



OPEN

Deep learning model for hair artifact removal and Mpox skin lesion analysis and detection

Edeh Michael Onyema^{1,2}, B. Gunapriya³, Balasubramanian Prabhu Kavin⁴, Mueen Uddin⁵, Priyan Malarvizhi Kumar⁶, Tehseen Mazhar^{7,9}✉ & Mamoon M. Saeed⁸✉

Accurate identification of Mpox is essential for timely diagnosis and treatment. However, traditional image-based diagnostic methods often struggle with challenges such as body hair obscuring skin lesions and complicating accurate assessment. To address this, the study introduces a novel deep learning-based approach to enhance Mpox detection by integrating a hair removal process with an upgraded U-Net model. The research developed the “Mpox Skin Lesion Dataset (MSLD)” by combining images of skin lesions from Mpox, chickenpox, and measles. The proposed methodology includes a pre-processing step to effectively remove hair from dermoscopic images, improving the visibility of skin lesions. This is followed by applying an enhanced U-Net architecture, optimized for efficient feature extraction and segmentation, to detect and classify Mpox lesions accurately. Experimental evaluations indicate that the proposed approach significantly improves the accuracy of Mpox detection, surpassing the performance of existing models. The achieved accuracy, recall, and F1 scores for Mpox detection were 90%, 89%, and 86%, respectively. The proposed method offers a valuable tool for assisting physicians and healthcare practitioners in the early diagnosis of Mpox, contributing to improved clinical outcomes and better management of the disease.

Keywords Deep learning, Hair removal, Mpox disease, Mpox diagnosis and therapy, U-Net model

Mpox is fast becoming a global concern due to the speed of its spread and symptoms¹. This virus shares evolutionary commonalities with smallpox and cowpox. Rats and monkeys are significant vectors for the illness². The virus was first discovered in monkeys in 1958³ at a lab in Copenhagen, Denmark. In 1970, while attempts to eliminate smallpox were ramping up, reports of Mpox emerged in the Democratic Congo⁴. Infections can be transferred by respiratory droplets, saliva, or nasal secretions⁵. Transmission via animal bites is also possible. Patients with Mpox encounter a wide range of pains, weariness, and red lumps on the skin over time⁶. Despite growing reported cases, Mpox has not caused disruptions like COVID-19⁷. The disease outbreak has been more devastating in developing countries, especially in Africa⁸, and the ailment currently has no cure⁹. However, immunization approaches are being used to mitigate the associated risks of Mpox¹⁰. The use of technology can offer assistance in curtailing or managing the disease. The use of Artificial intelligence (AI) algorithms has the potential to improve the treatment of chronic diseases. This study explored the possibility of a Deep learning algorithm in diagnosing or detecting Mpox.

The similarity of Mpox with other related diseases like chicken pox makes it more challenging to detect. However, lymph node swelling often distinguishes it from other poxes^{11,12}. To diagnose Mpox, medical personnel will take a tissue sample from a patient with an active illness in one part of their body. The material is subsequently transferred to a lab where polymerase chain reaction (PCR) is used to analyze it¹³. This examination is notoriously time-consuming and costly. The current state of Mpox treatment is hopeless due to

¹Department of Science Education, Faculty of Education, Alex Ekwueme Federal University, Ndufu-Alike, Abakaliki, Nigeria. ²Research Institute of IoT Cybersecurity, Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung City, Taiwan. ³New Horizon College of Engineering, Ring Road, Bellandur Post, Bengaluru 560103, India. ⁴Department of Data Science and Business Systems, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu 603203, India. ⁵College of Computing and Information Technology, University of Doha for Science and Technology, Doha, Qatar. ⁶Department of Information Science, University of North Texas, Denton, TX 76205, USA. ⁷Department of Computer Science and Information Technology, School Education Department, Government of Punjab, Layyah 31200, Pakistan. ⁸Department of Communications and Electronics Engineering, Faculty of Engineering, University of Modern Sciences (UMS), Sana'a 00967, Yemen. ⁹School of Computer Science, National College of Business Administration and Economics, Lahore 54000, Pakistan. ✉email: tehseenmazhar719@gmail.com; dr.mamoon@ums-edu.com

the lack of an effective antiviral medication^{14,15}. Researchers must devise a reliable method of diagnosing Mpox before proceeding with data collection and a clinical trial. Technology advancements offer more effective ways to improve monkey pox care¹⁶. Clinical adoption of Machine learning as a valuable decision-making tool has increased the ability of doctors to provide safe, accurate, and quick imaging solutions¹⁷. For instance, a study by¹⁸ created CAD that showed promise in detecting breast cancer quickly. The potential of emerging technologies could also be leveraged in managing the threat of Mpox¹⁹.

This study utilized a deep learning approach. Deep Learning is a branch of AI that helps develop models by automatically extracting features, training them without human input, and generating output. Medical practitioners are already using imaging techniques of many types to aid in the diagnosis of a wide range of disorders, from brain cancer to other respiratory ailments like pneumonia TB, and even COVID. Recently, image interpretation has been the subject of substantial research²⁰. There are many Mpox cases but not enough diagnostic tools. Thus, this is an area that needs attention. The need for more highly trained medical professionals has made it challenging to staff every hospital with a specialist. The deep learning model may also help with other problems, such as a shortage of RT-PCR kits, unreliable test findings, expensive prices, and lengthy wait periods. Researchers have looked into deep learning strategies to discover if machine learning may help build an effective triage method for diagnosing Mpox disease²¹.

The recent global outbreaks of Mpox have highlighted the importance of rapid, accurate diagnostic methods, particularly those based on dermatological examinations. Mpox is often characterized by discrete skin lesions, making visual diagnosis the preferred method; however, overlapping symptoms with other skin disorders might complicate clinical identification. The traditional manual Hair Removal and Skin Lesion Analysis in Mpox Detection is labor-intensive and prone to subjective mistakes, highlighting the importance of automated, dependable techniques. In this area, deep learning algorithms provide promising answers by allowing for efficient lesion identification and categorization. Nonetheless, skin lesion analysis faces major hurdles, including differences in lesion appearance, the existence of occlusions such as body hair, and uneven picture quality. These difficulties may mask important traits required for an appropriate diagnosis. Also, Despite substantial advances in medical image processing, hair removal, and lesion segmentation in dermoscopic images continue to pose significant hurdles. Traditional image processing algorithms frequently struggle with the complexity and diversity of skin pictures, and many contemporary deep learning approaches either ignore hair artifacts or fail to obtain exact lesion boundaries. This gap underlines the critical need for more advanced, robust deep learning models that can manage both hair removal and accurate lesion segmentation, boosting diagnostic reliability and practical application. Addressing these problems, this work introduces a unique deep learning strategy that uses an upgraded U-Net model created specifically for effective hair removal and robust skin lesion segmentation, with the goal of improving the precision and reliability of Mpox identification from dermatological images.

Problem statement

The superior learning potential of especially variants of Networks (CNNs) has revolutionized several arenas of medical research in recent years. These deep networks can analyze pictures in layers, automatically identify prominent characteristics, and learn to determine the ideal task illustrations by training them with vast data. The usefulness is constrained, however²², by the essential for massive volumes of data and lengthy training using specialized computer resources. Since impartial, uniform medical data might be hard to come by, employing accelerators (e.g., GPU, TPU) to speed up the process is a partial solution. Adding new samples to an existing dataset by making small adjustments to the current data is known as data augmentation. When there is not enough information, researchers often turn to transfer learning²³. This technique uses a (CNN) model with a dataset and uses that model's prior knowledge to learn about a new context using a much smaller dataset²⁴. Challenges arise when questions of confidentiality and reliability are factored in. In addition, relative variability may persuade a bias in the dataset due to the high frequency of Mpox in under-developed African countries. This study attempted to leverage the power of algorithms to improve Mpox detection and also support previous efforts²⁵ toward harnessing the potential of technology to enhance the timely detection of diseases and interventions.

The objective of the research work

This study attempted to detect Mpox using photographs from a recently created collection aptly titled Mpox Skin Photographs collection. The key contributions of the study are:

- Create a novel picture collection that can train an algorithm to identify Mpox patients from photographs or patient's skin reliably.
- Development of an improved U-Net to classify Mpox from fused features.
- Early detection of Mpox using deep learning technique.
- Presents a model to help health experts to improve treatment outcomes.

Organization of the research work

This paper is organized as follows: The history of the illness and its detection methods are presented in Section "Introduction". In Section "Related works", we deliberate relevant works of current models that highlight the gaps in our knowledge. Section "Related terminologies" provides a comprehensive overview of the available pre-trained models, while Section "Materials and methods" explains and illustrates the suggested model. Section "Experimentation and results" depicts the validation investigation of the proposed technique using current models and discusses the consequences. In Section "Conclusion and future scope", the consequences are obtainable.

Related works

Sitaula et al.²⁶ conducted a study to test Deep learning models for identifying the Mpox virus. Their model showed a precision score of 85.44%, a recall score of 85.47%, an F1 score of 85.40%, and an accuracy score of 87.13%. The study proved that machine learning models have the potential for the management of Mpox.

Jaradat et al.⁹ investigated and compared the accuracy of two pre-trained models in recognizing pox. These models were MobileNetV2, and EfficientNetB3. To assess the usefulness of the models, several metrics, such as accuracy, recall, precision, and F1-score, were utilized. During our tests, the MobileNetV2 model had the best levels of accuracy (98.16%), recall (0.96), and precision (0.99), as well as the highest F1-score (0.98). During the validation process using a variety of datasets, they achieved the highest level of accuracy, scoring 0.94%. Regarding the classification of mpox photos, our research indicates that the MobileNetV2 method is superior to the performance of other models presented in the scholarly literature. These findings suggest that machine learning technologies may be effective in the early detection of pox. Because our system worked well on both the training set and the test set for pox classification, it is probably a helpful tool for efficient clinical diagnosis. This is because pox did well on both sets.

The dataset produced by Ahsan et al.²⁷ is now obtainable for transfer from our shared repository on GitHub. This provides a more secure manner of accessing and spreading such data to create and deploy machine learning models since it allows the dataset to be created using photographs gathered from various open-source and internet portals that do not levy any limitations on use, even for profitable reasons. Across the course of two independent research, we propose and evaluate an updated version of the VGG16 model. Our initial computational results allow us to suggest a model that has the potential to accurately identify Mpox patients with an accuracy of 971.8% (area under the curve=97.2) and 880.8% (area under the curve=0.867), respectively, for Studies 1 and 2. We use a technique called Explanations, or LIME for short, to explain our model's prediction and feature extraction to shed more light on the peculiarities of the Mpox virus's emergence.

Haque et al.²⁸ have linked deep-based procedures with a convolutional block attention area of the feature maps to the behavior of an image-based classification illness. This was done to do an image-based classification of the Mpox sickness. VGG19, Xception, and DenseNet121 are the names of the five deep-learning models that we created and evaluated as part of this work. A construction consisting of Xception-CBAM-Dense layers has a validation accuracy of 83.89% compared to other models to diagnose human Mpox and other infections.

Alrusaini²⁹ has implemented two different convolutional neural network (CNN) models, GoogLeNet and ResNet50, to extract information from the photographs. The photographs were analyzed using the principal component method to identify characteristics, and classification system models such as InceptionV3, ResNet50, VGG-16, SqueezeNet, and Support Vector Machines were utilized, among others. The results showed that both strategies successfully achieved the desired results. Despite this, the VGG-16 model yielded the best results (accuracy=0.96, F1-score=0.92). The findings of this study add to the growing body of data suggesting that AI may be useful in diagnosing Mpox. If authorization is given by the relevant, it can facilitate the quicker and more convenient diagnosis of illnesses. When it is developed into a smartphone app, regular people may use it to make preliminary diagnoses of themselves before going to a medical professional or a hospital. The study's author proposes broadening the scope of already available picture datasets and undertaking more research into the models to provide more reliable statistical evaluations.

Deep learning models pre-trained, including VGG-19, VGG-16, and EfficientNet-B0, were utilized to determine whether cowpox, smallpox, or chickenpox cases included Mpox. MobileNet V2 achieved the highest accuracy in terms of classification, reaching 99.25%, according to the results of experimental studies conducted on both the primary and supplementary datasets. On the other hand, the VGG-19 model was successful in accurately classifying 78.82% of the raw data. Since it considered these results, the shallow model performed better when smaller image datasets were used. As more data was added, the deep learning models' weights were fine-tuned to provide the best possible results, which led to an increase in the effectiveness of deep neural networks.

Altun et al.³⁰ came up with the idea for the hybrid function learning transfer learning model with changeable hyperparameters. The MobileNetV3-s model, the EfficientNetV2 model, the ResNET50 model, the Vgg19 model, the DenseNet121 model, and the Xception model have all been modified to utilize this approach and in our research, evaluation, and comparison criteria included the (AUC), accuracy, recall, and loss scores, as well as the F1 score. The optimized hybrid MobileNetV3-s model had the greatest performance overall, with a mean F1-score of 0.98 and recall of 0.97, respectively. During this study, a custom CNN model was developed using convolutional neural networks, hyperparameter optimization, and a hybrid model. The results obtained from this model were very impressive. The unique CNN model architecture that we have developed demonstrates the efficiency and usefulness of deep learning approaches for classification and discrimination.

Uysal²² has developed a hybrid artificial scheme that can recognize photographs of Mpox lesions on the skin. These pictures of skin were obtained from an image repository that is open to the public. The multi-class structure of this dataset includes categories of information about chickenpox, measles, Mpox, and normal. It can be seen in the original dataset that there is an imbalance in the distribution of the data between the classes. This inconsistency was resolved by employing several different strategies for enhancing and processing data. After completing these stages, cutting-edge deep-learning models were utilized to look for indications of Mpox. For this work in particular, we produced a hybrid model by combining the two deep learning models that performed the best with the (LSTM) model. The result was enhanced classification results. The hybrid AI system recommended for detecting Mpox achieved an accuracy rate of 87% during testing, and it obtained a score of 0.8222 on Cohen's kappa.

Uzun Ozsahin et al.²¹ showed a study in which they analyzed two digitized skin photographs. One of the photos depicted Mpox, while the other one depicted chickenpox. In this particular instance, we employed a CNN with a depth of two and four convolutional layers. We used three layers, one each after the additional, third,

and layers. Ultimately, we examined how well our model achieved compared to the most advanced deep-learning skin lesion detection algorithms currently available. The performance of every DL model was outperformed by our proposed CNN model, which achieved an accuracy of 99.60% on the test. In addition, a total score of 99.00% was achieved when accuracy, recall, and F1 were all added together. Because of this, Alex Net was able to attain a level of accuracy of 98.00%, which was higher than that of any other pre-trained model.

The VGGNet trained with VGG16 and VGG19 achieved the poorest overall, with an accuracy of only 80%. Because it is based on an innovative mix of model and image augmentation practices, the CNN model that has been proposed is generalized, and it prevents overfitting. Using this method, digital skin photos taken from patients with Mpox may be evaluated promptly and with high precision. It has been recommended and evaluated that a deep learning-based CNN perfect called MonkeyNet, a modified version of Bala et al.'s²⁴ DenseNet-201 model, be used. In this study, a deep network was suggested, which attained an accuracy of 93.19% when using the original dataset and 98.91% when using the upgraded dataset to diagnose Mpox sickness. This version displays the Grad-CAM, which indicates how effective the model is and displays the infected regions in each class photo. The proposed approach would not only help to stop the spread of Mpox, but it will also make it easier for medical personnel to make an accurate analysis of the sickness at an earlier stage.

Almufareh et al.³¹ present a way for diagnosing non-contact and non-invasive MPX, in addition to being more intelligent and secure than traditional methods by evaluating pictures of skin lesions. This method may be found in their work. Deep learning algorithms are utilized in the proposed method to determine whether or not skin lesions are positive for MPXV. We use two datasets (MSLD) and Dataset (MSID) to evaluate the proposed strategy's effectiveness in skin lesions caused by Mpox. When comparing the results of several deep learning models, sensitivity, specificity, and total accuracy were the metrics that were utilized. The technique recommended to identify Mpox has yielded extremely positive results, demonstrating that it can be used broadly. Even in undeveloped countries where laboratory facilities are scarce, this ingenious and inexpensive solution has the potential to be put to good use.

An ensemble learning-based methodology is proposed by Pramanik et al.³² to determine whether skin lesions in photographs contain the Mpox virus. Initially, we evaluate three pre-trained base learners to acquire the best possible performance on the Mpox dataset. These three base learners are Inception V3, Xception, and DenseNet169. In addition, we feed the probabilities derived from these deep models into our ensemble system in the form of inputs. We present a Beta function-based normalization scheme of probabilities to combine the outcomes to learn an effective aggregation followed by the ensemble. This is done to achieve the goal of learning a compelling combination. Using a methodology known as five-fold cross-validation and a lesion available to the general public, the system is subjected to a comprehensive test. The model received average scores of 93.39 for accuracy, 88.91 for precision, 96.78 for recall, and 92.35 for F1 overall. Yasmin et al.³³ are working to find a solution to this problem by developing a diagnostic model for Mpox through the application of image-processing procedures. To achieve this goal, many data augmentation procedures have been utilized to remove the likelihood of the perfect overfitting. After cleaning the data, we used the transfer-learning method to input it into six distinct Deep Learning (DL) models. After analyzing each model's precision, recall, and accuracy performance matrices, we selected the one that provided the best overall results. "PoxNet22" was proposed as a solution to improve upon the model that PoxNet22 is superior to other methods of Mpox classification because it achieves perfect scores on all three measures of precision, recall, and accuracy. When it comes to classifying and diagnosing Mpox sickness, medical professionals will find the findings of this study to be quite helpful.

Ariansyah et al.³⁴ provide an image classification that may be utilized to identify the symptoms of the Mpox strategy to model pictures. This approach was built on a Network architecture and included VGG-16 transfer learning. A model developed on one dataset may be adapted to another with the help of a transfer learning technique. Because of this, the model could gain knowledge from the previous data and then apply that knowledge to the new data. Researchers have recommended employing deep learning to anticipate new data because it is so excellent at spotting patterns in images that are similar to one another. As a direct result of this, the VGG-16 model can achieve a highly acceptable 83.333% accuracy at the epoch of 15.

REN and Emrullah³⁵ have attempted to identify Mpox using photo datasets available on Kaggle by utilizing Convolutional Neural Network models. The models are referred to by their names, which are EfficientNetB3, ResNet50, and InceptionV3. The findings from the three models show that resNet50 achieves the highest performance levels overall. The accuracy of resNet50, 94.000%, makes it the most accurate model. It is possible to evaluate the usefulness of the models along four different aspects. Some of the terms that are commonly used for them are precision, recall, F1, and accuracy. These models demonstrate that it is feasible to classify Mpox accurately. Therefore, these prototypes can be utilized for the projects now being worked on.

Farouk and Abd Elaziz³⁶ have attempted to employ various models in artificial intelligence to help many medical professionals categorize Mpox. Deep learning (DL), an area of intelligence, has been suggested as a potential solution to this problem. To improve the data and draw out more features for the suggested model, certain VGG16 and Inception models are applied. While (SVM) was used for classification, Particle Swarm Optimization (PSO) was used for feature selection and optimization of the neural network's parameters. Both of these methods were used in conjunction with one another. In the end, the model was evaluated using a confusion matrix, which showed that although the VGG16 model had an accuracy of 85%, the new accuracy had grown to 94.5% following enhanced PSO. The model's accuracy increases to 90.2% once improved PSO is adopted, from 75.2% in the Inception state when it was first used. These findings can help classify Mpox instances and diagnose those patients. It is common knowledge that maintaining good hygiene, avoiding contact with animal faeces, and being vaccinated to strengthen one's immune system effectively prevent disease transmission.

Research gap

If we want to lessen the sum of resources wanted for dispensation without losing the relevant feature dataset, a feature extraction approach that extracts novel features from the original dataset is highly useful. As a bonus, feature extraction can cut down on the amount of unnecessary features for a specific research project. Early features undergo a stunning metamorphosis into more significant characteristics when they are subjected to feature extraction. To reduce the high dimensionality vector, feature extraction is used to generate new features dependent on the original set. Algebraic transformation is used in the transformation process, and optimization criteria are used to determine the final form³⁷. In addition, feature extraction can manage crucial data while confronting problems with many dimensions³⁸. These dimensionality reduction methods preserve the relative distance among features and the underlying potential structure of the original data to minimize the sum of information lost during the feature translation process³⁹. Compared to feature selection approaches, feature extraction is more robust against overfitting and yields accurate classification results. However, sometimes the data description is lost in specific datasets, and the cost of this operation is high⁴⁰. Techniques now demonstrate conclusively that pre-trained models may serve as feature extraction models. However, only a single model is considered for feature extraction. However, the study also considered an ensemble of pre-trained feature extraction and classification models. Furthermore, current models are considered part of the hair removal process in skin photos. This is a crucial pre-processing step in improving categorization. Because hair in photos might mislead the model to incorrectly label cases of Mpox as healthy and prevent an appropriate diagnosis, to prevent this, researchers thought of using a deep learning model to filter out hair before using the image. As a result, the study successfully diagnosed Mpox and improved the classification accuracy rate relative to previous models.

Related terminologies

A technique used in deep learning and machine learning, transfer learning replicates the results of one model in another. By reusing a model that has already been trained for a different purpose, we can address all or part of a problem. In transfer learning, machines apply their knowledge to make more accurate predictions in unrelated tasks. The most notable advantages of transfer learning are its short training time, improved neural network performance, and lack of need for massive volumes of data. The state- that we have chosen is briefly described below.

VGGNet

Andrew Zisserman and Oxford University introduced the VGGNet CNN architecture in 2014. To create it, we deepened the publicly available CNN architecture to 16 or 19 layers, thus the names VGG-16 and VGG-19. There are 138 parameters in the VGG-16 architecture and 144 million in the VGG-19. A total of 13 convolutional layers make up the VGG-16 model. VGG uses a smaller filter size (3 3) than other methods. Its effective receptive field is equivalent to a single 77 convolutional layer. Layers make up the VGG-19 model. Different implementations of VGGNet use two 4096-channel layers and a further 1000-channel layer to prepare for 1000 labels.

ResNet

In 2015, Kaiming Huang established ResNet. Each framework layer is then decided from the residuary function by referencing its input layer, which is the concept behind it. With a reduced error rate of 3.57%, this network triumphed in the 2015 ILSVRC competition. At this stage, the ResNet model is optimal, and further improvements in accuracy are possible. We have utilized ResNet-18 and ResNet-101, all of which have been explicitly pre-trained for this purpose. The fundamental design of these networks is outlined in Table 1. The network takes in data at a 224-by-224 input size. The first convolutional layer and the last three layers are attached in the specified topologies. The depth of a deep network may be adjusted by increasing the number of its internal convolution layers.

Layer name	ResNet-18	ResNet-50	ResNet-101	Size of output
Conv1	7×7, 64-Stride 2 3×3=-max pool, Stride 2	7×7, 64 Stride 2 3×3 max pool, Stride 2	7×7, 64 Stride 2 3×3 max pool, Stride 2	112×112
Conv2_x	[(3×3,256@3×3,256)]×2	[(1×1,256@3×3,256@1×1,1024)]×6	[(1×1,256@3×3,256@1×1,1024)]×23	14×14
Conv3_x	[(3×3,512@3×3,512)]×2	[(1×1,512@3×3,512@1×1,2048)]×3	[(1×1,512@3×3,512@1×1,2048)]×3	7×7
Conv4_x	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	56×56
Conv5_x	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 18 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	28×28
Pooling	Average pooling			
FC	Fully connected			
Softmax	Softmax			

Table 1. The RestNet-18, 50 and RestNet-101 constructions.

GoogLeNet

CNN, developed by Google using the Inception framework, is called GoogLeNet. The ILSVRC 2014 competition was won by GoogLeNet (Inception V1). The 6.67% mistake rate it attained is almost as low as the human performance judges must now evaluate for competitions. The network may pick from various convolution filter blocks thanks to the Inception module. The maximal max-pooling layer with stride 2 is occasionally used to halve the grid's resolution by the Inception network. GoogLeNet has 22 deep CNN layers in its design but only 4 million parameters, down from 60 million in AlexNet. Four separate branches grow out of it at the same time. Convolution layers are utilized in the first three levels with kernel sizes of 1, 3, and 5 pixels, respectively. Using a window size of 11, two intermediary branches in the input channel can be convolved to simplify the frame. Each of the four branches is given sufficient padding to ensure the inputs and outputs have identical heights and widths. When the output from each branch is connected, an inception block is formed. There are over 6.8 million tunable values within it. Six convolutional layers, three 11 (for dimensionality reduction), three 33, and four 77 convolutional layers are included in the GoogLeNet architecture's nine inception blocks, along with four layers of max-pooling, two levels of normalization, average pooling, and an FC layer. All convolution layers use the ReLU activation function, while the FC layer employs drop regularisation. In the last layer, we employ a softmax function.

Inception-ResNet-v2

This network is a variation of Inception-v3 that incorporates elements of the ResNet architecture. Batch normalization is only implemented in the Inception-ResNet-v2 standard layers. Residual modules can expand the networks' depth and the sum of inception blocks. One stem block (6 convolutional blocks) and three inception sets (max pooling layer) comprise the Inception-ResNet-v2 architecture. There are five modules and seven convolution blocks in the first layer. Two depletion blocks with unique convolutional layers, average pooling layers, and FC layers, as well as ten inception modules with five convolution blocks in the first block and five inception units with four blocks in the third and final block. The function is used at the output layer.

Inception-v3

Standardization for the side head layer, improved label smoothing, and the use of auxiliary classifiers to further broadcast label info across the network are all features of the CNN architecture of the Inception series that are also present in this network. To train, Inception-v3 employed 1,000,000 photos from thousands of classes in the ImageNet datasets. Inception v3 achieved an accuracy of over 78.1% on the ImageNet dataset. This structure represents the culmination of several theories developed by many scientists throughout time. Convolutional layers, average pooling layers, maximum pooling layers, concatenation layers, dropout layers, and FC layers are a few of the symmetric and asymmetric building elements that comprise the architecture. Since it is well suited for activation input, the design extensively uses the batch norm distribution norm. Softmax is used to determine the loss. Both the number of learnable limits and the degree of network complexity have been reduced in this model.

Materials and methods

Dataset

As a result of its fast spread, Mpox has caused public health issues in more than 65 countries. Stopping its fast development requires prompt clinical identification. Many biochemical assays, including (PCR) tests, are not readily available⁴¹. Although, using computer vision techniques, Mpox can be detected by looking at pictures of skin lesions. However, at present, no such data is available. Thus, the "Mpox Skin Lesion Dataset (MSLD)" was sourced from Kaggle and is publicly available at <https://www.kaggle.com/datasets/nafin59/Mpox-skin-lesion-dataset>. The study acknowledges that Chickenpox and Measles lesions visually resemble Mpox lesions (specifically the rash and pustules in the early stages). As a of this resemblance, we included images of Chickenpox and Measles in the Mpox classification task—not because those diseases are Mpox, but to help the classifier learn to distinguish between similar-looking skin conditions. Thus, the dataset includes 228 total images: 102 images are labeled as Mpox (positive cases). At the same time, 126 images are labeled as other skin lesions, which include Chickenpox and Measles (negative or non-Mpox cases). Including these similar diseases makes the classifier more robust. It ensures that the model learns not just to recognize Mpox but also to differentiate it from visually similar diseases. The images of actual Mpox cases used in the study are shown in Fig. 1

Data preprocessing using

This study employs a data augmentation approach for preprocessing and a deep learning model for hair removal.

Augmentation technique

The ImageDataGenerator from the Keras image processing package is used to upsurge the extent of the dataset. The ImageDataGenerator function allows various manipulations, including mirroring, flipping, and resizing. The ImageDataGenerator facility is described in⁴². The picture data in this work is enhanced using the parameters listed in Table 2. As proposed in⁴³, the generator and facility types are randomly picked. Images that have been enhanced using Image Generator or another tool are saved here so that the process may be repeated easily. After processing, there was a 14-fold increase in the number of photographs. There are 1428 images in the "Mpox" category and 1764 in the "Others" category.

Algorithm 1 shows the algorithm for data augmentation procedures used in this study.

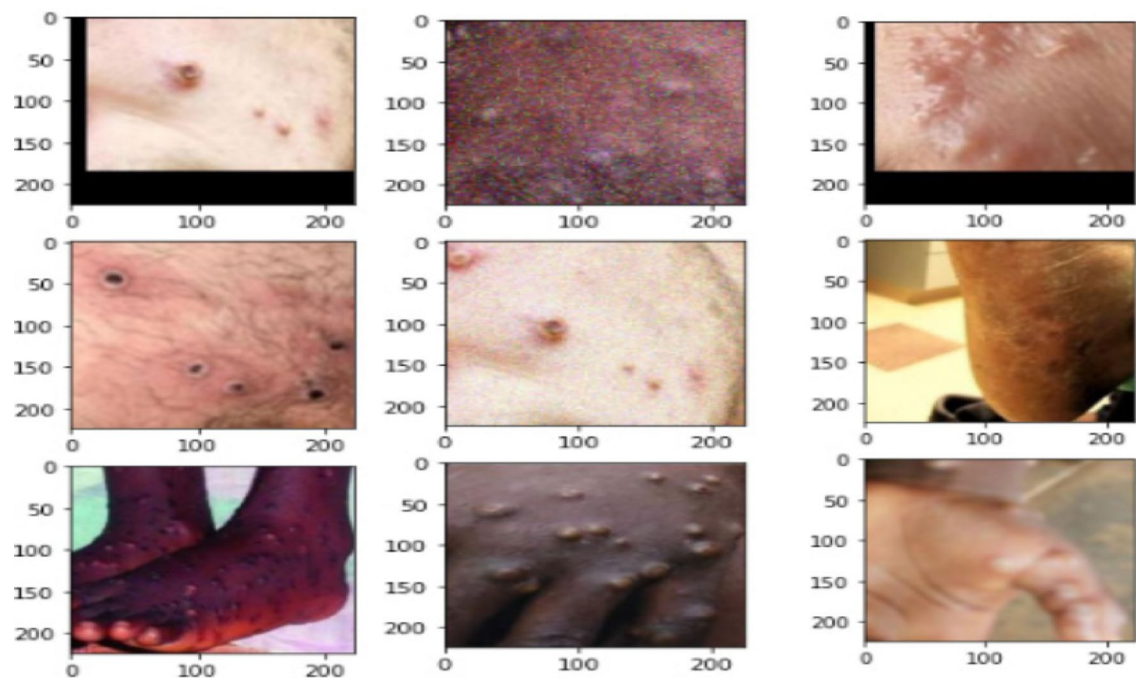


Fig. 1. Mox Images adapted from Kaggle dataset⁴⁹.

Generator Kind	Facility
Width shift	Up to 2%
Fill mode	Reflective
Horizontal flip	True
Rotation range arbitrarily	0–45°
Zoom variety	2%
Height shift Up to	2%
Shear range	2%

Table 2. Data augmentation procedures used in this study.

Input: read original image samples x using OpenCV.
Resize the image to 128 × 128.
Store resize image as an array inside a list.
Call Image data generator function
for n 1 to 20 do
Batch size = 16
Save to directory
Save the format as "png."
end for
End

Algorithm 1. For Code Augmentation.

Hair removal process using deep learning model
We provide a reconstruction loss function and discuss our proposed deep-learning model for removing hair from the input photos. We have developed and proposed a convolutional encoder-decoder architecture to remove hair from input pictures. Figure 2 shows the construction of a 12-layer model. To train our model, we

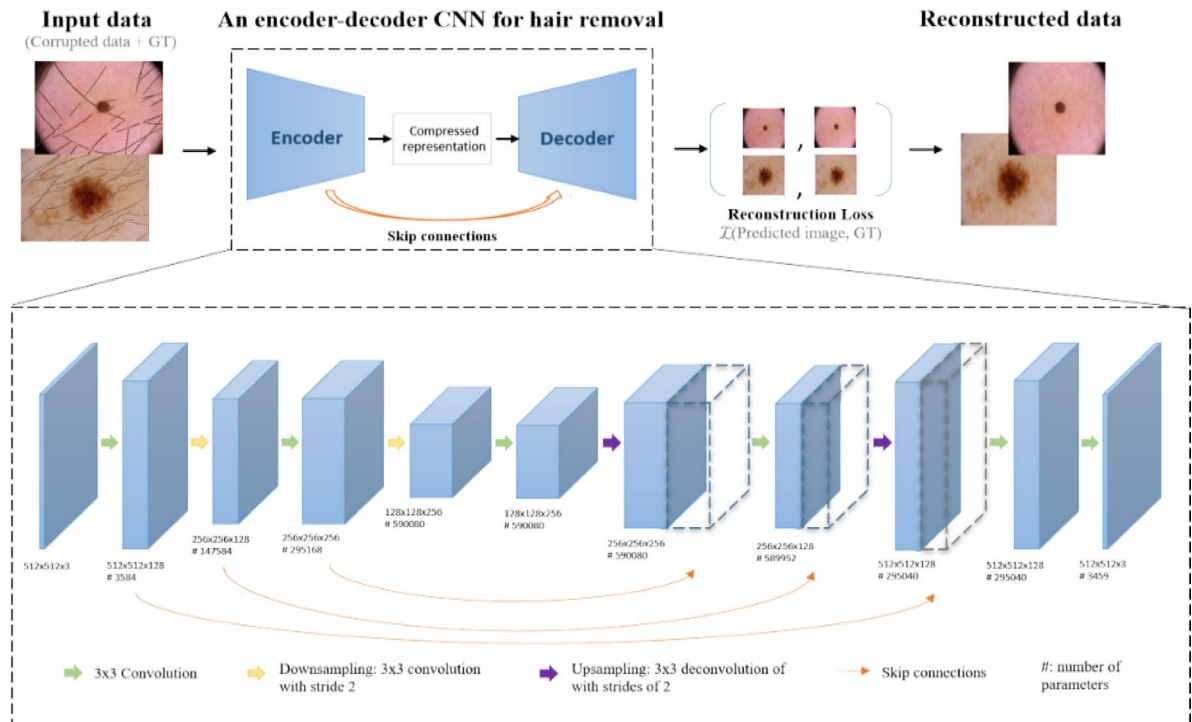


Fig. 2. The architecture of the proposed network.

employ picture pairings consisting of a hairless reference image and an image with synthetic hair. The result is a hair-free version of the rebuilt image. The network uses our suggested loss function during training to determine how well the output picture compares to the hairless reference. We then detail the suggested model and loss functions for this important job.

Our model begins with an encoder network as its initial component. The output is a concealed characteristic, and the input size is 512 by 512 by 3. In its search for these traits, the encoder often disregards background noise. We want the network to disregard the hair as background noise and provide a picture of smooth skin as its output. A decoder, which attempts to fill in the blanks in the model's high-level feature representation, is located in the model's second layer. The final product is a skin-hair-free, 512-by-512-by-3 replica of the original image. There are two blocks in both the encoder and decoder. To lower the spatial resolution, the encoder uses a down-sampling operation performed by a two-stride 3 3 convolution on the output of the first block and the output of the second block, respectively.

In contrast, the decoder up-samples the feature map in each block using a 3×3 deconvolution with 2×2 strides in each dimension. The up-sampled output is then joined with the feature map from the encoder's layer with the same resolution through a skip link. As a result, the decoder's ability to restore visual information is enhanced. The combined data is then subjected to a convolution with three identical terms. The feature map is then convolved by three times the sum of output channels in the final block. This design was selected because we want to see how well autoencoders perform as a denoising solution, which is why we use them. Our little data suggests that a smaller network is better suited to learn the task at hand successfully.

Reconstruction loss function The network is instructed in its learning by a criteria called the loss function, which is a numerical representation of the model's faults. It is calculated by comparing the network's prediction to the matching hairless Ground Truth (GT) picture. Several different loss functions have been put to use in picture restoration projects. The (MSE) and the (MAE) are examples of common types of loss. The disparity between the two pictures' corresponding pixels is the sole determinant of these metrics. Since a pixel's noise should be treated equally from the inaccuracy of its adjoining pixels, the results may need more quality from a human's perspective. Other loss measures, such as the (SSIM) and the Multiscale Structural Similarity Metric (MSSSIM), dependent on local brightness, contrast, and structure, have been proposed to address these shortcomings.

Inspired by the findings of Liu et al. in⁴⁴, we suggest capturing the best aspects of the loss functions that assess statistical variables locally combined with losses to create outcomes that appeal to a human observer.

This leads us to the following definition of our reconstruction loss:

$$\mathcal{L}_{rec} = \alpha L_1^{foreground} + \beta L_1^{background} + \gamma L_2^{composed} + \delta \mathcal{L}_{SSIM} + \lambda \mathcal{L}_{tv} \quad (1)$$

where $\alpha, \beta, \gamma, \delta$ and λ are linear mixture weights that characterize the renovation loss function. Due to the high cost of doing a grid search and the large sum of searchable parameters, we decided on a random hyperparameter

search. Here is how we did it: We ran our model 10 times, each giving each weight a random value between 0 and 10.

The term $L_1^{foreground}$ is the L_1 distance among the original and the forecast of the network only among those pixels fitting to the hair areas.

Next, $background$ estimates the $L1$ distance among the network's forecast only among the contextual pixels, which is for regions.

Then, $L_2^{composed}$ computes the $L2$ limited to the hair areas but standardizes over all pixels rather than just the number of hair pixels. The loss function is computed using the SSIM metric for the entire picture, thus the name LSSIM. Last but not least, we utilize a total variation to ease the change in projected values for hair patches based on their surroundings.

Features extraction using deep learning model

After the photos have been preprocessed, features are extracted using one of five distinct pre-trained models and then fused into a single form before being passed on to the classifier. Pre-trained models are described in the section on related terminology, where the fusion process is also discussed. In pattern recognition, here is where the spotlight shines brightest. Fusion of characteristics from many levels or branches. Concatenation and summing are two distinct techniques that can be used to conduct feature fusion. We use the (CSID) fusion method to establish connections between the feature vectors we have chosen in the matrix. Equation (2) provides an assessment of the merged features:

$$\chi_{o,p} = \min_{\xi_{o,p}} \left\| O_l - \sum_{p=1}^{Q_t} h_{t,p} * \xi_{t,p} - \sum_{m=1}^{M_o} l_{o,p} * \xi_{o,p} \right\|_2^2 + \diamond_0 \sum_{p=1}^{Q_t} \|\xi_{t,0}\|_1 \quad (2)$$

This operation endures till all pairs are associated. $\chi_{o,p}$ is the last fused vector. The fused features from the five pre-trained models serve as input to the encoder part of the improved U-Net architecture.

Classification using improved U-Net architecture with VGG-16

Two- and three-dimensional U-Net modified models^{45–50} have been used in research on Mpox. The fundamental capability of U-Nets to generate pixel-wise masking for each duplicate object is the driving force behind their use. By assigning each pixel in each image to one of two categories—positive or negative—the goal is to notice and identify the positions and forms of various items in the image. In this way, the model establishes a region-wise boundary as interest. The Dense blocks' contributions to feature mapping in the enhanced U-Net counters the vanishing gradient issue in convolutional blocks. To improve model stability and performance, the proposed change (a) uses dense-convolutional blocks rather than convolutional sampling, which increases feature re-usability; (b) employs (c) incorporates Batch Normalisation (BN) layers in dense blocks. Tensor Processing Unit (TPU) environments are used to develop these models. It is an ASIC, or application-specific integrated circuit, developed for a particular purpose and built to scale computationally. It was utilized to make model preparation more effective, cutting down on training time by a factor of 10, and it operates ten times quicker than GPUs.

U-Net design has contributed much to the development of biomedical image processing. Encoder, bottleneck, and decoder are the three components that make up U-Net's entire architecture. The first step is to feed the photos into the encoder. A 256-by-256 grid served as the input. Downsampling occurs in the encoder due to the usage of convolution and pooling. As a result of this processing, the original image size is decreased, but the spatial characteristics are preserved. Images are downsampled many times before being fed into a deep neural network. A decoder receives the pictures as input. The pictures are scaled up by the decoder, which employs upsampling methods. Upscaling photos often requires a deconvolution technique. Jump links connect the upsampling and downsampling nodes. Following upsampling operations, the network produces a picture containing feature masks. The optimizer, which minimizes the loss, updates the model weights with a batch scope of 32.

$$loss(h, \tilde{h}) = - \sum_{i=1}^C \tilde{h}(i) \log(h(i)) \quad (3)$$

where $\tilde{h}(i)$ is the likelihood of the i th class. Classification (C) was binary: either positively or negatively labeled areas. Here is how to use the Adam optimizer to refresh weights and bias:

Booting of weights:

$$\rho_x \leftarrow 1, \rho_y \leftarrow 1, x \leftarrow 0, y \leftarrow 0 \quad (4)$$

Inform rubrics for Adam Optimizer:

Where

$$\rho_x \leftarrow \beta_x \rho_x \quad (5)$$

$$\rho_y \leftarrow \beta_y \rho_y \quad (6)$$

$$x \leftarrow \beta_x x + (1 - \beta_x) \nabla_x J \quad (7)$$

$$y \leftarrow \beta_y y + (1 - \beta_y) (\nabla_z J \odot \nabla_x J) \quad (8)$$

$$z \leftarrow z - \alpha \left(\frac{x}{\sqrt{y + \epsilon}} \frac{\sqrt{1 - \rho_y}}{1 - \rho_x} \right) \quad (9)$$

x , y represent vectors, respectively; β_x and β_y characterize exponential decay charges for the moment vector, correspondingly; ρ_x and ρ_y refer to learning rates time decay factor, correspondingly, which are similar to momentum and connected to the updates; α represents the learning rate (Eq. 9); $\nabla_x J$ characterizes the gradient of the cost function, J ; ϵ is a considerable value taken to avert separation by zero illness (Eq. 9); \odot mentions multiplication (Eq. 8) and the processes under the root are also handled element-wise (Eq. 9).

The (LR) was set to $1 \times e^{-4}$ in this procedure. The sum of epochs was set to 40 per fold.

The enhanced U-Net consists of input and output layers an Encoder, Skip connections, and a Decoder. Upsampling and downsampling are used to create the encoder and decoder zones. Figure 3 depicts the full design architecture. In our improved U-Net, we incorporated *dense blocks* into the encoder to enhance feature reuse and gradient flow, inspired by the DenseNet architecture. For the decoder, we utilize a modified version of the VGG-16 model as a backbone for upsampling and refinement of features, leveraging its well-established ability to extract and preserve semantic information. While this configuration is indeed unconventional, it attempted to improve standard decoder designs in our application context. The layers from VGG-16 used in the decoder include layers such as conv3_1, conv4_1, conv5_1] adapted for upsampling via transposed convolutions. A layer and a batch normalization (BN) layer with a ReLU activation function make up the dense blocks. For the decoder part, we used VGG-16 architecture. The VGG-16 is used as a backbone within the U-Net framework, which receives the concatenated feature maps from the fused models. This is accomplished by imagining the pre-training of the aforementioned architecture in Fig. 3.

A few blocks shortened the encoder section of the current VGG-16 model. The pre-trained model was missing its first few layers. Horizontal and vertical edges are detected in the first layer of the perfect. The model becomes more sophisticated as one travels further into the network, making feature extraction easier. For nested layers, we use concatenation skip connections. With dense blocks, accessing many feature channels in the network's final layers is possible while maintaining small models and high feature re-usability. Transpose layers, as seen in decoder blocks, increase the dimensionality of feature maps. The model is fed the segmentation map produced at the end as the region to be predicted in the output region. Finally, the model employs a softmax layer to determine whether a pixel has a MpoX.

Experimentation and results

Python 3.7.0 was used for our investigations, with the deep learning basis implemented in PyTorch 1.2.0.

Data splitting: The dataset was divided into 70% training, 15% validation, and 15% testing sets, ensuring class balance across all splits.

Parameter tuning: Hyperparameters such as learning rate, batch size, and regularization weight were optimized using a grid search on the validation set.

Convergence criteria: Training was stopped early if the validation loss did not improve for 10 consecutive epochs (early stopping), with the best model weights retained.

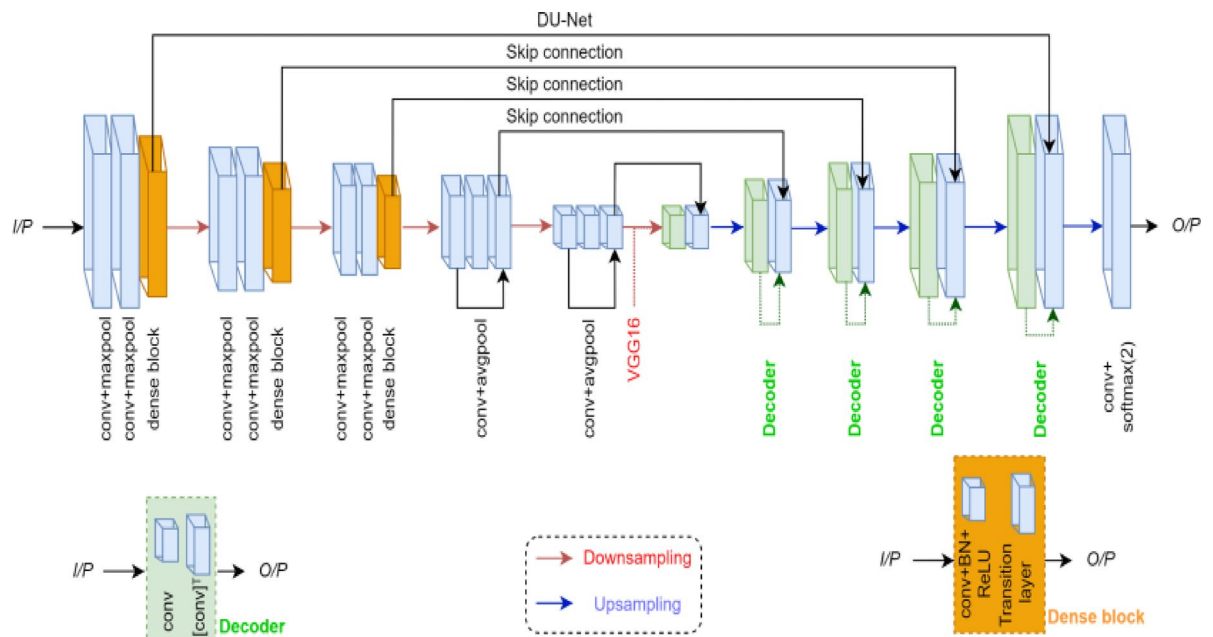


Fig. 3. Proposed architecture: improved U-Net.

Model	Recall	Acc	AUC	F1	Precision	AP
Swin transform	0.8383	0.8477	0.8807	0.8284	0.8187	0.7572
U-Net	0.8263	0.8479	0.8913	0.8263	0.8263	0.7590
ResNet	0.7964	0.8346	0.8745	0.8085	0.8210	0.7431
VGGNet	0.7844	0.8478	0.8851	0.8187	0.8562	0.7661
Improved U-Net	0.8503	0.8688	0.9058	0.8503	0.8503	0.7886

Table 3. Analysis of proposed classifier based on 60–40.

Model	Acc	AUC	F1	Precision	Recall	AP
ResNet	0.8482	0.8797	0.8095	0.8151	0.8041	0.7339
VGGNet	0.8451	0.8649	0.8162	0.8506	0.7844	0.7618
Swin transform	0.8583	0.8787	0.8333	0.8599	0.8084	0.7791
U-Net	0.8688	0.9058	0.8503	0.8503	0.8503	0.7886
Improved U-Net	0.8766	0.9135	0.8563	0.8750	0.8383	0.8044

Table 4. Comparative Analysis of proposed classifier based on 70–30.

Training duration: On average, training took approximately 3 h per fold using an RTX2060 graphics processing unit (GPU) with 6 GB of RAM and an AMD CPU R7-4800 with 2.9 GHz and 16 GB of RAM. with convergence typically achieved within 40–50 epochs.

Performances metrics

We utilized the measures of accuracy characteristics (AUC), F1 score (F1), precision (P), recall (R), and average precision (AP) to gauge how well our model performed. The e-Time, or average training time per epoch, was used to calculate the computational cost of training. Using Eq. (10), we can get the precision:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

where TP characterizes a positive result, TN a result, and FN a negative result.

Equation (11) may be used to determine the level of accuracy:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

The recall is intended by the subsequent Eq. (12):

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

The F1 is intended by the subsequent Eq. (13):

$$F1 = 2 \frac{Precision * recall}{Precision + recall} \quad (13)$$

The AUC curves evaluate the false positive and true positive rates at various cutoffs. Precision at each threshold is weighted equally, which is how AP summarises a precision-recall curve.

Validation analysis of proposed classifier

Table 3 represents the analysis of a proposed classifier based on 60–40. The analysis of the ResNetmodel reaches an Acc value of 0.8346, an AUC rate of 0.8745, an F1 score value of 0.8085, a precision value of 0.8210, a recall value of 0.7964, and an AP value of 0.7431, respectively. Also, the VGGNet model reaches an Acc value of 0.8478, an AUC value of 0.8851, an F1 score value of 0.8187, a precision value of 0.8562, a recall value of 0.7844 and an AP value of 0.7661, respectively. In the next Swin Transform model reaches an Acc value of 0.8477, an AUC value of 0.8807, an F1 score value of 0.8284, a precision value of 0.8187, a recall value of 0.8383 and an AP value of 0.7572, respectively. The U-Net model reaches an Acc value of 0.8479, an AUC value of 0.8913, an F1 score value of 0.8263, a precision value of 0.8263, a recall value of 0.8263, and an AP value of 0.7590, respectively. Moreover, the Improved U-Net model reaches an Acc value of 0.8688, an AUC value of 0.9058, an F1 score value of 0.8503, a precision value of 0.8503, a recall value of 0.8503 and finally, an AP value of 0.7886, respectively.

Table 4 signifies that the Comparative Analysis of the proposed classifier is based on 70–30. The analysis of the ResNetmodel reaches an Acc value of 0.8482, an AUC rate of 0.8797, an F1 score value of 0.8095, a precision value of 0.8151, a recall value of 0.8041, an AP value of 0.7339 respectively. After the VGGNetmodel reaches an

Model	Acc	AUC	F1	Precision	Recall	AP
ResNet	0.8482	0.8797	0.8095	0.8151	0.8041	0.7339
VGGNet	0.8241	0.8446	0.8069	0.7778	0.8383	0.7229
Swin Transform	0.8451	0.8649	0.8162	0.8506	0.7844	0.7618
U-Net	0.8661	0.8904	0.8440	0.8625	0.8263	0.7888
Improved U-Net	0.8871	0.9088	0.8693	0.8827	0.8563	0.8189

Table 5. Experimental analysis of proposed classifier based on 80–20.

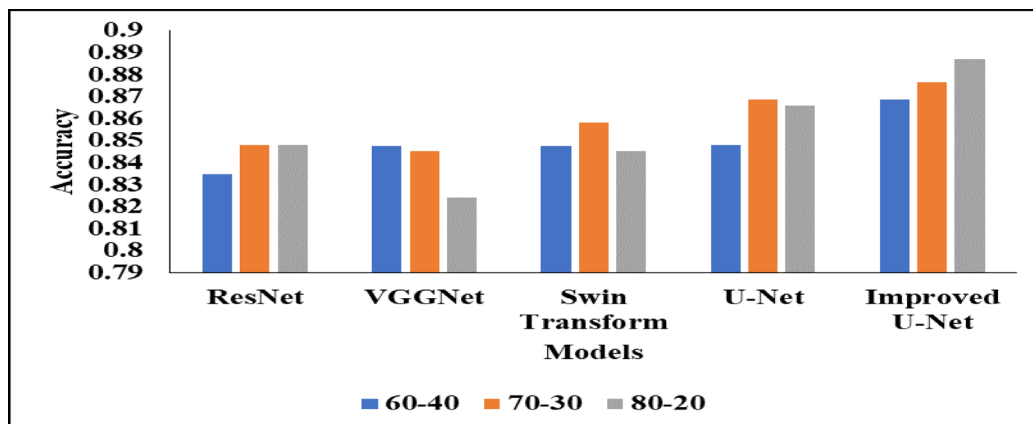


Fig. 4. Accuracy Comparison.

Acc value of 0.8451, an AUC value of 0.8649, an F1 score value of 0.8162, a precision value of 0.8506, a recall value of 0.7844, and an AP value of 0.7618 respectively. Next Swin Transform model reaches an Acc value of 0.8583, an AUC value of 0.8787, an F1 score value of 0.8333, a precision value of 0.8599, a recall value of 0.8084, and an AP value of 0.7791 respectively. U-Net model reaches an Acc value of 0.8688, an AUC value of 0.9058, an F1 score value of 0.8503, a precision value of 0.8503, a recall value of 0.8503, and an AP value of 0.7886, respectively. Finally, the Improved U-Net model reaches an Acc value of 0.8766, an AUC rate of 0.9135, an F1 score value of 0.8563, a precision value of 0.8750, a recall value of 0.8383, an AP value of 0.8044, respectively.

Table 5 above represents the experimental analysis of the proposed classifier based on 80–20. In an experimental analysis of the ResNet model reaches an Acc value of 0.8482, an AUC value of 0.8797, an F1 score value of 0.8095, 0.8151, a recall value of 0.8041, and an AP value of 0.7339, respectively. Next, the VGGNet model reaches an Acc value of 0.8241, an AUC value of 0.8446, an F1 score value of 0.8069, a precision value of 0.7778, a recall value of 0.8383, and an AP value of 0.7229, respectively. The Swin Transform model reaches an Acc value of 0.8451, an AUC value of 0.8649, an F1 score value of 0.8162, a precision value of 0.8506, a recall value of 0.7844, and an AP value of 0.7618, respectively. U-Net 0.8661 model reaches an Acc value of 0.8904, an AUC value of 0.8440, an F1 score value of 0.8625, a precision value of 0.8263, and an AP value of 0.7888, respectively. While the Improved U-Net model reaches an Acc value of 0.8871, an AUC value of 0.9088, an F1 score value of 0.8693, a precision value of 0.8827, a recall value of 0.8563, and an AP value of 0.8189, respectively. Figure 4 depicts the accuracy level in comparison to the various models. Figures 5 and 6 highlight the graphical representation of the proposed model in terms of AUC and F1 score analysis, respectively. Similarly, Fig. 7 indicates the model precision, while Fig. 8 shows the analysis of the classifier in terms of recall, and Fig. 9 depicts the graphical comparison in terms of AP.

Validation analysis of proposed hair removal technique

Table 6 signifies that the Analysis of DL is based on Multiclass classification. The analysis of the LR model reaches an accuracy of 0.8864 ± 0.0028 , a precision value of 0.8604 ± 0.0032 , a recall value of 0.8748 ± 0.0041 , and an F1-score value of 0.8675 ± 0.0032 respectively. DNN model reaches an accuracy of 0.8977 ± 0.0033 , a precision value of 0.8693 ± 0.0055 , a recall value of 0.8890 ± 0.0029 , and an F1-score value of 0.8790 ± 0.0040 respectively. After that, the MLP model reaches an accuracy of 0.7197 ± 0.0025 , a precision value of 0.6364 ± 0.0044 , a recall value of 0.6730 ± 0.0038 , and an F1-score value of 0.6542 ± 0.0034 respectively. AE model reaches an accuracy of 0.8335 ± 0.0020 , a precision value of 0.7870 ± 0.0029 , a recall value of 0.8158 ± 0.0030 , and an F1-score value of 0.8011 ± 0.0024 respectively. CNN model reaches an accuracy of 0.9021 ± 0.0039 , a precision value of 0.8941 ± 0.0047 , a recall value of 0.8778 ± 0.0049 , an F1-score value of 0.8859 ± 0.0046 respectively.

Table 7 represents the analysis of the DL model based on Binary classification. The analysis of the LR model reaches an accuracy of 0.8933 ± 0.0029 , a precision score of 0.8656 ± 0.0039 , a recall value of 0.8795 ± 0.0037 , and an F1 score of 0.8725 ± 0.0035 respectively. DNN model reaches an accuracy of 0.8990 ± 0.0063 , a precision value of 0.8791 ± 0.0065 , a recall value of 0.8833 ± 0.0079 , and an F1 score of 0.8812 ± 0.0069 respectively. Next,

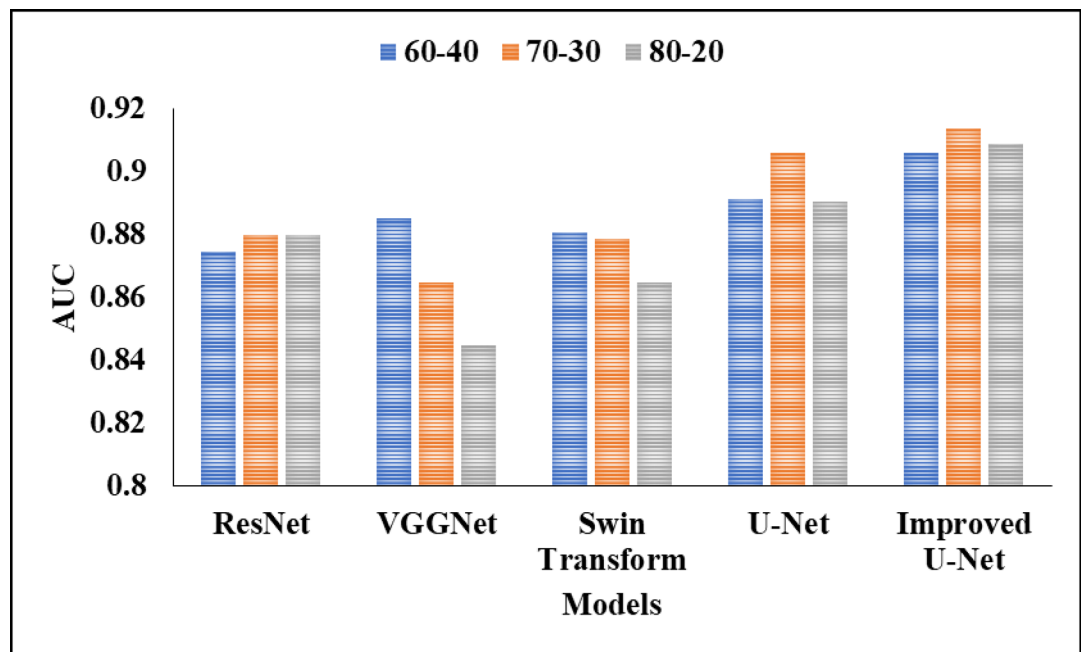


Fig. 5. Graphical representation of the proposed model in terms of AUC.

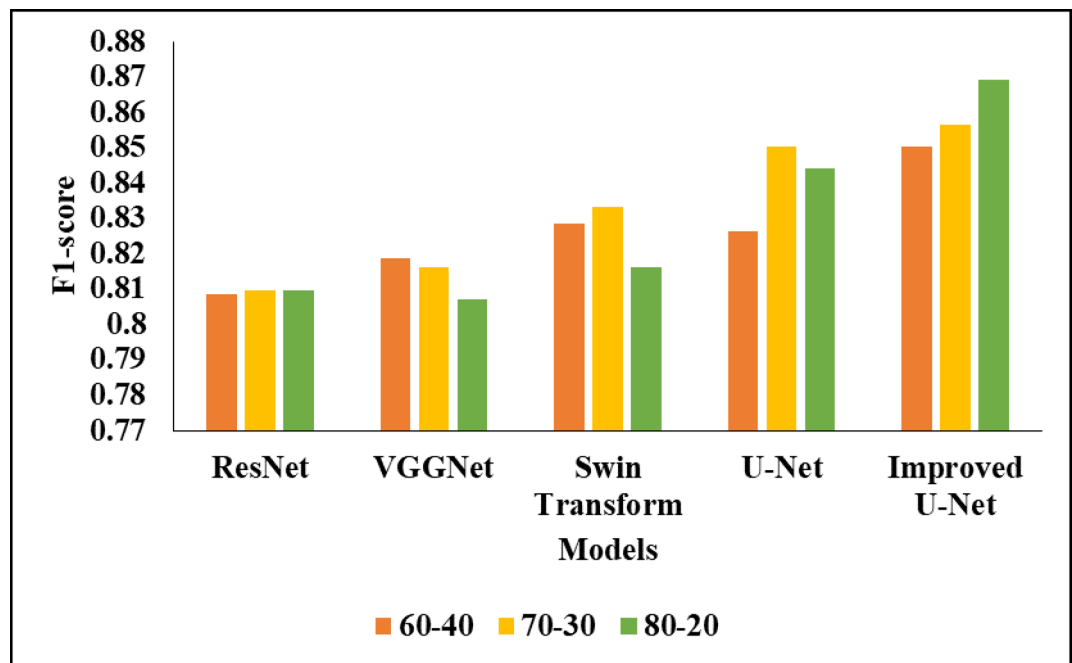


Fig. 6. F1-score analysis.

MLPmodel reaches an accuracy of 0.7355 ± 0.0028 0029, a precision value of 0.6453 ± 0.0047 , a recall value of 0.6850 ± 0.0045 , and an F1 score of 0.6646 ± 0.0038 respectively. AEmodel reaches the accuracy of 0.8438 ± 0.0064 0029, a precision value of 0.8004 ± 0.0095 , a recall value of 0.8254 ± 0.0072 , and an F1 score of 0.8127 ± 0.0081 , respectively. Then, CNNmodel reaches an accuracy of 0.9179 ± 0.0044 0029, a precision value of 0.9083 ± 0.0059 , a recall of worth 0.8917 ± 0.0057 , and the F1 score of 0.8992 ± 0.0053 respectively.

Quantitative segmentation results

We evaluated the segmentation performance of our U-Net-based model on a subset of the dataset with ground truth lesion annotations as can be seen in Table 8 based on the following metrics:

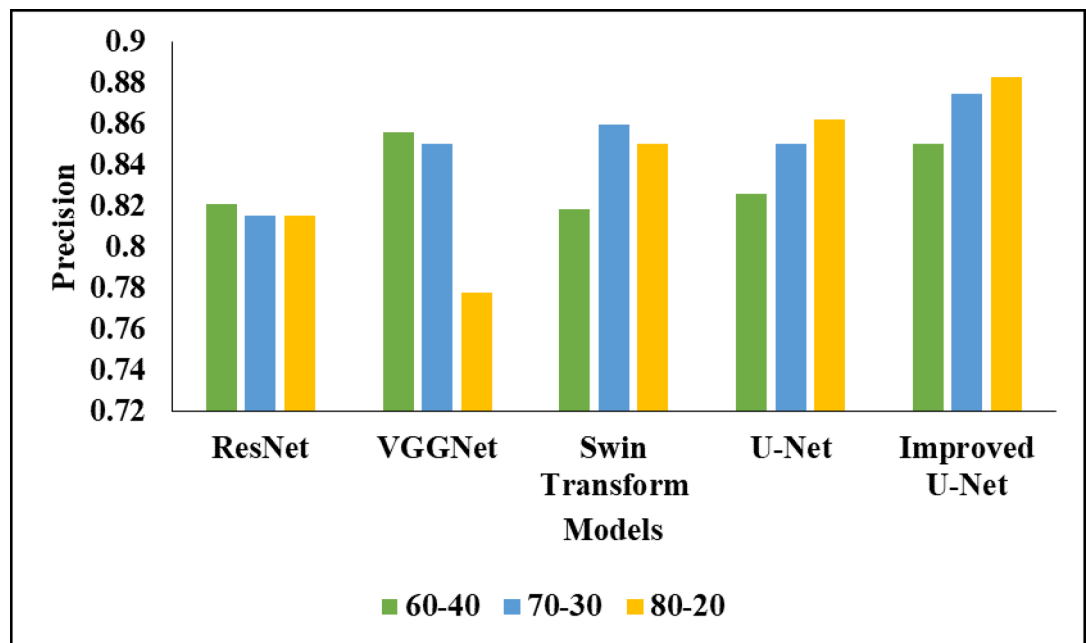


Fig. 7. Validation based on precision.

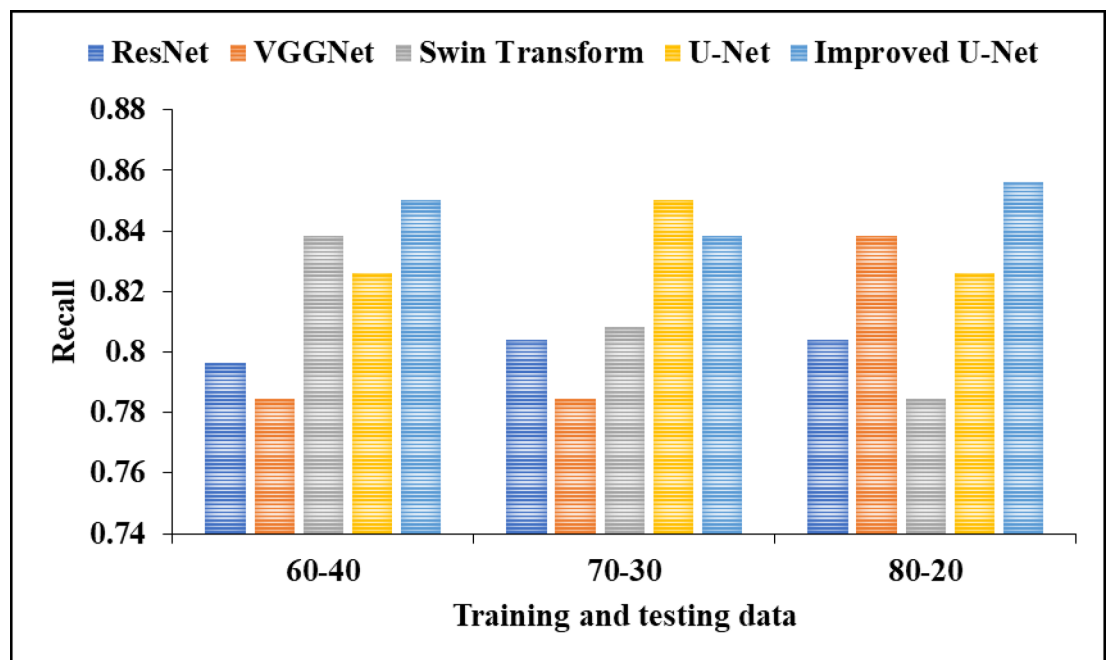


Fig. 8. Analysis of classifier in terms of recall.

- **IoU (Intersection over Union):** Measures the overlap between the predicted segmentation and the ground truth.
- **Dice score:** A harmonic mean of precision and recall, widely used in medical image segmentation.

These results demonstrate that the U-Net architecture achieves strong segmentation performance, supporting its use as an effective feature extractor in our pipeline.

Discussion

The study presented a model for the detection of Mpox. The proposed model would contribute to the efforts towards combating the disease using technology and improving the accuracy of Mpox illness diagnosis

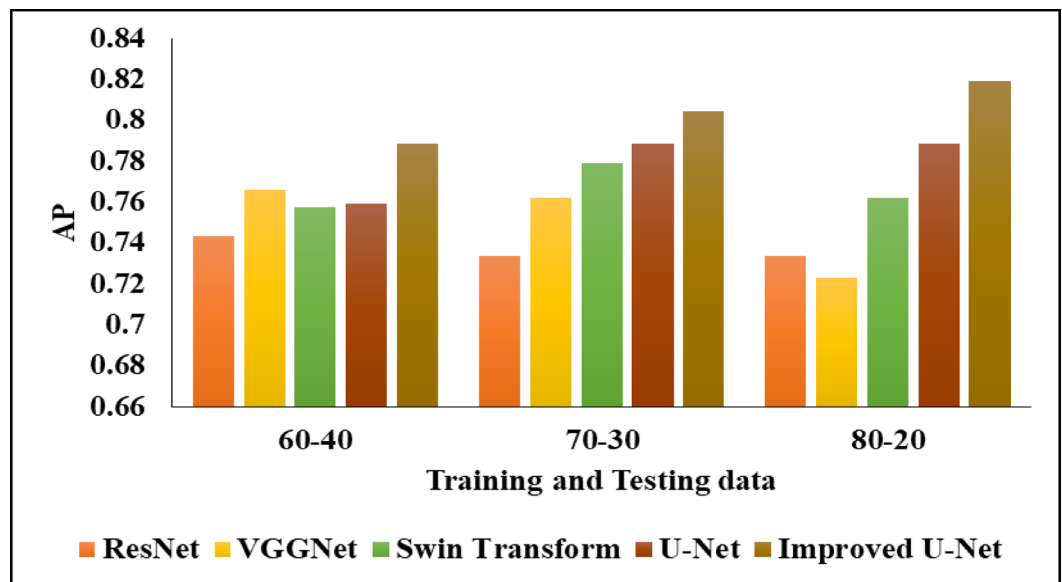


Fig. 9. Graphical Comparison in terms of AP.

Models	Accuracy	Precision	Recall	F1 score
LR	0.8864 ± 0.0028	0.8604 ± 0.0032	0.8748 ± 0.0041	0.8675 ± 0.0032
DNN	0.8977 ± 0.0033	0.8693 ± 0.0055	0.8890 ± 0.0029	0.8790 ± 0.0040
MLP	0.7197 ± 0.0025	0.6364 ± 0.0045	0.6730 ± 0.0039	0.6542 ± 0.0034
AE	0.8335 ± 0.0020	0.7870 ± 0.0029	0.8158 ± 0.0030	0.8011 ± 0.0024
CNN	0.9021 ± 0.0039	0.8941 ± 0.0047	0.8778 ± 0.0049	0.8859 ± 0.0046

Table 6. Analysis of DL based on Multiclass classification.

Models	Accuracy	Recall	Precision	F1 score
LR	0.8933 ± 0.0029	0.8795 ± 0.0037	0.8656 ± 0.0039	0.8725 ± 0.0035
DNN	0.8990 ± 0.0063	0.8833 ± 0.0079	0.8791 ± 0.0065	0.8812 ± 0.0069
MLP	0.7355 ± 0.0028	0.6850 ± 0.0045	0.6453 ± 0.0047	0.6646 ± 0.0038
AE	0.8438 ± 0.0064	0.8254 ± 0.0072	0.8004 ± 0.0095	0.8127 ± 0.0081
CNN	0.9179 ± 0.0044	0.8917 ± 0.0057	0.9083 ± 0.0059	0.8992 ± 0.0053

Table 7. Analysis of DL model based on Binary classification.

Metric	Mean + Std Dev.
IoU	0.76 ± 0.08
Dice score	0.85 ± 0.06

Table 8. Segmentation performance of the U-Net-based model on a subset of the dataset with ground truth lesion annotations.

by combining a hair removal process with an improved U-Net model. The modified U-Net model was subsequently used to diagnose Mpox disease with increased precision, making use of the hair-free images' superior segmentation capabilities. The study found that including the hair removal process in the diagnostic pipeline considerably improved performance metrics like accuracy, sensitivity, and specificity when compared to traditional models that did not include this preprocessing step. The study contributes to the field of medical imaging by proposing a more reliable and effective method for early Mpox detection, which can potentially improve diagnostic outcomes and patient care. Future work could explore the adaptation of this technique for other skin-related diseases and conditions, further broadening its applicability.

Conclusion and future scope

Leveraging state-of-the-art architectures and transfer learning strategies, we introduce the open-source MSLD dataset for lesion analysis and conduct an initial feasibility study. This work proposes a novel convolutional neural network (CNN)-based method for removing hair artifacts from input images, utilizing an encoder-decoder architecture proven effective in similar reconstruction tasks. A key highlight of our design is the integration of skip connections, which significantly enhance data retrieval. To extract and merge features, we employ pre-trained CNN models. These fused features are input into a custom classifier—an improved U-Net model enhanced with VGG-16—for Mpox detection. Despite the limitations of a small dataset, the encouraging outcomes from fivefold cross-validation suggest promising potential for AI-assisted preliminary analysis of Mpox. The MSLD dataset and our methods aim to support the development of remote, scalable analytic tools, which are particularly valuable in areas where traditional testing (like PCR or microscopy) is inaccessible. Furthermore, we are preparing to launch a web application prototype, allowing individuals to conduct at-home preliminary screenings for Mpox and seek early medical intervention when necessary. Our proposed methodology is efficient, cost-effective, and adaptable, enabling rapid deployment by healthcare providers without the need for specialized laboratory equipment. This opens possibilities for real-time patient screening, particularly for those presenting early symptoms of Mpox.

However, the study is not without limitations. The dataset size is relatively small and may not represent broader diversity across race, gender, and geographic location. Future work should include more heterogeneous samples to enhance the reliability of results across different populations. Additionally, while pre-training on ImageNet provided a strong starting point, training with a lesion-specific dataset could further improve model accuracy and generalizability. Lastly, because the dataset primarily comprises web-scraped images, it lacks critical metadata that could enrich the analysis. Expanding the dataset with well-annotated samples is essential for achieving robust and generalizable conclusions.

Data availability

Data will be provided upon sufficient request by corresponding author.

Received: 7 March 2025; Accepted: 2 June 2025

Published online: 01 July 2025

References

- Qiao, G., Song, H., Hou, S. & Jinyi, X. Enhancing literature review and understanding under global pandemic. *Risk Manag. Healthc. Policy* **16**, 143–158. <https://doi.org/10.2147/RMHP.S393293> (2023).
- Sookaromdee, P. & Wiwanitkit, V. N. Infected mother and mpox: The present concern. *J. Perinat. Med.* **51** (3), 437–437 (2023).
- Tiwari, A. et al. Mpox outbreak: Wastewater and environmental surveillance perspective. *Sci. Total Environ.* **856**, 159166 (2023).
- Sobral-Costas, T. G. et al. Human Mpox outbreak: Epidemiological data and therapeutic potential of topical Cidofovir in a prospective cohort study. *J. Am. Acad. Dermatology*. **88** (5), 1074–1082 (2023).
- Bengesi, S., Oladunni, T., Olusegun, R. & Audu, H. A machine learning-sentiment analysis on Mpox outbreak: an extensive dataset to show the polarity of public opinion from Twitter tweets. *IEEE Access* **11**, 11811–11826 (2023).
- Priyadarshini, I., Mohanty, P., Kumar, R. & Taniar, D. Mpox outbreak analysis: An extensive study using machine learning models and time series analysis. *Computers* **12** (2), 36 (2023).
- Long, B., Tan, F. & Newman, M. Forecasting the Mpox outbreak using ARIMA, prophet, neuralprophet, and LSTM models in the United States. *Forecasting* **5** (1), 127–137 (2023).
- Gao, J. et al. Mpox outbreaks in the context of the COVID-19 pandemic: Network and clustering analyses of global risks and modified SEIR prediction of epidemic trends. *Front. Public Health* **11**, 1052946 (2023).
- Jaradat, A. S. et al. Automated Mpox skin lesion detection using deep learning and transfer learning techniques. *Int. J. Environ. Res. Public Health* **20** (5), 4422 (2023).
- Meena, G., Mohbey, K. K., Kumar, S. & Lokesh, K. A hybrid deep learning approach for detecting sentiment polarities and knowledge graph representation on Mpox tweets. *Decis. Anal. J.* **7**, 100243 (2023).
- Li, X., Yang, X., Ma, Z. & Xue, J.-H. Deep metric learning for few-shot image classification: A review of recent developments. *Pattern Recogn.* **138**, 109381. <https://doi.org/10.1016/j.patcog.2023.109381> (2023).
- Yao, D. et al. Deep hybrid: Multi-graph neural network collaboration for hyperspectral image classification. *Def. Technol.* **23**, 164–176 (2023).
- Arco, J. E. et al. Uncertainty-driven ensembles of multi-scale deep architectures for image classification. *Inform. Fus.* **89**, 53–65 (2023).
- Oza, P., Sharma, P. & Patel, S. Deep ensemble transfer learning-based framework for mammographic image classification. *J. Supercomput.* **79** (7), 8048–8069 (2023).
- Alqahtani, T. M. Big data analytics with optimal deep learning model for medical image classification. *Comput. Syst. Sci. Eng.* **44** (2), 1433–1449 (2023).
- Gupta, K. & Bajaj, V. Deep learning models-based CT-scan image classification for automated screening of COVID-19. *Biomed. Signal Process. Control* **80**, 104268 (2023).
- Ugboaja, S. G. et al. Advanced diabetes prediction using supervised machine learning technique: Random forest. *Trop. J. Appl. Nat. Sci.* **2** (3), 1–14 (2024).
- Diao, S. et al. Deep multi-magnification similarity learning for histopathological image classification. *IEEE J. Biomed. Health Inform.* **27**(3), 1535–1545. <https://doi.org/10.1109/JBHI.2023.3237137> (2023).
- Iqbal, T. & Wani, M. A. Weighted ensemble model for image classification. *Int. J. Inform. Technol.* **15** (2), 557–564 (2023).
- Gao, L. et al. Research on image classification and retrieval using deep learning with attention mechanism on diaspora Chinese architectural heritage in Jiangmen. *China Build.* **13** (2), 275 (2023).
- UzunOzsahin, D., Mustapha, M. T., Uzun, B., Duwa, B. & Ozsahin, I. Computer-Aided Detection and Classification of Mpox and Chickenpox Lesion in Human Subjects Using Deep Learning Framework. *Diagnostics* **13**(2), 292 (2023).
- Uysal, F. Detection of Mpox disease from human skin images with a hybrid deep learning model. *Diagnostics* **13** (10), 1772 (2023).
- Lakshmi, M. & Das, R. Classification of Mpox images using LIME-enabled investigation of deep convolutional neural network. *Diagnostics* **13** (9), 1639 (2023).
- Bala, D. et al. MonkeyNet: A robust deep convolutional neural network for Mpox disease detection and classification. *Neural Netw.* **161**, 757–775 (2023).

25. Siva Shankar, G., Onyema, E. M., Kavin, B. P., Venkataramaiah Gude, B. V. V. & Prasad, S. Breast cancer diagnosis using virtualization and extreme learning algorithm based on deep feed forward networks. *Biomed. Eng. Comput. Biol.* <https://doi.org/10.1177/11795972241278907> (2024).
26. Sitaula, C. & Shahi, T. B. Mpox virus detection using pre-trained deep learning-based approaches. *J. Med. Syst.* **46** (11), 78 (2022).
27. Ahsan, M. M., Uddin, M. R., Farjana, M., Sakib, A. N., Momin, K. A., & Luna, S. A. Image Data collection and implementation of deep learning-based model in detecting Mpox disease using modified VGG16. arXiv preprint [arXiv:2206.01862](https://arxiv.org/abs/2206.01862). (2022).
28. Haque, M.E., Ahmed, M.R., Nila, R.S. and Islam, S. Classification of human Mpox disease using deep learning models and attention mechanisms. arXiv preprint [arXiv:2211.15459](https://arxiv.org/abs/2211.15459). (2022).
29. Alrusaini, O. A. Deep learning models for the detection of Monkeypox skin lesion on digital skin images. *Int. J. Adv. Comput. Sci. Appl.* <https://doi.org/10.14569/IJACSA.2023.0140170> (2023).
30. Altun, M. et al. Mpox detection using CNN with transfer learning. *Sensors* **23** (4), 1783 (2023).
31. Almufareh, M. F., Tehsin, S., Humayun, M. & Kausar, S. A transfer learning approach for clinical detection support of Mpox skin lesions. *Diagnostics* **13** (8), 1503 (2023).
32. Pramanik, R., Banerjee, B., Efimenko, G., Kaplun, D. & Sarkar, R. Mpox detection from skin lesion images using an amalgamation of CNN models aided with a Beta function-based normalization scheme. *Plos One.* **18** (4), e0281815 (2023).
33. Yasmin, F. et al. PoxNet22: A fine-tuned model for the classification of Mpox disease using transfer learning. *IEEE Access* **11**, 24053–24076 (2023).
34. Ariansyah, M. H., Winarno, S. & Sani, R. R. Mpox and measles detection using CNN with VGG-16 transfer learning. *J. Comput. Res. Innov.* **8** (1), 32–44 (2023).
35. Örenç, S., Acar, E. & Özerdem, M. S. Utilizing the ensemble of deep learning approaches to identify Monkeypox disease. *DÜMF Mühendislik Dergisi* <https://doi.org/10.24012/dumf.1199679> (2023).
36. Farouk, R. M., & Abdelaziz, M. A framework of deep learning for mpox image classification based on particle swarm optimization (PSO). *Tobacco Regulatory Science (TRS)* 470–485 (202).
37. Zebbari, D. A., Haron, H., Zeebaree, S. R. & Zeebaree, D. Q. Enhance the mammogram images for both segmentation and feature extraction using wavelet transform. In: *International Conference on Advanced Science and Engineering (ICOASE)* 100–105 (IEEE, 2019).
38. Moghaddam, S. H. A., Beirami, B. A. & M.Mokhtarzade, and A feature extraction method based on spectral segmentation and integration of hyperspectral images, international. *J. Appl. Earth Observ. Geoinf.* **89**, 102097 (2020).
39. Selvaraj, D. et al. Super learner model for classifying leukemia through gene expression monitoring. *Discover Oncol.* **15**, 499. <https://doi.org/10.1007/s12672-024-01337-x> (2024).
40. Aziz, R., Verma, C. & Srivastava, N. Dimension reduction methods for microarray data: A review. *AIMS. Bioeng.* **4**(1), 179–197 (2017).
41. Nolen, L. D. et al. Extended Human-to-Human transmission during a monkey pox outbreak in the Democratic Republic of Congo. *Emerg. Infect. Dis.* **22**, 1014–1021 (2016).
42. Imagedatagenerator. (Accessed on May 10, 2022). https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator, 2022.
43. Sreenivas Bhattiprolu. Data augmentation. (Accessed on May 10, 2022). <https://github.com/bnsreenu>, 2020.
44. Liu, G. et al. Image inpainting for irregular holes using partial convolutions. In *European Conference on Computer Vision (ECCV)* 85–100 (2018).
45. Dolz, J., Desrosiers, C. & Ayed, I. S. Ivd-net: Intervertebral disc localization and segmentation in MRI with a multi-modal unit. In *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging* 130–143 (Springer, 2018).
46. Lachinov, D., Vasiliev, E. & Turlapov, V. Glioma segmentation with cascaded unit. In *International MICCAI Brainlesion Workshop* 189–198 (Springer, 2018).
47. Hwang, H., Rehman, H. Z. U. & Lee, S. 3d U-net for skull stripping in brain MRI. *Appl. Sci.* **9**(3), 569 (2019).
48. Edeh, M. O. et al. Potential of internet of things and semantic web technologies in the health sector. *Nigerian J. Biotechnol.* **38** (2), 73–83. <https://doi.org/10.4314/njb.v38i2.8> (2021).
49. Mpox skin lesion dataset. Kaggle. <https://www.kaggle.com/datasets/nafin59/Mpox-skin-lesion-dataset>
50. Chaudhary, V., Lucky, L., Sable, H. & Bhalla, N. Interdisciplinary approach to Monkeypox prevention: Integrating nanobiosensors, nanovaccines, artificial intelligence, visual arts, and social sciences. *Small Struct.* <https://doi.org/10.1002/sstr.202400647> (2025).

Author contributions

Edeh Michael Onyema and Mueen Uddin: Conceptualization, Methodology, Writing-Original draft preparation, Edeh Michael Onyema, B.Gunapriya, Balasubramanian Prabhu Kavin: Data curation and Visualization: Tehseen Mazhar and Mueen Uddin: Writing-Reviewing and Editing, Priyan Malarvizhi Kumar and, Mamoon M. Saeed Investigation, Software. B.Gunapriya, Balasubramanian Prabhu Kavin: Validation, Data curation, Formal analysis. Tehseen Mazhar and Priyan Malarvizhi Kumar: resources, and analysis. All authors reviewed the manuscript.”

Funding

No.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Informed consent

Not Applicable.

Clinical trial number

Not Applicable.

Additional information

Correspondence and requests for materials should be addressed to T.M. or M.M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025