018). Measles: A disease often forgotten but not gone. *Hong Kong Medical Journal*, 24(5): 512-520. <https://doi.org/10.12809/hkmj187470>
 - [24] Bojarska, M., Domańska, N., Brzyska, A., Bogucka, J., Wilczek, N., Pieciewicz-Szczęsna, H. (2022). Measles: A disease that's making a comeback. *Journal of Education, Health and Sport*, 12(9): 273-287. <https://doi.org/10.12775/JEHS.2022.12.09.032>
 - [25] Fraser, C., Riley, S., Anderson, R.M., Ferguson, N.M. (2004). Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences*, 101(16): 6146-6151. <https://doi.org/10.1073/pnas.0307506101>
 - [26] Pereira, M.R. (2024). Measles outbreak in the United States. *American Journal of Transplantation*, 24(5): 708. <https://doi.org/10.1016/j.ajt.2024.03.014>
 - [27] Baxby, D., Bennett, M., Getty, B. (1994). Human cowpox 1969–93: A review based on 54 cases. *British Journal of Dermatology*, 131(5): 598-607. <https://doi.org/10.1111/j.1365-2133.1994.tb04969.x>
 - [28] Hazel, S.M., Bennett, M., Chantrey, J., Bown, K., Cavanagh, R., Jones, T.R., Baxby, D., Begon, M. (2000). A longitudinal study of an endemic disease in its wildlife reservoir: Cowpox and wild rodents. *Epidemiology and Infection*, 124(3): 551-562. <https://doi.org/10.1017/S0950268899003799>
 - [29] Mauldin, M.R., Antwerpen, M., Emerson, G.L., Li, Y., Zoeller, G., Carroll, D.S., Meyer, H. (2017). Cowpox virus: What's in a Name? *Viruses*, 9(5): 101. <https://doi.org/10.3390/v9050101>
 - [30] Goyal, M., Knackstedt, T., Yan, S., Hassanpour, S. (2020). Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in Biology and Medicine*, 127: 104065. <https://doi.org/10.1016/j.compbiomed.2020.104065>
 - [31] Liu, Y., Jain, A., Eng, C., Way, D.H., Lee, K., Bui, P., Kanada, K., De Oliveira Marinho, G., Gallegos, J., Gabriele, S., Gupta, V., Singh, N., Natarajan, V., Hofmann-Wellenhof, R., Corrado, G.S., Peng, L.H., Webster, D.R., Ai, D., Huang, S.J., Coz, D. (2020). A deep learning system for differential diagnosis of skin diseases. *Nature Medicine*, 26(6): 900-908. <https://doi.org/10.1038/s41591-020-0842-3>
 - [32] Sethi, Y., Nambiar, A.R. (2018). iSkin specialist – an artificial intelligence aided diagnostic support system for dermatology. In *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, pp. 5399-5407. <https://doi.org/10.1109/BigData.2018.8622499>
 - [33] Han, J., Kamber, M., Pei, J. (2012). *Data Mining: Concepts and. Techniques*, Waltham: Morgan Kaufmann Publishers. <https://homes.di.unimi.it/ceselli/IM/2012-13/slides/02-KnowYourData.pdf>.
 - [34] Abdollahi, B., Tomita, N., Hassanpour, S. (2020). Data augmentation in training deep learning models for medical image analysis. In *Deep Learners and Deep Learner Descriptors for Medical Applications*, pp. 167-180. https://doi.org/10.1007/978-3-030-42750-4_6
 - [35] Shaziya, H., Zaheer, R. (2020). Impact of hyperparameters on model development in deep learning. In *Proceedings of International Conference on Computational Intelligence and Data Engineering: ICCIDE 2020*, pp. 57-67. https://doi.org/10.1007/978-981-15-8767-2_5
 - [36] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
 - [37] Saidani, T. (2023). Deep learning approach: YOLOv5-based custom object detection. *Engineering. Technology & Applied Science Research*, 13(6): 12158-12163. <https://doi.org/10.48084/etasr.6397>
 - [38] Lakhotiya, R., Chavan, M., Divate, S., Pande, S. (2023). Image detection and real time object detection. *International Journal for Research in Applied Science and Engineering Technology*, 11(5): 2785-2790. <https://doi.org/10.22214/ijraset.2023.51839>
 - [39] Lavanya, G., Pande, S.D. (2023). Enhancing real-time object detection with YOLO algorithm. *EAI Endorsed Transactions on Internet of Things*, 10: 1-9. <https://doi.org/10.4108/eetiot.4541>
 - [40] Huang, Z., Li, L., Krizek, G.C., Sun, L. (2023). Research on traffic sign detection based on improved YOLOv8. *Journal of Computer and Communications*, 11(7): 226-232. <https://doi.org/10.4236/jcc.2023.117014>
 - [41] Sohan, M., Sai Ram, T., Rami Reddy, C.V. (2024). A review on yolov8 and its advancements. In *International Conference on Data Intelligence and Cognitive Informatics*, pp. 529-545. https://doi.org/10.1007/978-981-99-7962-2_39
 - [42] Nanni, L., De Luca, E., Facin, M.L., Maguolo, G. (2020). Deep learning and handcrafted features for virus image classification. *Journal of Imaging*, 6(12): 143. <https://doi.org/10.3390/jimaging6120143>
 - [43] Ji, M., Wu, Z. (2022). Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic. *Computers and Electronics in Agriculture*, 193: 106718. <https://doi.org/10.1016/j.compag.2022.106718>
 - [44] Ali, S.N., Ahmed, M.T., Paul, J., Jahan, T., Sani, S.M., Noor, N., Hasan, T. (2022). Monkeypox skin lesion detection using deep learning models: A feasibility study. *arXiv preprint arXiv:2207.03342*. <https://doi.org/10.48550/arXiv.2207.03342>
 - [45] Lakshmi, M., Das, R. (2023). Classification of monkeypox images using LIME-enabled investigation of deep convolutional neural network. *Diagnostics*, 13(9): 1639. <https://doi.org/10.3390/diagnostics13091639>
 - [46] Khan, G.Z., Ullah, I. (2023). Efficient technique for monkeypox skin disease classification with clinical data using pre-trained models. *Journal of Innovative Image Processing*, 5(2): 192-213. <https://doi.org/10.36548/jiip.2023.2.009>