PJET, Vol. 4, No. 1, July 2026: 95-110
Poly Journal of Engineering and Technology
journal homepage : https://journals.bdu.edu.et/index.php/PJET/index
ORGINAL ARTICLE

A Deep Learning Based Approach for Detection and Severity Classification of Mpox Disease

Dires Workie Sefineh[1]*, Bantesew Muluye Eneyew[2], Demeke Endalie[1], Afework Abiye Jenber[3], and Tesfa Tegegne[1]

[1] Faculty of Computing, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia

[2] Faculty of Civil and Water Resource Engineering, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia

[3] Faculty of Mechanical and Industrial Engineering, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia

ABSTRACT

Mpox disease has become one of the serious global public health outbreaks globally. It is a major concern due to its rapid transmission, increasing outbreak frequency, and potential to cause severe illness and death. Detecting mpox disease in the early stage before widespread community transmission is a major challenge for clinicians in health care system settings. This study aims to design an early detection and severity level classification of mpox disease model using a deep learning approach. We used a deep Convolutional Neural Network (CNN) for detection and severity classification of mpox disease into different predefined class levels. A total of 3017 customized images were collected from online health sources and Kaggle, then annotated by radiologists at Felegehiwot Comprehensive Specialized Hospital in Bahir Dar, Ethiopia. The dataset was divided into training (70%), validation (15%), and testing (15%) subsets, with data augmentation applied to increase the limited dataset size. Segmentation techniques were applied to improve image quality case. The experimental result shows that the proposed model achieved an accuracy of 95.63%, Precision 95.73%, Recall 95.73%, and F1-Score 95.67% severity classification performance result. This study has great significance by enabling clinically meaningful stage-specific diagnosis, offering doctors a practical tool for early and efficient detection and management of mpox disease. The proposed model is a lightweight, faster, smaller-sized deep learning-based model that uniquely classifies mpox disease severity stages by integrating a threshold segmentation step before feature extraction, improving both accuracy and efficiency compared to existing approaches.

Keywords: Deep learning; Feature extraction; Segmentation; Severity; Mpox

DOI: https://doi.org/10.20372/pjet.v4i1.3056

Corresponding Author: **Dires Workie Sefineh**

Email: Corresponding author: *dires.workie@bdu.edu.et* or *abushuworkie@gmail.com*

1. Introduction

Mpox is a serious zoonotic viral infection discovered in Denmark (1958). The first reported human case of mpox disease was in African continent, in the Democratic Republic of Congo, in a 9-month-old boy [1]. Mpox virus is now a sensitive and significant concern outbreak quickly spread across the continent of Africa [2]. According to the World Health Organization (WHO) on August 26, 2024, report, it is the second viral outbreak potential epidemic disease next to the COVID-19 pandemic [3]. The common symptoms of the mpox viral disease include fever, intense headache, lymphadenopathy, back pain, muscle pain, severe asthenia, and skin rash. The evolution of skin lesions from a mpox rash over several clearly defined stages is one of the most distinguishing clinical presentations. After the rash appears, lesions go through a series of stages: macules, papules, vesicles, pustules, and eventually to scabs [4]. In each stage, there are specific morphological features, which express the development of the disease.

It is usually a self-limiting disease with an incubation period of 3-17 days (about 2 and a half weeks), progressing through several stages [5]. During this time, a person does not have symptoms and may feel fine. In Figure 1, the mpox disease stages are depicted graphically and summarized in Table 1. Hence, the disease is a severe and self-limiting condition; an effective diagnosis with complete knowledge of viral strains and genetic concepts is crucial in health care systems. Detecting mpox disease in the early stage before widespread community transmission is a major challenge for clinicians, particularly in resource-limited regions [6].

Table 1: Description of Mpox Disease Severity Stages

Stages	Severity type	Description (incubation period)
1	Macules	Appear like flat, round pink spots with no bump, lasting for 1to 2 days
2	Papules	Go from a flat, pink spot to a raised bump that typically lasts 1 to 2 days
3	Vesicles	The bumps go from raised to filled with a clear fluid for another 1 to 2 days.
4	Pustules	Progress from a clear fluid-filled bump to an opaque pus bump, about a week
5	Scabs	*Scabs* Over a week or two, the pus bumps will crust and scab over.

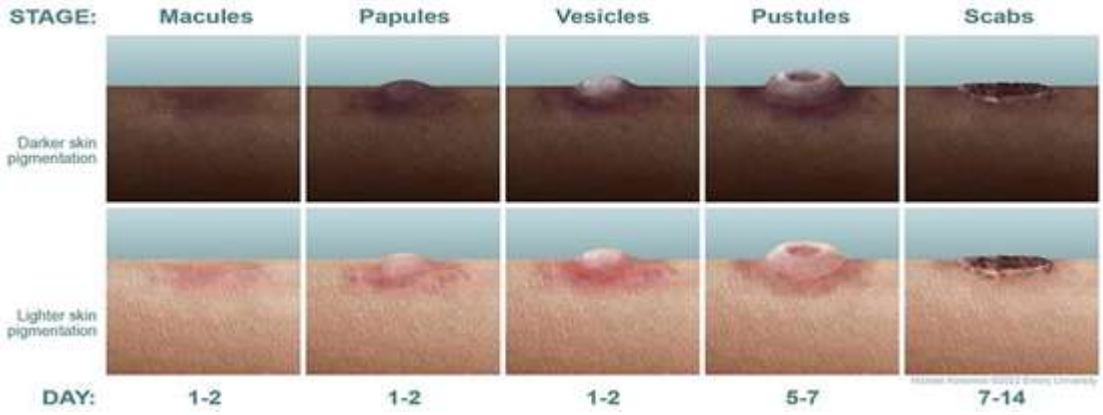

Figure 1: Mpox Disease Severity Stages samples

Recent studies in deep learning and machine learning have a significant role in medical image analysis. Deep convolutional neural networks have achieved remarkable success in skin disease diagnosis, medical image classification, and skin lesion detection by automatically learning discriminative features from medical images [7].

Several studies have investigated deep convolutional neural network-based mpox skin lesion disease detection and classification. However, most previous studies focus on distinguishing mpox disease from other skin lesion diseases, including chickenpox, measles, herpes simplex, and hand-foot-and-mouth disease. In this study, we propose a multi-level severity classification of a mpox disease model using a deep learning approach capable of categorizing mpox disease severities into six class levels (macules, papules, vesicles, pustules, scabs, and normal). This approach aims to improve clinical interpretability by enabling automated assessment of disease progression and supporting more informed decision-making in patient care and management.

The main contribution of this work is the design of a deep learning-based model for the automatic detection and severity classification of the mpox disease stages (macules, papules, vesicles, pustules, scabs, and normal), enabling automated analysis of disease progression beyond binary detection. The proposed approach enhances clinical interpretability and helps in making clinical decisions by quantifying the severity of mpox disease from different stages. Furthermore, the research investigates complex deep learning architectures for precise classification of visually similar lesion phases and discusses the possibility of incorporating explainable AI in prediction to increase transparency.

Related works

The study [8], machine learning-based mpox virus image diagnosis with feature selection and an advanced statistical loss function was used. This used manual or handcrafted-based feature extraction, which leads to time-consuming and is expensive. The main goal of this study was to extract significant features and classify mpox from non-mpox cases using MXGBoost based on a statistical loss function to obtain better classification. The authors used a dataset of measles, chickenpox, and mpox from various kinds of web-scrapping, such as websites, publicly available cases, and news portals. Finally, the suggested framework was trained based on the dataset and achieved 97.41% classification accuracy. Another study [9] proposed a machine learning-based approach to diagnose mpox disease. The authors in this study used a total of 2187 images as an open-source dataset. From this, 1000 were labeled as mpox, and the rest of 1187 labeled as others, i.e., chickenpox and measles. A comparison was conducted with machine learning algorithms, including Support Vector Machine, K-Nearest Neighbor, Random Forests, Logistic Regression, Decision Tree, and Naive Bayes, and three CNN models, including AlexNet, GoogleNet, and VGG16, achieving an accuracy success rate of 95.55%. In the study [10], the authors used a deep learning-based approach to diagnose mpox from skin lesion images. They used a publicly available dataset and tested with pretrained models AlexNet, Places365-GoogLeNet, ResNet-18, GoogLeNet, and SqueezeNet. After applying hyperparameters to choose the best parameter, ResNet18 was able to obtain the highest accuracy of 99.49%. This result proved that the proposed model based on ResNet-18 is a crucial way in classifying mpox virus.

The study in [11] also developed a deep neural network-based mpox disease detection. In this study, the authors used a publicly available dataset formed of skin images of three disease types: mpox, chickenpox, measles, and normal cases, each consisting of 43 Mpox, 47 Chickenpox, 27 Measles, and 54 normal images. Finally, they achieved the best outperformed model of DenseNet201, with an accuracy of 97.63%. Another study in [12] developed a deep learning-based classification model to classify mpox lesions and rash stages. The dataset was a collection of images of skin lesions and various stages of rashes. After data augmentation was applied in the training and testing phases, the

experimental result achieved an accuracy of 90.68% and 90.62% for pox and for the pox rash stage, respectively. Another study [13] investigated a deep learning-based mpox disease detection classification model using feature selection and a fusion approach. Two pretrained models were trained using transfer learning. The authors used a publicly available dataset and achieved an accuracy of 98.59%.

Despite recent advances in deep learning, mpox severity classification remains challenging. To the best of our knowledge, no previous work introduced a multilevel mpox disease severity stages detection and classification by integrating a threshold segmentation step before feature extraction. In this work, we propose a new CNN-based mpox disease detection and severity classification model by applying a threshold segmentation process before the feature extraction phase and optimized filter size selection, resulting in a lightweight, faster, and more accurate model.

2. Materials and Methods

2.1 Proposed system architecture

The proposed system model encompasses three major components: image pre-processing, image segmentation, feature extraction and classification. In pre-processing we normalized the images into a size of the standard level in which Keras can understand and fit closely. In segmentation, we used a threshold segmentation algorithm to choose the ROI of the actual mpox disease image. In feature extraction, we used CNN to detect and extract the relevant texture features that reveal the actual image. Finaly, we used a CNN based SoftMax classifier to classify a mpox disease into six severity levels (Macules, Papules, Vesicles, Pustules, Scabs and Normal). The overall proposed model architecture is illustrated in Figure 2

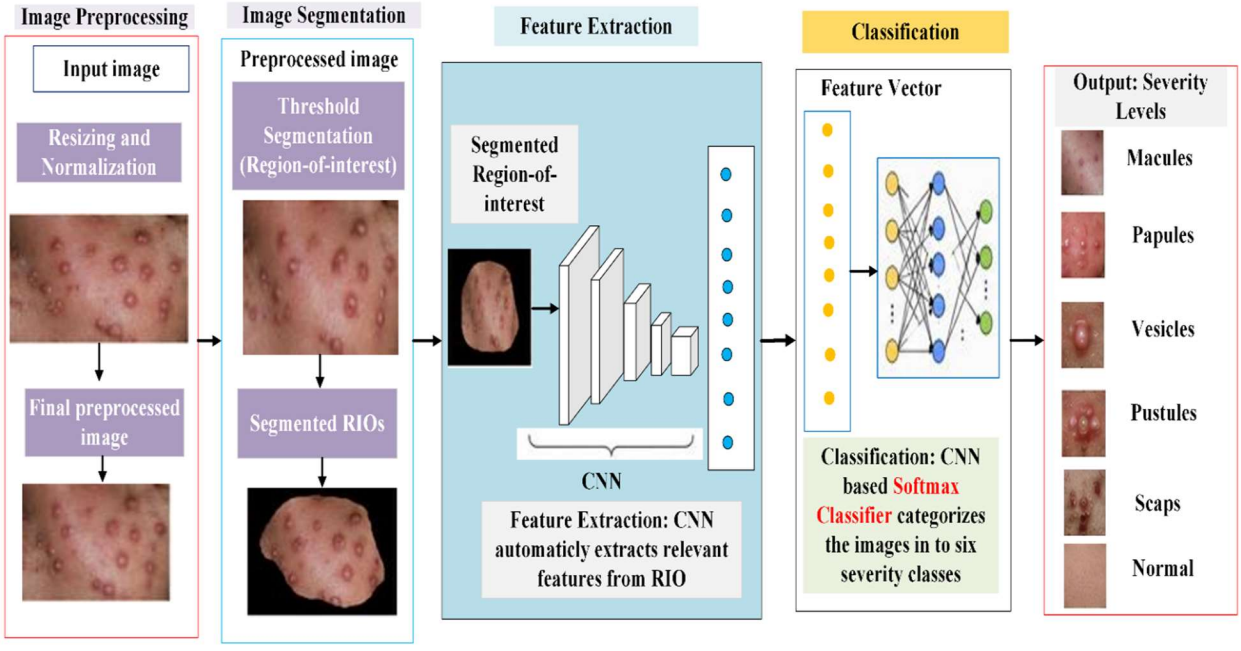

Figure 2: Proposed Model Architecture for Mpox Detection and Severity Classification

2.2 Pre-Processing

The image might be subjected to various background noises (*noise, illumination, and shadow)*. As a result, the necessary information might not be visible. Pre-processing is essential for enhancing image quality since it reduces extraneous noise and boosts contrast without sacrificing information. The median filtering and histogram equalization has been applied to the mpox image. For Keras to function properly, the images are resized into standard size (224 × 224) pixels, like ultramodern models [14, 15] and then translated into NumPy arrays. In this case OpenCV library is used to resize images and converting them into NumPy array so that Keras can work smoothly. Another technique applied in this study is data augmentation algorithm. This algorithm is used for dataset limitations, regularization, and improved generalization. To do this, we have randomly transformed it using techniques including translation, rotation, shearing, and resizing.

2.3 Segmentation

We applied a global threshold segmentation step prior to CNN feature extraction to explicitly isolate lesion region of interest (ROI) from surrounding healthy skin and background image (hair, shadows, textured skin). Standard preprocessing (resizing, histogram equalization, denoising) improves overall image quality but does not remove non-lesion features that can confuse the network [16].Threshold segmentation reduces irrelevant background information and forces the network to learn discriminative lesion texture and morphology, which (a) improves feature signal-to-noise, (b) reduces overfitting to background images, and (c) shortens convergence time. Empirically, integrating segmentation produced more stable training and improved classification performance in our experiment. The image was first converted to grayscale, then blurred to reduce noise. A mask based on the threshold value of 125 was then made after preliminary experiments evaluating several threshold values (100–150), and it was chosen using the binary mask. Finally, the segmented image was fed into the CNN algorithm. The algorithm below in Table 2 shows a straightforward iterative algorithm for selecting the threshold value (T) in an image.

Table 2: Image ROI segmentation algorithm

Input image	
1.	Select a first estimate for the global threshold T.
2.	Segment the image using T. This will produce two groups of pixels: G1. consisting of all pixels with intensity values greater than T and G2, consisting of pixels with values less than or equal to T.
3.	Compute the average intensity values m1 and m2 for the pixels in regions G1 and G2, respectively.
4.	Compute a new threshold value: T=1/2(m1+ m2).
5.	Repeat steps 2 through 4 until the difference in T in successive iterations is smaller than a predefined value, ΔT.
6.	Segment the image using function g(x,y) = 1 if f(x,y) > T and g(x,y) = 0 if f(x,y) ≤ T.

The segmented image is then calculated as follows in equation (1) below.

$$g\ (x,y)\ = 1\ if\ f\ (x,y) > Txy\ g\ (x,y)\ = 0\ if\ f\ (x,y)\ < Txy \quad (1)$$

2.4 Convolutional Neural Network (CNN)

CNN is a deep learning model architecture that includes three distinct types of processing learning layers: the convolution layer, pooling layer, and the full connection layer [17].Different from traditional machine learning, CNN can be better applied on image and time series data processing, especially for image classification and language recognition. This network encompasses different learning layers throughout the model as illustrated in Figure 3.

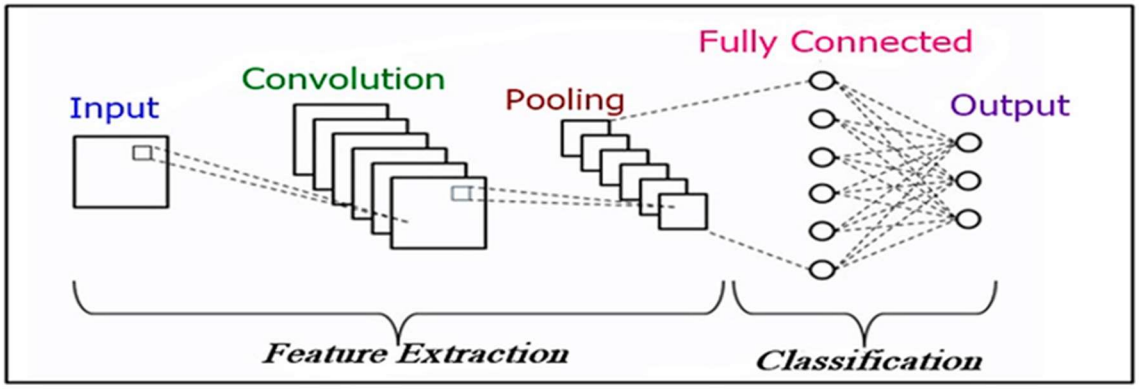

Figure 3: Standard Cconvolutional Nneural Nnetwork Aarchitecture

The deep neural network convolution operation can be defined mathematically in equation 2 below.

$$S(i,j) = \Sigma(m,n)\, K(m,n) * I(i-m+1, j-n+1) \quad (2)$$

2.5 Classification

Convolutional neural network was used for classification. This classification's process is composed of three distinct phases: training phase, validation phase, and testing phases. The operations we applied were convolution, activation, pooling, dropout, flatten, SoftMax classification, learning rate, and batch normalization. At last, after the distinctive texture features were extracted and learned, the SoftMax layer classified the mpox disease into six class severity levels (Macules, Papules, Vesicles, Pustules, Scabs and Normal).

2.5.1 Training phase

Convolutional layer: several convolutional layers are used to detect and extract image features during the training phase. The input size of the image for the first convolution layer is (224,224,3). Different parameters, such as the number of filters, filter size, stride size, and zero paddings, are needed for this convolution layer. The number filter is used to control the depth of the output volume. Following a series of tests with different numbers of filters and convolutions, in this model authors used those with high accuracy (16, 32, 64, and 128) filters. The filter size (3,3) was used in our model to decide the size of the filter or kernel. The stride size (2,2) was used in our model to figure out the number of pixels skipped horizontally and vertically. The last parameter applied in our model was, the "same" padding operation, to decide or control the size of the input and output equality. Equationally, the cross-correlation convolutional neural operation can be defined as follows in equation (3) below.

$$s(i,j) = (K * I)(i,\mathrm{j}) = \sum_m \sum_n I(i+m, j+n) K(m,n) \quad (3)$$

Pooling layer: the pooling operation is used to reduce or down sampling the size of the input volume and remove redundant spatial information**.** We have used max-pooling a (2,2) size operation to reduce the width and the height of the input volume by half each time convoluted.

Fully connected layer: This layer combines all the features learned by the earlier layers across the image to find the larger patterns. We applied one fully connected layer to compute the extracted feature. This is the previously convoluted and pooling layer of high-level features of an image. Therefore, vectors gained in the last fully connected layer are equal to the number of features extracted from the image. This layer is done before the SoftMax classifier that classifies to the specific classes.

SoftMax*:* In this layer the output of the fully connected layer features is normalized and classified into a specific severity class levels (Macules, Papules, Vesicles, Pustules, Scabs and Normal).

$$SoftMax(x)_i = \frac{e^z i}{\sum_{j=1}^{n} e^z j} \tag{4}$$

2.5.2 Validation Phase

Activation function: in the activation layer, the rectified linear unit (ReLU) activation function was used throughout our model. This function sets the zero threshold, mathematically it can be defined as follows in equation (5) below.

$$ReLU = Y = max\,(0, x) \tag{5}$$

If x > 0 — the volume of the array of pixels stays the same and

If x < 0 — it cuts off unnecessary details in the channel.

Batch normalization: we applied the batch normalization activation to normalize the activation of the given inputs in every neural batch of training phase. This method significantly reduces the number of training epochs and is used to lessen internal covariant shift issues, which has a significant impact on the learning process. However, the benefits of the added time exceed the drawbacks. Additionally, it reduces the volatility of tuning learning rate and regularization. If we consider *B* to be denote our mini-batch activations of size *m* of the entire training set, the empirical mean and variance of *B* then can be computed as follows in equation (6).

$$\mu B = \frac{1}{m}\sum_{i=1}^{m} x_i \;\; and \;\; \delta^2 B = \frac{1}{m}\sum_{i=1}^{m}\big((x_i - \mu B)\big)2 \tag{6}$$

For each neural layer with d dimensional input, $x= (x^{(1)}, \ldots \ldots, x^{(d)},) xi$ can be normalized via the following equation (7).

$$x^k i = \frac{x^{(k)} i - \mu^{(k)} B}{\sqrt{\left(\delta^{(k)} B\right)^2 + \epsilon}} \tag{7}$$

where $k \in [1, d]$ and $i \in [1, m]$; $\mu^{(k)}B$ and $\delta^{(k)}B$ are the per- dimension mean and standard deviation, respectively.

Dropout: to reduce or remove the occurrence of overfitting (*When the model performs well during training but performs poorly during testing*), we applied a dropout probability of ***0.5*** in each layer of the model.

2.5.3 Testing phase

Unlike the training and validation phases we used a different procedure to test our model. After the model is learned with the training and validation phase via the dataset and the algorithm, then, it is tested with previously unseen data. The testing samples are different from the training dataset. We used 15% of the entire dataset to test the trained model's performance.

2.6 Severity classification

The mpox disease detection and severity classification is now done by using the actual high-level learned model developed in the training and validation phases. Upon the learning model, the SoftMax layer classified the mpox disease into six predefined severity class levels (Macules, Papules, Vesicles, Pustules, Scabs and Normal).

3. Results and Discussion

In this section, the simulation result of the proposed model is described in detail. The datasets used, the different phases (training phase, validation phase, and testing phase) are described in detail in this subsection bellow. The overall simulation procedure performed are: loading image, building model architecture, determining the training parameters, training the model, validating the model, and evaluating the trained model with unseen data.

3.1 Dataset Acquisition

To detect and classify the mpox disease severity stages the authors grasp relevant mpox disease images from online available Kaggle mpox skin lesion datasets [18] and internet-based online health sources. After collecting relevant datasets, the image preprocessing (*noise, illumination, and shadow*) data augmentation and segmentation (t*o get a particular ROI image feature*) technique were applied to improve the quality of the image. 3017 mpox dataset samples making up 433 Macules, 594 Papules, 426 Vesicles, 456 Pustules, 506 Scabs and 602 Normal were collected. These images are then proportionally divided into 70%, 15%, and 15% for training, validation, and testing of our model, respectively. The overall dataset description is described in Table 3 below.

Table 3: Mpox Disease Dataset Description

No.	Mpox severity stages	Image format	Source	Total dataset
1.	Macules	JPEG	Kaggle & online data	433
2.	Papules	JPEG	Kaggle & online data	594
3.	Vesicles	JPEG	Kaggle & online data	426
4.	Pustules	JPEG	Kaggle & online data	456
5.	Scabs	JPEG	Kaggle & online data	506
6.	Normal	JPEG	Kaggle & online data	602
7.	Total	**3017**		

3.2 Implementation Setup

The Experiments were carried out on a local Intel® Core i7 PC with 16 GB of RAM and Windows 10, supplemented by cloud resources via Google Colab with GPU support. Using TensorFlow and Keras, the model was trained with optimum parameters and tested in real time, providing consistency, scalability, and quick mpox severity identification. Authors first prove the difference in the model's accuracy during the training and validation phases before comparing it to the most advanced (state-of-the-art) models. we also show the performance of the model while applying segmentation. Additionally, we display the model's performance when segmentation is applied. Training phase:

Convolution, activation function pooling, and fully connected layers come in order throughout the training phase before SoftMax classifier. However, in the validation phase: the model was trained with dropout (at the first stages) and batch normalization to keep it from overfitting. In this phase additionally, data augmentation technique is applied. The model was trained with epochs of 100 iterations, 32 batch size and with a start of 0.01 learning rate. The adaptive moment estimation (Adam) optimizer was used, since most deep-learning models applied this optimizer than SGD and others [19]. The model was optimized with different learning rates (0.01, 0.001, 0.0001). When a 0.1 learning rate applied overfitting occurred and underfitting occurred while a small learning rate is applied. After applying those learning rates, the 0.001 learning rate achieves a better result. Better results have been obtained after training and confirmed the model with a momentum value of 0.9. and 0.99. Thus, momentum accelerates the network training and keeps well from underfitting and overfitting.

The model trained and confirmed with different 20, 50,100, 150, and 200 epochs. The model shows overfitting when trained over 150, 200 epochs, meaning that it performs admirably on training data but has high error rates on test data. However, our model shows poor fitting on the training data and high error rates on both the train and test data when we train with 50 and lower epochs. The authors finally determined the improved results at epoch of 100 and managed both overfitting and underfitting. The authors trained the model and applied a dropout value of between 0.1 and 0.5 rates in each layer in the network (a recommended value and most used by many researchers [20, 21]). A dropout unit value of 0.25 achieved better results by controlling underfitting and overfitting.

3.3 Training Performance Analysis

The training dynamics of the proposed deep learning model for mpox severity classification demonstrate robust learning behavior with excellent convergence properties. As illustrated in Figure 4 presents the combined training progress, showing accuracy and loss curves over 100 epochs. Training and Validation Accuracy: The model achieved rapid initial learning, with training accuracy rising from approximately 45% to over 85% within the first 30 epochs. The validation accuracy closely tracked the training accuracy throughout the training process, indicating minimal overfitting. The best validation accuracy of 95.63% was achieved at epoch 65, after which the model maintained stable performance. The narrow gap between training and validation accuracy (approximately 2-3%) suggests that the model generalizes well to unseen data, which is critical for clinical deployment where variability in image quality and lesion presentation is expected.

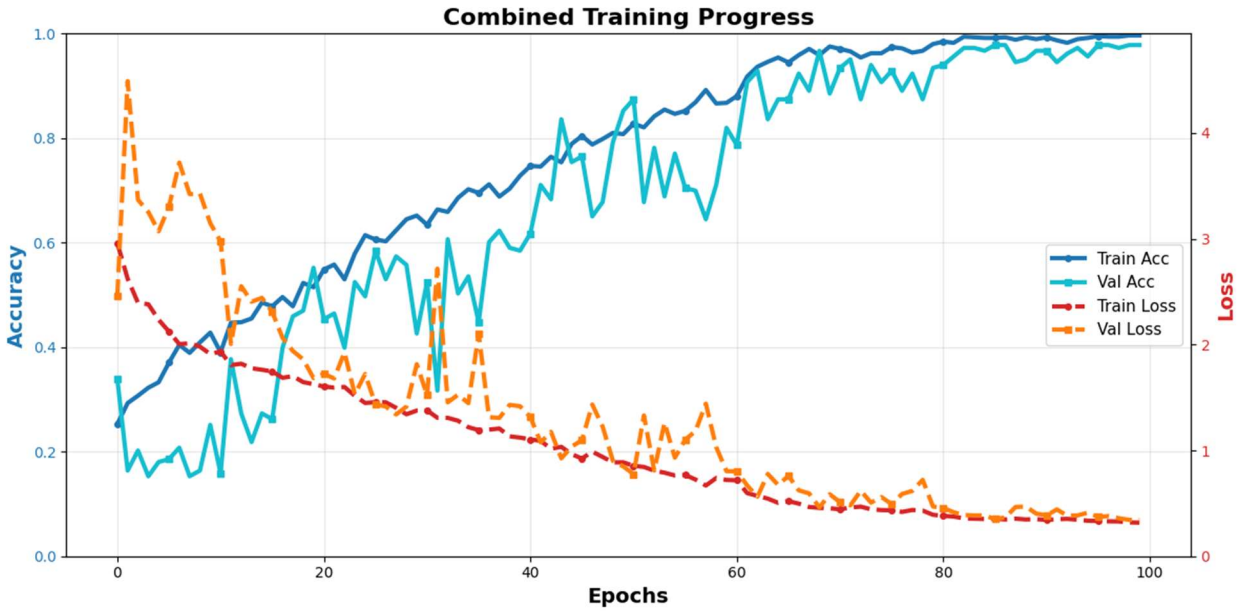

Figure 4: Combined Training Accuracy and Loss Performance

The training and validation loss curves in Figure 4 demonstrate consistent convergence behavior, with both training and validation loss decreasing sharply during the initial 30 epochs before plateauing. The validation loss reached its minimum value of 0.124 at epoch 68, corresponding closely with the peak validation accuracy. The absence of a widening gap between training and validation loss confirms that the model is not overfitting, despite the relatively complex architecture with approximately 12.5 million parameters. This can be attributed to the effective use of regularization techniques including L2 regularization, dropout layers (0.25-0.5), and batch normalization.

3.4 Overall Model Performance

The proposed model achieved outstanding performance across all evaluation metrics, substantially exceeding the baseline performance reported in previous studies, described in Table 6. Table 4 summarizes the comprehensive evaluation results:

Table 4: Overall model performance

Model	Evaluation Metrics			
	Accuracy	Precision	Recall	F1-score
Proposed model	95.63%	95.73%	95.73%	95.67%

The proposed model achieved outstanding performance across all evaluation metrics, with 95.63% accuracy, 95.73% precision, 95.73% recall, and 95.67% F1-score. The proposed model achieved a remarkable performance in classifying all six mpox disease severity stages, demonstrating its exceptional capability for reliable stage-specific diagnosis and revealing the model's potential as a reliable clinical decision-support tool for mpox severity classification. This improvement is attributed to threshold segmentation preprocessing, which effectively isolates lesion regions and reduces background noise. Precision and Recall: The balanced scores (95.73% each) indicate optimal trade-off between false positives and false negatives. High precision avoids unnecessary treatment and patient anxiety, while high recall prevents missed severe cases and delayed intervention. F1-Score: The 95.67% F1-score confirms robust

and balanced performance across all classes, maintaining high reliability even in transitional stages or suboptimal image quality.

3.5 Class-wise Performance

The per-class performance analysis provides critical insights into the model's diagnostic capabilities across the six mpox severity stages. As illustrated in Figure 5 and summarized in Table 5, the model demonstrates consistently high performance across all severity classes, with notable variations that reflect the clinical characteristics and visual distinctiveness of each stage.

Table 5: Per-Class Performance Summary

Severity Classes	Performance Metrics		
	Precision	Recall	F1-Score
Macules	87.5%	93.3%	90.3%
Papules	93.3%	87.5%	90.3%
Vesicles	96.8%	100%	98.4%
Pustules	100%	100%	100%
Scabs	96.8%	100%	98.4%
Normal	100%	95.6%	96.7%

The per-class performance analysis reveals that the model achieves consistently high classification accuracy across all six severity stages, with performance variations reflecting the clinical distinctiveness of each stage. Early-stage classes (macules and papules) demonstrated slightly lower but clinically acceptable performance (F1-scores: 90.3%), with high recall for macules (93.3%) ensuring early detection, and high precision for papules (93.3%) minimizing unnecessary interventions. The model excelled in intermediate and advanced stages, achieving perfect performance for pustules (100% across all metrics) and near-perfect performance for vesicles and scabs (98.4% F1-score), which is clinically critical as these stages represent peak infectivity and require urgent clinical decisions.

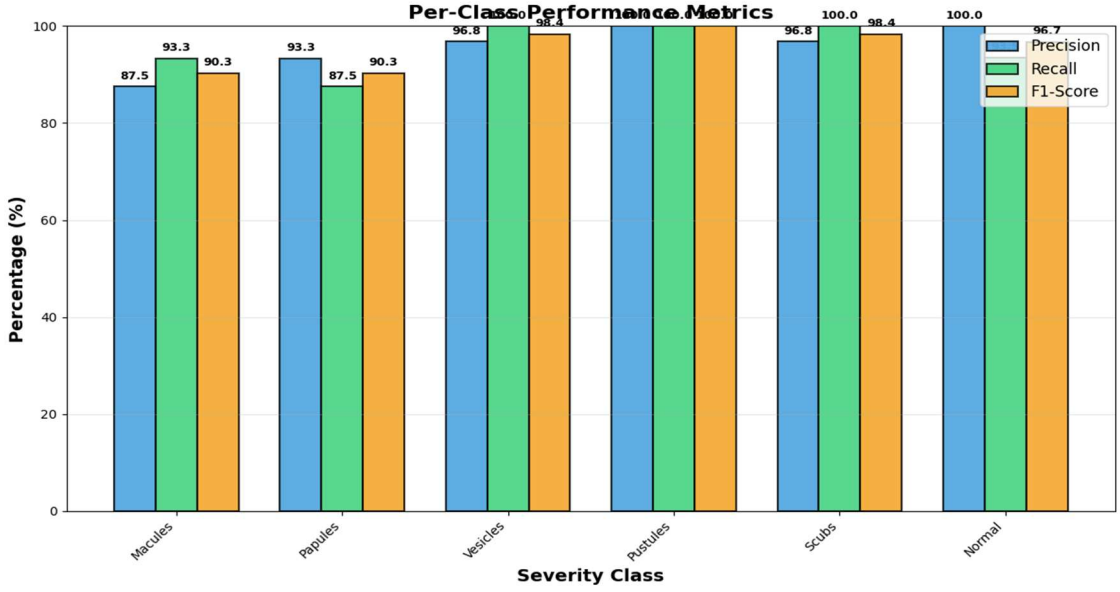

Figure 5: Per-Class Performance Metrics of the Proposed* Mpox *Severity Classification Model

The normal class achieved perfect precision (100%) with 95.6% recall, ensuring healthy individuals are not falsely diagnosed while maintaining high sensitivity. The overall pattern demonstrates that the model's performance improves with disease progression, which is clinically advantageous as accurate identification becomes increasingly critical for patient management and outbreak control.

3.6 Confusion Matrix Analysis

The confusion matrix presented in Figure 6 provides comprehensive insights into the model's classification performance across all six mpox severity classes (Macules, Papules, Vesicles, Pustules, Scabs, and Normal). The matrix reveals the model's exceptional discriminative capabilities, with minimal misclassifications concentrated primarily between adjacent severity stages.

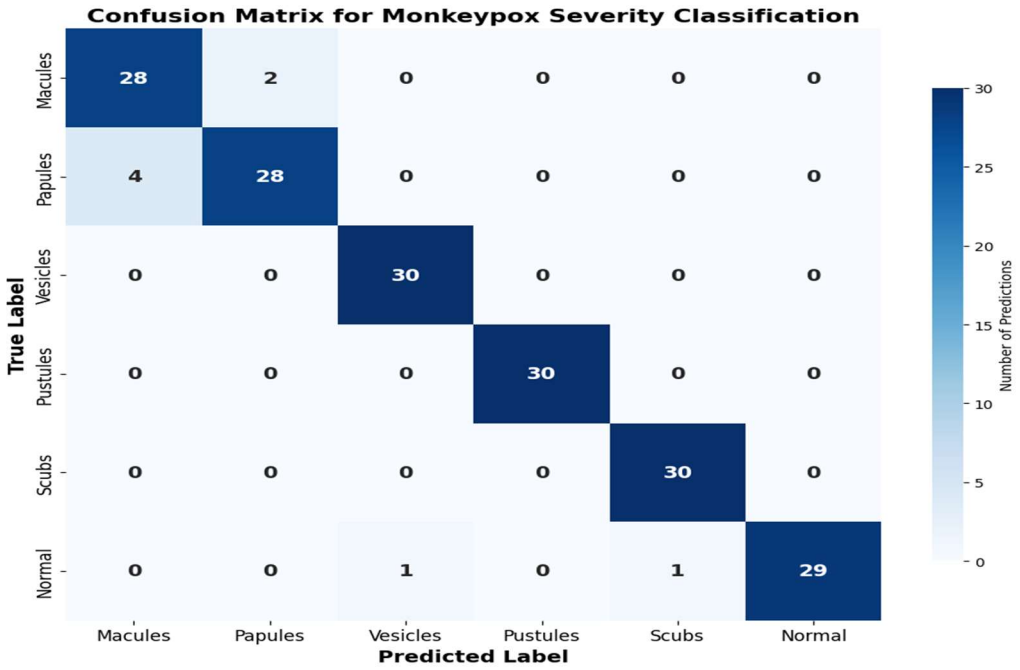

Figure 6: Confusion Matrix of Mpox Severity Classification

The confusion matrix demonstrates exceptional classification performance, with diagonal values ranging from 28-30 out of 30 samples per class (93.3-100% accuracy). Misclassifications were minimal (≤6.7% per class) and exclusively occurred between adjacent severity stages: macules and papules (2 errors each), vesicles and pustules (1 error), and scabs and normal (1 error each). Critically, no misclassifications occurred between distant stages, confirming the model's robust understanding of the hierarchical lesion progression. The perfect classification of vesicles (100%), a highly infectious stage requiring immediate isolation, and near-perfect performance for pustules (96.7%) and scabs (96.7%), demonstrates the model's clinical reliability for critical decision-making. The confusion matrix provides strong validation of the model's quantitative metrics and supports its clinical utility as a trustworthy diagnostic tool for mpox severity classification.

3.7 Sample Predictions

Figure 7 presents representative predictions from the test set, illustrating the model's classification performance across the six severity classes. The visual inspection of these sample predictions provides qualitative validation of the model's quantitative metrics and offers insights into its real-world clinical applicability.

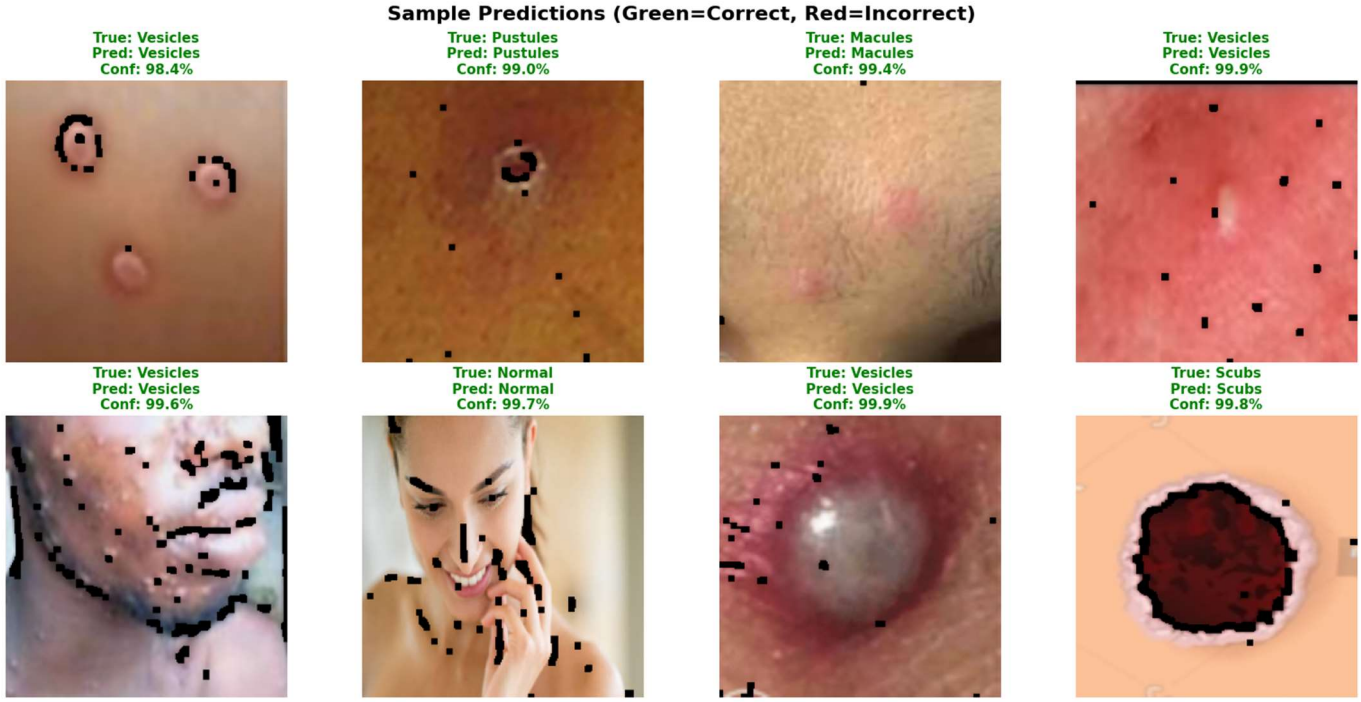

Figure 7: Sample Predictions on Test Set

All eight displayed samples were correctly identified, including early-stage macules (99.4% confidence), intermediate vesicles (up to 99.9%), advanced pustules (99.0%), healing scabs (99.8%), and normal skin (99.7%). The proposed model perfect prediction rate and consistently high confidence scores provide strong qualitative validation of the quantitative metrics (95.63% accuracy), particularly confirming the model's reliability for clinically critical stages requiring urgent intervention. These findings demonstrate the model's exceptional discriminative capabilities and support its clinical utility as a trustworthy decision-support tool for mpox severity classification.

3.8 Comparison with State-of-the-Art Architectures

Since there are currently no widely established state-of-the-art models specifically designed for multi-class mpox severity classification, we compare it with widely used deep convolutional neural network architectures, including ResNet50, EfficientNet-B3, DenseNet121, GoogLeNet, and MobileNetV2. To ensure fairness in the evaluation, each baseline model was carefully tuned to its optimal configuration through hyperparameter optimization. A systematic grid search was conducted to explore alternative parameter values and identify those that yielded the most accurate classification. All models were trained and evaluated on the same six-class severity classification task under identical experimental conditions and with our dataset.

Table 6: Performance Comparison with sate of the art Models

Baseline Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Size (mb)	Time (min)
GoogLeNet	94.89	94.05	93.78	93.67	99.01	84.20	3.28
MobileNetV2	93.97	94.76	94.12	93.77	98.45	9.27	2.06
DenseNet121	94.32	93.48	94.33	94.19	98.27	27.37	4.14
VGG16	90.16	91.48	90.06	89.53	99.35	56.76	6.31
ResNet50	40.98	42.79	40.67	39.90	79.22	91.01	3.31
EfficientNet-B3	27.86	18.21	27.90	19.27	72.28	41.91	3.96
Proposed Model	**95.63%**	**95.73%**	**95.73%**	**95.67%**	**99.76**	**1.64**	**1.05**

The comparison results described in Table 1 show significant performance differences between the baseline models and the proposed model. Among all the evaluated baseline models, the proposed model consistently achieved the highest performance across all metrics, demonstrating its superior capability for mpox disease severity classification. The proposed model achieved outstanding performance with 95.63% accuracy, 95.73% precision, 95.73% recall, 95.67% F1-score, and 99.76% ROC-AUC. These results demonstrate the model's exceptional discriminative ability in distinguishing between the six severity stages. The near-perfect ROC-AUC of 99.76% indicates that the model has excellent class separation capability, with minimal overlap between severity stages. Notably, our model achieved these results with the most lightweight architecture (1.64 MB) and the shortest training time (1.05 minutes), making it highly suitable for deployment in resource-constrained clinical settings.

4. Conclusion

This study introduced a deep learning-based framework for detecting mpox disease and classifying its severity stages across six class levels. In this study, we have developed a high-performance model using a CNN algorithm that can grade or classify the mpox disease stages (Macules, Papules, Vesicles, Pustules, Scabs, and Normal) into a predefined severity class level. The model was trained with a total of 3017 Mpox image dataset samples. The experimental result shows that the proposed model achieved an accuracy of 95.63%, Precision 95.73%, Recall 95.73%, and F1-Score 95.67% severity classification performance result. This shows that the proposed mpox disease severity classification model can be used to classify the stages of the disease more accurately than prior models. The class-wise analysis confirmed perfect classification for pustules (100%) and near-perfect performance for vesicles and scabs (98.4% F1-score), with minimal misclassifications (≤6.7%) exclusively between adjacent severity stages. Comparative evaluation against seven established architectures (GoogLeNet, ResNet50, EfficientNet-B3, DenseNet121, GoogLeNet, MobileNetV2, VGG16) validated the proposed study superior performance, while its lightweight design (1.64 MB, 1.05 MS inference) ensures practical clinical deploy ability. This framework offers a reliable tool for early detection, stage-specific treatment diagnosis, and effective outbreak management, with significant potential to support healthcare centers in improving patient outcomes during mpox outbreaks.

For the future, we planned to extend the model's performance by further analyzing different parameters (filter size, number of filters, learning rate, number of learning rate, etc.) and different CNN algorithms, as well as increasing the dataset from various diverse sources. We further planned to extend this study by integrating Explainable Artificial

Intelligence (XAI) techniques into the proposed model to improve transparency, interpretability, and clinical trust in health care systems.

Data Availability

The datasets and source code supporting this study are publicly available at https://github.com/Dires-colab-pro/Mpox-disease-stages-dataset.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

The authors received no funding for this work.

Ethics and Consent to Participate Declarations

Not applicable.

Acknowledgment

The authors gratefully acknowledge the Faculty of Computing at Bahir Dar Institute of Technology for their support throughout this study. We extend our appreciation to the radiologists at Felegehiwot Comprehensive Specialized Hospital in Bahir Dar, Ethiopia for their invaluable assistance in annotating of the dataset. We also recognize the contributions of open-source platforms, including Kaggle and other health data repositories, which provided essential resources for this research.

References

1. I. D. Ladnyj, P. Ziegler, and E. Kima, "A human infection caused by mpox virus in Basankusu Territory, Democratic Republic of the Congo," *Bulletin of the World Health Organization*, vol. 46, no. 5, pp. 593–597, 1972, doi: 10.2471/BLT.72.46.5.593
2. H. F. Alhasson, E. Almozainy, M. Alharbi, N. Almansour, S. S. Alharbi, and R. U. Khan, "A Deep Learning-Based mobile application for mpox detection," *Applied Sciences*, vol. 13, no. 23, p. 12589, Nov. 2023, doi: 10.3390/app132312589.
3. S. Khattak *et al.*, "The mpox diagnosis, treatments and prevention: A review," *Frontiers in Cellular and Infection Microbiology*, vol. 12, p. 1088471, Feb. 2023, doi: 10.3389/fcimb.2022.1088471.
4. Y. Ortiz-Martínez and A. F. Henao-Martínez, "Time-related change in mpox (mpox) skin lesions and their progression," in *Elsevier eBooks*, Elsevier, 2025, pp. 269–290. doi: 10.1016/b978-0-443-22123-1.00016-8.
5. S. Srivastava *et al.*, "The Global Mpox (MPOx) Outbreak: A Comprehensive review," *Vaccines*, vol. 11, no. 6, p. 1093, Jun. 2023, doi: 10.3390/vaccines11061093.
6. A. J. Porzucek *et al.*, "Development of an accessible and scalable quantitative polymerase chain reaction assay for mpox virus detection," *The Journal of Infectious Diseases*, vol. 227, no. 9, pp. 1084–1087, Oct. 2022, doi: 10.1093/infdis/jiac414.
7. A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv (Cornell University)*, Oct. 2020, doi: 10.48550/arxiv.2010.11929.

8. A. Gessain, E. Nakoune, and Y. Yazdanpanah, "Mpox," *New England Journal of Medicine*, vol. 387, no. 19, pp. 1783–1793, Oct. 2022, doi: 10.1056/nejmra2208860.
9. S. Rampogu, "A review on the use of machine learning techniques in mpox disease prediction," *Science in One Health*, vol. 2, p. 100040, Jan. 2023, doi: 10.1016/j.soh.2023.100040.
10. A. K. Gairola and V. Kumar, "Mpox Disease Diagnosis using Machine Learning Approach," *2022 8th International Conference on Signal Processing and Communication (ICSC)*, pp. 423–427, Dec. 2022, doi: 10.1109/icsc56524.2022.10009135.
11. T. Nayak *et al.*, "Deep learning based detection of mpox virus using skin lesion images," *Medicine in Novel Technology and Devices*, vol. 18, p. 100243, Jun. 2023, doi: 10.1016/j.medntd.2023.100243.
12. A. S. Azar, A. Naemi, S. B. Rikan, J. B. Mohasefi, H. Pirnejad, and U. K. Wiil, "Mpox detection using deep neural networks," *BMC Infectious Diseases*, vol. 23, no. 1, p. 438, Jun. 2023, doi: 10.1186/s12879-023-08408-4.
13. O. Chunhapran, M. Maliyeam, and G. Quirchmayr, "Mpox lesion and rash stage classification using deep learning technique," in *Lecture notes in networks and systems*, 2024, pp. 141–149. doi: 10.1007/978-3-031-58561-6_14.
14. S. Maqsood and R. Damaševičius, *Mpox Detection and Classification Using Deep Learning Based Features Selection and Fusion Approach*. IEEE (Institute of Electrical and Electronics Engineers), 2023, pp. 1–8. doi: 10.1109/syscon53073.2023.10131067.
15. H. Talebi and P. Milanfar, "Learning to resize images for computer vision tasks," *arXiv.org*, Mar. 17, 2021. https://arxiv.org/abs/2103.09950
16. C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, Jul. 2019, doi: 10.1186/s40537-019-0197-0.
17. F. M. Mustafa, "Image Enhancement based on the Histogram Equalization and Multiresolution Discrete Stationary Wavelet Transforms," *Academic Journal of Nawroz University*, vol. 11, no. 2, pp. 50–59, Apr. 2022, doi: 10.25007/ajnu.v11n2a1323.
18. X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artificial Intelligence Review*, vol. 57, no. 4, Mar. 2024, doi: 10.1007/s10462-024-10721-6.
19. S. N. Ali *et al.*, "Mpox skin lesion detection Using Deep Learning Models: A Feasibility study," *arXiv.org*, Jul. 06, 2022. https://arxiv.org/abs/2207.03342
20. E. Hassan, M. Y. Shams, N. A. Hikal, and S. Elmougy, "The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16591–16633, Sep. 2022, doi: 10.1007/s11042-022-13820-0.
21. I. Salehin and D.-K. Kang, "A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain," *Electronics*, vol. 12, no. 14, p. 3106, Jul. 2023, doi: 10.3390/electronics12143106