International Research Journal of Multidisciplinary Scope (IRJMS), 2025; 6(1):1420-1439 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2025.v06i01.02810

Revolutionizing Monkey Pox Diagnosis: A Cutting-Edge Deep Learning Pipeline for Advanced Lesion Segmentation and Classification Using CoAtNet

A Bamini[1]*, J Naveen Ananda Kumar[2], C Jayapratha[3], GB Govindaprabhu[4]

[1]Department of Computer Applications, The Standard Fireworks Rajaratnam College for Women, Sivakasi, Tamilnadu, India, [2]Tekinvaderz LLC, Florida, USA, [3]Department of Computer Science and Engineering, Karpaga Vinayaga College of Engineering and Technology, Madhuranthagam, Tamilnadu, India, [4]Madurai Kamaraj University (MKU), Madurai, Tamilnadu, India. *Corresponding Author's Email: drbaminia@gmail.com

Abstract

Monkeypox is a rapidly spreading virus which poses significant diagnostic challenges, due to its overlap with other viral illnesses. The availability of polymerase chain reaction (PCR) assays in resource-constrained environments is often hindered. In this work, deep learning is used to automate monkeypox detection by using skin lesion images. To enhance the quality and volume of the available dataset, advanced deep learning algorithms are combined with image augmentation techniques. By combining Flip, Mirror, rotate (FMR) random image augmentation with Automated White Balance Correction (AWBC), the detection model becomes more robust. With CoAtNet, a synthesis of convolutional neural networks and transformers, lesion images are captured both locally and globally. By using a hybrid architecture, the visual data can be analyzed more comprehensively and diagnostic errors may be reduced. This model is the most accurate, precise, recall, and F1-score with other existing models. An automated monkeypox detection system can be improved significantly by incorporating CNNs and transformers. Data augmentation strategies are also recommended as a way to enhance these models. The training dataset enhancing the model's ability to generalize to new cases. There is substantial promise in deep learning-based diagnostic tools for monkeypox, especially in areas with limited access to traditional laboratory testing. This work can support healthcare systems in combating the spread of this virus. Diagnostic gaps in such regions can be bridged with such systems, thereby contributing to the global health community's collective response.

Keywords: CoAtNet, Deep Learning Pipeline, FMR Random Image Augmentation, Monkey Pox Diagnosis, MSLD, Multi-scale Attention-guided Lesion Segmentation, White Balance Correction.

Introduction

As a zoonotic virus that is spreading rapidly and showing symptoms that overlap with other viruses, such as chickenpox, measles, and hand-foot-and-mouth disease (HFMD), monkeypox has resurfaced in recent months. A timely and accurate diagnosis is crucial to controlling outbreaks, especially in regions without adequate healthcare infrastructure. It is often difficult for resource-constrained environments to access traditional diagnostic methods, such as Polymerase Chain Reaction (PCR), because they require specialized laboratory settings and are sometimes time-consuming. An emerging solution to these challenges is the integration of artificial intelligence (AI) and deep learning into medical diagnostics. By applying deep learning models to skin lesion images, monkeypox has been demonstrated to be distinguished from other co-infections with similar symptoms through automated diagnosis. The widespread deployment and effectiveness of these models, however, are hindered by several challenges. In low-resource settings, real-time, deployable solutions are needed because there is a shortage of labeled monkeypox images, there is data imbalance, and it is difficult to distinguish between diseases with similar symptoms.

Several medical imaging tasks have shown remarkable success with deep learning, a subset of machine learning. These techniques could provide a quick, non-invasive screening method, assist healthcare workers in resource-limited settings, and possibly improve early detection rates for monkeypox. Such systems are, however, faced with

(Received 11th October 2024; Accepted 25th January 2025; Published 31st January 2025)

several challenges, including the limited availability of high-quality, diverse datasets, the need to distinguish monkeypox from visually similar conditions, as well as the need for models that can differentiate skin tones and lesion presentations accurately. A deep learning-based model for the early detection of monkeypox using skin lesion images is developed and enhanced in this paper. With FMR-based data augmentation and hybrid model architectures like CoAtNet, the work solves the critical challenges of data scarcity, model deployment and multi-modal image fusion. By applying advanced artificial intelligence techniques to monkeypox diagnosis. The initiative improves accuracy, robustness, and accessibility. Healthcare ethics are also discussed, especially privacy and security as work moves towards mobile-based, real-time diagnosis. Utilize deep learning to address these challenges of monkeypox detection on skin lesion images. The key aims in the work:

- Multimodal imaging has the potential to improve the understanding of the disease.
- The advanced techniques like GANs are employed to enhance and overcome data shortages.
- A COATNET architecture combines CNN and transformer models to capture local and global image features.
- Image quality is improved by automatic white balance, which reduces lighting variations.
- An integrated healthcare model that is widely applicable to multiple settings both on the desktop and in the community.

Among the evaluated models in the study, ResNet50, EfficientNet-B3, VGG16, and InceptionNetV3 (1). The researchers utilized Kaggle image datasets and applied data augmentation to enhance the sample size. The highest accuracy of 93% was achieved with the use of the EfficientNet-B3 model. Data sets with a variety of characteristics would produce more accurate and appropriate results. The work proposed using deep learning approaches to assist in early diagnosis of monkeypox. Monkeypox Skin Lesion Dataset (MSLD) was analyzed using a variety of pre-trained models, including EfficientNetB3, ResNet50V2, VGG16, DenseNet121, and InceptionV3 (2). By using pre-trained models in the framework, healthcare providers in resource-limited settings will be able to use it more easily without needing large amounts of training data and specialized computer resources. The framework proposed represents a potential strategy for improving monkeypox detection and management. With a 98% accuracy rate, the EfficientNetB3 model achieved the highest result.

According to the novel work, deep learning models can be used to automate the diagnostic process. A performance comparison between ResNet50, EfficientNetB3 and EfficientNetB7 is presented in this paper (3). A method is suggested for detecting monkey pox skin lesions early in the course of the disease. Even though a large dataset containing images from various countries of the world needs to be examined with other models, this study's results on a limited set of images are promising. The latest work suggested that a polymerase chain reaction (PCR) test could be used to diagnose monkeypox, but that the test takes time to determine the outcome (4). It would be beneficial if there was a non-clinical test that could help identify monkeypox in suspected patients. When sufficient training data is available, a variety of deep learning models can be used for this purpose. A sufficient amount of data has been added to existing datasets by extending them and adding new data. A pre-trained deep-learning model is then used to analyze this dataset, including ResNet50, EfficientNetB3, InceptionV3, and MobileNet2. These models have been compared for accuracy using this tool.

Using Transfer Learning, the work proposes a method of classifying monkeypox skin lesions from chickenpox and normal skin lesions (5). Using skin lesion image datasets from news reports, public health websites and case studies, five Transfer Learning models were trained: MobileNetv2, ResNet50, Inceptionv3, EfficientNetB5 and Xception. The trained models are compared in order to select the most effective model, which can be used in any application that requires quick, automated detection of monkeypox skin lesions. According to the results of the classification of monkeypox skin lesions, MobileNetv2 had the best model accuracy of 98.78%. The study used the Kaggle Monkeypox Image dataset, which is open source (6). A data replication method was first applied to the images to increase the sample dataset. In this study, five deep learning models are compared for detecting monkeypox virus

(DesNet121, ResNet50, Xception, EfficientNetB3, and EfficientNetB7). The accuracy, recall, precision, F1 score, and confusion matrix demonstrate the effectiveness of the methods. A DesNet121, ResNet50, Xception, EfficientNetB3, EfficientNetB7 method is 72% accurate, while a ResNet50 method is 73% accurate.

The accuracy metrics of the three methods were compared using previously trained CNN networks MobileNetV2, VGG16, and VGG19 on the Monkeypox Skin Image Dataset (7). Among the highest performance scores were 91.38 percent accuracy, 90.5% precision, 86.75 percent recall, and 88.25 percent f1 score obtained with MobileNetV2. According to the VGG16 method, the accuracy was 83.62%, while according to the VGG19 method, the accuracy was 78.45%. The research team classified monkeypox skin lesions with CNNs (8). The Grey Wolf Optimizer (GWO) improved CNN performance. The results improved accuracy, precision, recall, and AUC substantially. With the GWO optimizer, positively and negatively classified monkeypox cases were distinguished 95.3% more accurately. Monkeypox diagnosis and monitoring can be improved by using this method. Patient outcomes may be improved by earlier detection of lesions. Additionally, this work invented a vision transformer technique based on patches. A technique is used to detect monkeypox and chickenpox in human skin images (9). Medical technology can enhance the diagnostic process for these diseases. The ViT model is tested using a transfer learning approach for identifying subtle monkeypox and chickenpox patterns. This model's generalization capability was improved through carefully selected image augmentation techniques. It achieved 93% accuracy, precision, and recall in an evaluation of the patch-based ViT model.

The technology-based approach used in the work, it detects skin lesions automatically with sufficient training examples (10). With the help of MobileNetV2, which is a Fully Connected Convolutional Neural Network (FCCNN), monkeypox diagnosis has been improved. It is more accurate than classical machine learning approaches at identifying monkeypox cases. Several measures of effectiveness were assessed, including recall, precision, F score, and accuracy. The precision score is 0.95% with 0.99% accuracy, the recall is 1.0%, the F-score is 0.98%, and the F-score is 0.99%.

According the study aims to improve feature selection and classification methods during a pandemic using metaheuristic optimization (11). Extracting the necessary features involves deep learning and transfer learning. Feature extraction is conducted using the GoogleLeNet network. In addition, features are selected using a binary implementation of the dipper throated optimization algorithm. After that, features are labeled by using a decision tree classifier. The work describes a hybrid artificial intelligence system that detects monkeypox in skin images (12). For skin images, a dataset of open sources was used. The dataset is multi-classed, containing chickenpox, measles, monkeypox, and normal. There is an unbalanced distribution of classes in the original dataset. In order to overcome this imbalance, several data augmentation procedures were used. A number of deep learning models, including CSPDarkNet, InceptionV4, MnasNet, MobileNetV3, RepVGG, SE-ResNet and Xception, were used for monkeypox detection after these operations. By combining the two most effective deep learning models with the short-term memory (LSTM) model, a unique hybrid deep learning model was created for this study.

By analyzing skin lesion images, the work present an elegant, smart, and secure noninvasive, non-contact method to diagnose MPX (13). To classify skin lesions as MPXV positive or negative, deep learning techniques are employed. A Kaggle monkeypox skin lesion dataset (MSLD) and a monkeypox skin image dataset (MSID) are used as evaluation datasets. Using sensitivity, specificity, and balanced accuracy, the work evaluated multiple deep learning models. It has demonstrated its potential for wide-scale deployment in monkeypox detection with highly promising results. By using deep learning approaches and classification models proposed a model for detecting mpox (14). Toward this goal, the work compared five common pretrained deep learning models for detecting MPox (VGG19, VGG16, ResNet50, MobileNetV2, and EfficientNetB3). A score for F1 was calculated based on accuracy, recall, precision, and precision of the models. In the experiments, the MobileNetV2 model performed the best with 98.16% accuracy, 0.96 recall, 0.99 precision, and 98 F1-score. Moreover, validation of the model with different datasets showed that the highest accuracy of 0.94% was achieved using the MobileNetV2 model. With deep-learning methods, the work aim to detect monkeypox disease rapidly and safely through skin lesions (15). The optimization of hyperparameters was supported by deep-learning tools and transfer learning tools. By customizing the transfer learning model together with hyper parameters, a hybrid function learning model was developed. A custom model was implemented for MobileNetV3-s, EfficientNetV2, ResNET50, Vgg19, DenseNet121, and Xception. Among the metrics evaluated and compared in this study were AUC, accuracy, recall, loss, and F1-score. A hybrid MobileNetV3-s model with optimal F1-score, AUC, accuracy, and recall achieved the best score, with an average F1-score of 0.98. It has been recommended the proposed work that anyone who is suspected to have monkeypox infection should undergo testing (16). It is advisable to collect samples from skin lesions or exudates, swabs, and crusts if these are available. The laboratory confirms suspected cases with nucleic acid amplification testing, such as real-time or conventional polymerase chain reactions.

According to the work, the approach involves normalizing data and then linearly transforming it to reduce covariance between features (17). In addition, the concrete variance remains the same. Using PCA (Principal Component Analysis), the features are fused. The study proposes MXGBoost (Modified eXtreme Gradient Boosting) based on statistical loss functions to classify monkeypox and other viral samples (chickenpox samples, smallpox samples) in order to acquire effective prediction results. According to the work a coalesced CAD system is used to classify monkeypox using deep learning (18). Firstly, the given dataset images are pre-processed using a proposed fusion-based contrast enhancement method. A second step involves modifying and training six deep learning models: Vision Transformers (ViT), Shifted Windows (Swin) Transformers, ResNet-50, ResNet-101, EfficientNetV2, and ConvNeXt-V2. A third step involves acquiring and integrating the deep feature vectors from all the deep learning networks.

Based on images of skin lesions, in the work proposed deep learning to diagnose monkeypox. Five pre-trained deep neural networks were used to test the dataset: GoogLeNet, Places365-GoogLeNet, SqueezeNet, AlexNet, and ResNet-18 (19). Choosing the best parameters was done using hyper parameters. F1-score, AUC, and precision/precision are performance metrics considered. The highest accuracy was achieved by ResNet18, which obtained 99.49%. A validation accuracy of more than 95% was achieved with the modified models. The results demonstrate that deep learning models such as the one proposed based on ResNet-18 are deployable and crucial in the fight against monkeypox.

Research Gap

There are several gaps to fill in application of deep learning to monkeypox diagnosis, even though existing research has made significant strides:

Limited Multi-Disease Dataset Availability: Current systems mainly focus on single-disease datasets, restricting generalization. Similar symptoms include chickenpox, measles, and hand-foot-and-mouth disease.

Scarcity of Low-Volume Datasets: Deep learning models often struggle with insufficient data, particularly small datasets. When collecting large datasets is difficult in low-resource settings, accuracy and generalization are reduced.

Absence of White Balance Correction: There is a lack of attention to lighting conditions and image quality in many studies. AWBC (Automatic White Balance Correction) reduces color distortions and improves image clarity.

Lack of Unified Segmentation for Multi-Disease Datasets: Existing methods do not provide segmentation techniques for handling multiple diseases in a single dataset. Systems that differentiate among infections are difficult to build because of this gap.

Inaccurate Disease Prediction: Monkeypox cannot be distinguished from other visually similar diseases by many systems. As a result, dataset diversity and segmentation strategies are limited.

Methodology

This work introduces an innovative approach to monkeypox detection using advanced deep learning techniques. Multi-scale attention-guided lesion segmentation, sophisticated image preprocessing, and state-of-the-art classification models form the core of the methodology. The Monkeypox Skin Lesion Dataset (MSLD v2.0) includes images of monkeypox, chickenpox, measles, cowpox, hand-foot-and-mouth disease, and healthy skin. Using FMR (Flip, Mirror, Rotate) random image augmentation the work enhance dataset diversity and standardize image quality with Automated White Balance Correction. Multi-scale attention-guided segmentation of skin lesions is the key innovation. A basic CNN, Inception V3, and the primary classifier, CoatNet, which combines convolutional and transformer architectures, are compared. With this comprehensive approach, monkeypox detection will be significantly improved, potentially leading to rapid diagnosis and outbreak management. Figure 1 shows the flow of the proposed work.

Dataset

Monkeypox Skin Lesion Dataset (MSLD v2.0) is a comprehensive collection of skin lesion images designed specifically for studying and detecting monkeypox and other skin conditions. Diversity and relevance to the current global health context make this dataset especially valuable (20). The dataset is described in detail here. Six classes of images are included in MSLD v2.0:

- The monkeypox (MPox): Skin lesions characteristic of the infection.
- Chickenpox: Varicella-zoster virus, often mistaken for monkeypox.
- Measles: Virus-induced skin manifestations.
- Cowpox: Infections caused by cowpox virus.
- HFMD (Hand, Foot, and Mouth Disease): Common viral infection characterized by skin symptoms.
- Healthy: Normal, unaffected skin images.

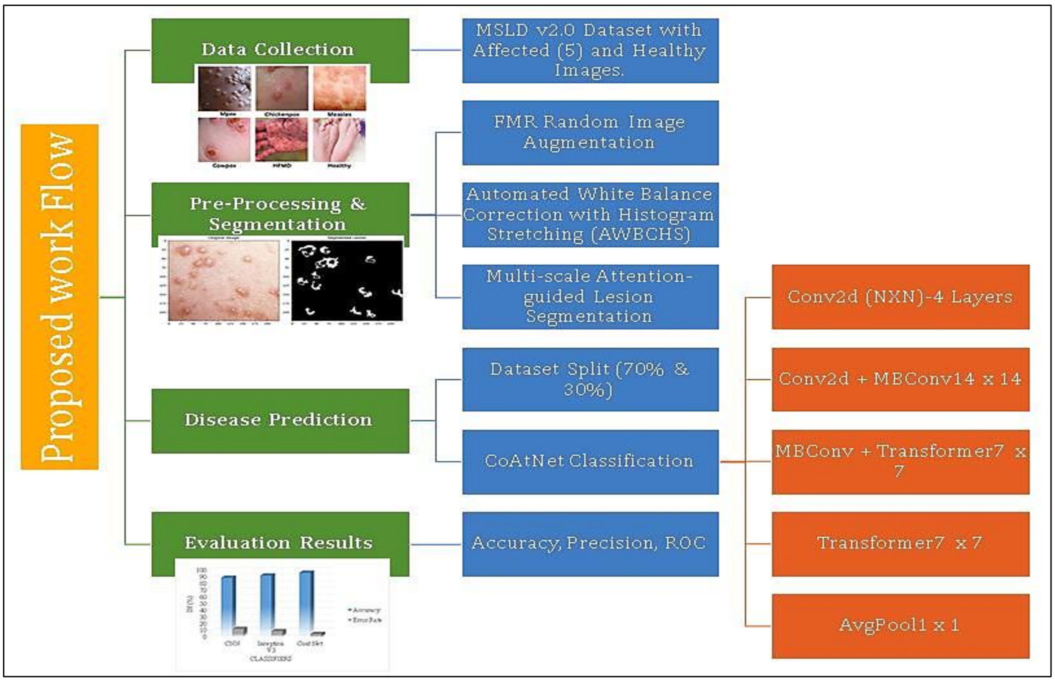
Proposed work Flow
Data Collection
Mpox
Chickenpox
Measles
Cowpox
HFMD
Healthy
MSLD v2.0 Dataset with Affected (5) and Healthy Images.
FMR Random Image Augmentation
Pre-Processing & Segmentation
Automated White Balance Correction with Histogram Stretching (AWBCHS)
Multi-scale Attention-guided Lesion Segmentation
Conv2d (NXN)-4 Layers
Conv2d + MBConv14 x 14
Dataset Split (70% & 30%)
Disease Prediction
CoAtNet Classification
MBConv + Transformer7 x 7
Evaluation Results
Accuracy, Precision, ROC
Transformer7 x 7
AvgPool1 x 1
CNN
Inception V3
CoatNet
CLASSIFIERS
Accuracy
Error Rate

Figure 1: Proposed Work Flow

Using this multi-class structure, the work can develop a diagnostic approach that not only detects monkeypox but also differentiates it from other visually similar skin conditions. The images in the dataset reflect the diverse presentation of these diseases across different populations based on skin tones, lesion stages, and imaging conditions. A variety of demographics and clinical settings is needed to train robust models. There are probably some imbalances in the dataset with certain conditions being rarer than others, even though the exact number is not specified. The use of stratified sampling and possibly class weighting or augmentation techniques to address imbalances is critical to the data handling and model training strategies. The sample dataset for the monkeypox is shown in Figure 2.

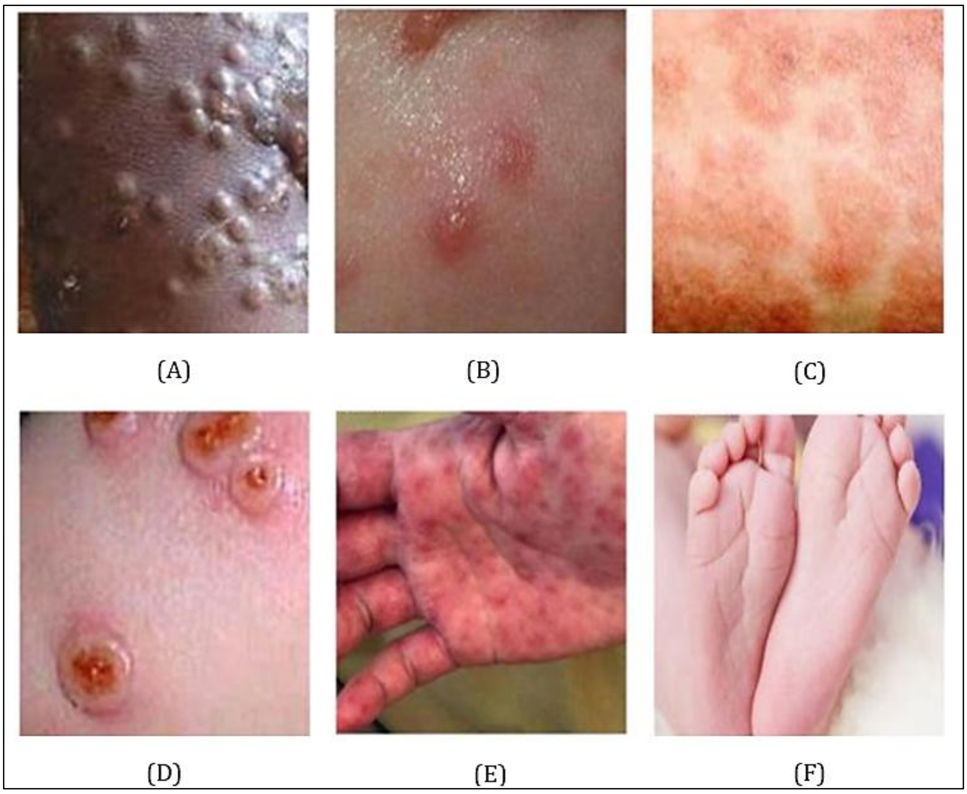

Figure 2: Sample Dataset. (A) Mpox, (B) Chickenpox, (C)Measles, (D) Cowpox, (E) HFMD, (F) Healthy

As in real-life clinical scenarios, MSLD v2.0 images vary in quality. It is possible that some images are high-resolution, well-lit photographs taken in controlled clinical environments, whereas others may be lower-quality snapshots taken by patients themselves. Image quality diversity underscores the importance of preprocessing steps like Automated White Balance Correction (AWBC). On the basis of this comprehensive and diverse dataset, it aims to develop a generalizable, accurate, and robust monkeypox detection system. Improved skin lesion classification may reduce misdiagnoses and improve diagnostic accuracy with multi-class datasets.

Multi-disease datasets allow the model to differentiate diseases with similar visual features, improving its clinical relevance. This approach has several limitations and challenges. In multi-disease datasets, certain conditions are often underrepresented, resulting in biased performance. Chickenpox and monkeypox are visually similar, but if subtle differences are not captured, misclassification is more likely. Furthermore, training a model to handle multiple classes increases computational requirements and may lead to longer training times and overfitting.

FMR Random Image Augmentation

In deep learning algorithms, such as CNNs, image augmentation is vital. This technique creates numerous variations of the training data to enlarge the dataset. To train thoroughly with limited data, augmentation is essential. The methods like flipping, mirroring, and rotation, input scenarios can be significantly expanded. Due of the exposure to various visual perspectives in such datasets, models trained on them are more robust and generalizable. When object orientation is not fixed, mirroring can provide symmetrical perspectives that can be especially useful. In this approach, overfitting is reduced, model performance is improved, and generalization to unseen data is enhanced. Nevertheless, these augmentations must be contextually appropriate for the task, as they may distort important features in some cases (e.g., medical images). As a result, flipping, mirroring, and rotating enhance the training set and enhance the model's robustness. The image augmentation using FMR is illustrated in Figure 3. The figure shows the flipped, rotated and mirrored image of the dataset.

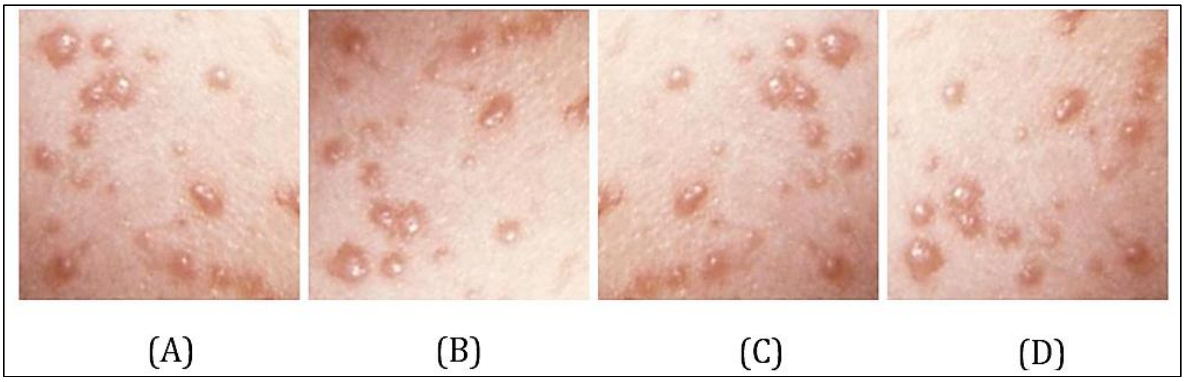

Figure 3: Image Augmentation. (A) Original, (B) Flipped, (C) Mirrored, (D) Rotated 90^0

```
Input: A set of input images, ImageSet = {I1, I2, ..., In}, Flip (Horizontal/Vertical): flip_horizontally, flip_vertically, mirror_axis (horizontal or vertical axis), rotation_range (e.g., from -180° to 180°)
Output: A set of augmented images, AugmentedImageSet
Algorithm FMR Random Image Augmentation
    • Initialize AugmentedImageSet = {} to store the augmented images.
    • Define augmentation parameters: flip_horizontally, flip_vertically, mirror_axis, and rotation_range.
    • For each image I in ImageSet:
    • If flip_horizontally:
    • I_flip_horiz = Flip(I, direction="horizontal")
    • Add I_flip_horiz to AugmentedImageSet
    • If flip_vertically:
    • I_flip_vert = Flip(I, direction="vertical")
    • Add I_flip_vert to AugmentedImageSet
    • If mirror_axis is defined:
    • I_mirror = Mirror(I, axis=mirror_axis)
    • Add I_mirror to AugmentedImageSet
    • Set angle = RandomValue(rotation_range)
    • I_rotated = Rotate (I, angle)
    • Add I_rotated to AugmentedImageSet
    • If flip_horizontally and mirror_axis:
    • I_combined1 = Rotate (Flip(Mirror(I, axis=mirror_axis), "horizontal"), angle)
    • Add I_combined1 to AugmentedImageSet
    • If flip_vertically and mirror_axis:
    • I_combined2 = Rotate (Flip(Mirror(I, axis=mirror_axis), "vertical"), angle)
    • Add I_combined2 to AugmentedImageSet
    • If flip_horizontally and flip_vertically:
    • I_combined3 = Rotate (Flip (Flip(I, "horizontal"), "vertical"), angle)
    • Add I_combined3 to AugmentedImageSet
    • Repeat the above steps for each image in the dataset.
    • Return the AugmentedImageSet.
End Algorithm
```

Algorithm 1: FMR Random Image Augmentation

Algorithm 1 shows the Image augmentation. Random image augmentation with FMR (Flip, Mirror, Rotate) addresses several key challenges in developing accurate monkeypox detection models. Using this method, the work can effectively mitigate the problem of data scarcity, a common problem when dealing with relatively rare diseases such as monkeypox. Multiple variations of each original image are created by flipping, mirroring, and rotating, to increase diversity without adding more real-world samples. The augmentation improves the models' generalizability by learning orientation-invariant features, as well as reducing the risk of overfitting.

Automated White Balance Correction with Histogram Stretching (AWBCHS)

The "White Lens Problem," a phenomenon where flash photography or specific lighting scenarios result in brilliant, white reflections on the skin, frequently hinders medical image processing, notably for skin lesion identification techniques. The presence of monkeypox affects skin lesion pictures negatively. Reasons to deal with this issue include:

- A white glare can obscure details in skin lesions, such as texture, borders, and color variation, hindering accurate diagnosis.
- Segmentation algorithms are hampered by bright spots interfering with the separation of lesion and healthy skin.
- A monkeypox lesion and surrounding skin can appear white because of reflections, making diagnosis easier.
- White lens problems may result in models relying on artifacts rather than actual lesion features.
- Some images may be affected while others are not in an inconsistent dataset.
- Mistaking a white glare for a lesion can result in an incorrect diagnosis.

The Automatic White Balance Correction (AWBC) is included in preprocessing. By reducing the effects of white lenses, the technique normalizes lighting and color conditions. By avoiding imaging conditions, deep learning models don't suffer artifacts. Monkeypox detection is therefore more accurate and reliable in real-world imaging. Inconsistent lighting conditions require correction of the lens' color cast. By correcting white balance and enhancing contrast, automated methods like Gray World Assumption and Histogram Stretching can significantly improve image quality. These techniques work and their key benefits are explained here.

Gray World Assumption

Using this method, it is assumed that an image should be neutral gray on average. There are times when certain color channels (Red, Green, and Blue) dominate an image, resulting in an off-balanced appearance when the lighting is poor or there is a color cast (e.g., white lens problem). In Gray World Assumption, each color channel (R, G, B) is scaled to equal intensity to achieve a neutral gray. By balancing the colors, any unnatural color dominance is removed and the overall white balance is improved.

Histogram Stretching

Using Histogram Stretching, the contrast and dynamic range of an image can be enhanced after the white balance has been corrected. By spreading the pixel intensity values across the entire range (0 to 255), details become more visible. A YUV color space is created by separating luminance (brightness) from chrominance (color) information from the image. Only the luminance channel (Y) is stretched in histogram stretching, which enhances brightness and contrast while largely leaving colors unchanged.

```
Input: image (RGB)
Output: enhanced_image (RGB)
Algorithm: AWBCHS
    • Load the image from file. image ← load_image("path_to_image")
    • Apply Gray World Assumption for white balance:
    • Calculate avgR ← mean_intensity(image[:, :, Red_Channel])
    • Calculate avgG ← mean_intensity(image[:, :, Green_Channel])
    • Calculate avgB ← mean_intensity(image[:, :, Blue_Channel])
    • Calculate avgGray ← (avgR + avgG + avgB) / 3
    • Scale Red channel: image[:,:,Red_Channel] ← clip(image[:, :, Red_Channel] * (avgGray / avgR), 0, 255)
    • Scale Green channel: image[:,:,Green_Channel] ← clip(image[:, :, Green_Channel] * (avgGray / avgG), 0, 255)
    • Scale Blue channel: image[:,:,Blue_Channel] ← clip(image[:, :, Blue_Channel] * (avgGray / avgB), 0, 255)
    • Apply Histogram Stretching:
```

- Convert image to YUV color space: img_yuv ← RGB_to_YUV(image)
- Perform histogram equalization on the Y channel: img_yuv[:, :, Y_Channel] ← equalize_histogram(img_yuv[:, :, Y_Channel])
- Convert the image back to RGB: enhanced_image ← YUV_to_RGB(img_yuv)
- Return or display the enhanced image. return enhanced_image

End Algorithm

Algorithm 2: Automated White Balance Correction

Algorithm 2 shows the automated white balance correction. Automated White Balance Correction begins by loading the image and applying the Gray World Assumption to correct white balance issues. To create a neutral gray, each channel's average intensity is calculated, then each channel's average intensity is adjusted to match that of the neutral gray. By neutralizing the color cast caused by improper lighting, color casts are removed. Following the white balance correction, the algorithm applies Histogram Stretching to enhance contrast. Image luminance channels are converted into YUV color space is shown in Figure 4. Improves contrast and brightness with histogram equalization. Color and contrast are improved by RGB conversion.

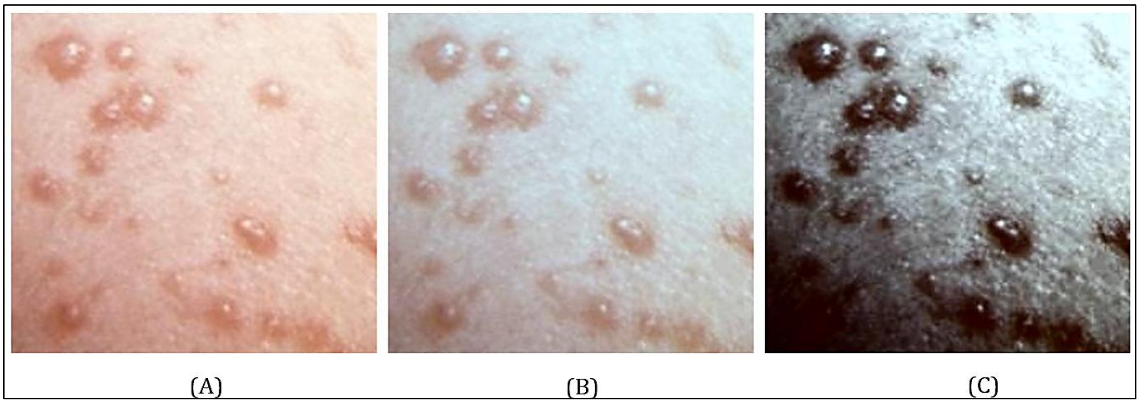

Figure 4: White Balancing and Histogram Stretching. (A) Original Image with White Lens Problem, (B) After AWB (Gray World), (C) After Histogram Stretching

Multi-Scale Attention-Guided Lesion Segmentation

It is essential to segment images to detect monkeypox and improve diagnostic efficiency. By segmenting monkeypox lesions from healthy tissue, the technique can analyze key monkeypox indicators. Input to subsequent algorithms is cleaner and more targeted with this technique. Quantitative data on lesions, such as size, shape, and distribution, can also be analyzed. Using these metrics, monkeypox must be distinguished from other diseases. It also reduces false positives by distinguishing lesion boundaries from irregularities on the skin and image artifacts. Different lighting and skin tones can be accounted for by segmentation. Various skin lesion images need to be standardized. In deep learning, monkeypox is more accurately detected.

This work, propose an integrated approach for segmenting lesions in medical images by integrating multiscale attention-guided techniques. Data representation is simplified by converting grayscale images to grayscale. CLAHE (Contrast Limited Adaptive Histogram Equalization) boosts the contrast. It improves contrast in localized areas to distinguish subtle lesions. The method filters lesions using multi-scales after preprocessing. A Gaussian blur with different kernel sizes is applied to enhance multi-scale features in the image. This method allows robust representation of features, which is vital to detecting lesions accurately.

A mechanism utilizing adaptive thresholding follows in order to detect potential lesion areas. Adaptive thresholding isolates regions of interest by creating binary attention maps with substantial intensity variations. Edge detection is completed using the Canny edge detector for accurate segmentation. The segmented results are refined using post-processing techniques. Using morphological operations, such as closing, small gaps can be filled and edges smoothed. While contour area filtering can reduce small noises. This process produces an accurate and noise-free segmented image. Multi-scale attention-guided lesion segmentation is effective when compared with original medical images. High precision lesion segmentation can be achieved using multi-scale and attention-guided techniques.

```
Input: A medical image (e.g., MRI or CT scan) in color format.
Output: Segmented lesion image.
Algorithm: Multi-Scale Attention-Guided Lesion Segmentation
   //Preprocessing
        • gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
        • Enhance contrast using CLAHE
        • Initialize CLAHE with a clip limit of 2.0 and tile grid size of (8, 8): clahe =
          cv2.createCLAHE(clipLimit=2.0, tileGridSize= (8, 8))
        • Apply CLAHE to the grayscale image: enhanced = clahe.apply(gray)
    // Multi-Scale Filtering
        • Apply Gaussian blur with different kernel sizes:
        • Small scale blur: blur_small = cv2.GaussianBlur(enhanced, (3, 3), 0)
        • Medium scale blur: blur_medium = cv2.GaussianBlur(enhanced, (5, 5), 0)
        • Large scale blur: blur_large = cv2.GaussianBlur(enhanced, (7, 7), 0)
        • Combine the blurred images using weighted addition:
        • combined = cv2.addWeighted(blur_small, 0.5, blur_medium, 0.3, 0)
        • combined = cv2.addWeighted(combined, 0.8, blur_large, 0.2, 0)
    //Generate an attention map:
        • attention_map = cv2.adaptiveThreshold(combined, 255, cv2.ADAPTIVE_THRESH_GAUSSIAN_C,
          cv2.THRESH_BINARY, 29, 15)
        • edges = cv2.Canny(attention_map, 100, 200)
        • Define a kernel: kernel = np.ones((5, 5), np.uint8)
        • Apply morphological closing: closing = cv2.morphologyEx(edges, cv2.MORPH_CLOSE, kernel)
        • Find contours: contours, _ = cv2.findContours(closing, cv2.RETR_EXTERNAL,
          cv2.CHAIN_APPROX_SIMPLE)
       // Remove small contours
        • for contour in contours:
        • if cv2.contourArea(contour) < 100:
        • cv2.drawContours(closing, [contour], -1, 0, -1)
        • Return the final segmented image:
        • return closing
End Algorithm
```

Algorithm 3: Lesion Segmentation using Multi-Scale Attention Guide

Following are the key steps in the algorithm:

- A grayscale conversion is followed by Contrast Limited Adaptive Histogram Equalization (CLAHE). For preprocessing, the contrast between lesions and healthy skin is enhanced.
- Various kernel sizes are used in Gaussian blurring to achieve multi-scale effects. This technique, lesion features can be captured at multiple levels of detail.
- By combining blurred images with weighted additions, fine details can be preserved.
- An adaptive thresholding technique generates an attention map. This reduces background noise and irrelevant features by focusing the algorithm on lesion-prone areas.
- Lesion boundaries are determined using a method called Canny edge detection. Using this method, one can precisely define the contours of an abnormality of the skin.
- After segmentation, morphological operations refine the results. To close small gaps between detected lesions, an operation is performed. The contour area also eliminates false positives and noise.

With multiple scales of lesion features, the algorithm is robust to changes in appearance and size. The mechanism focuses on relevant images, increasing accuracy and efficiency. Monkeypox can be more accurately detected using this segmentation method. The segmentation results are illustrated in Figure 5. Using segmented images, classification algorithms or quantitative assessments of lesion characteristics can be performed. Algorithm 3 shows the Lesion Segmentation using Multi-Scale Attention Guide. Figure 5(A) shows the Original Image and Figure 5(B) shows the segmented Lesion.

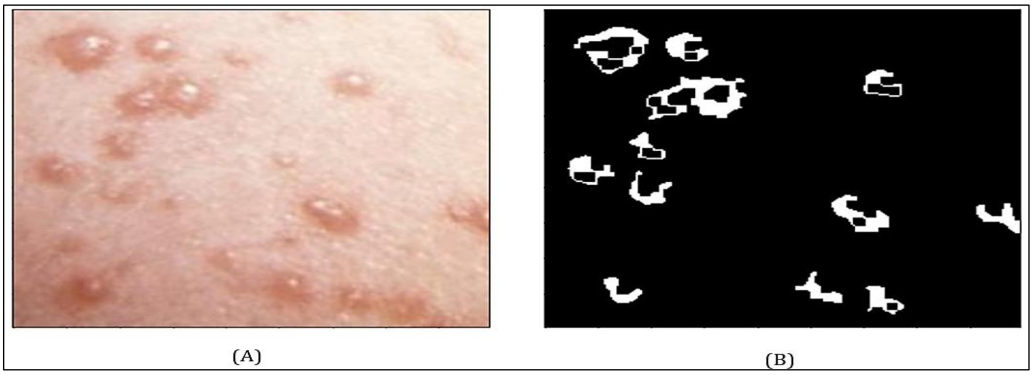

Figure 5: Segmentation of Lesion. (A) Original Image, (B) Segmented Lesion

Disease Prediction

For disease forecasting, the work applied and compared three advanced deep learning tools: a simple CNN, Inception V3, and COATNet. The work trained each model on a specially prepared dataset to effectively sort images of skin conditions.

Dataset Split

The work initially drew upon the Monkeypox Skin Lesion Dataset, version 2.0, which includes various images of skin rashes such as monkeypox, chickenpox, measles, cowpox, and hand-foot-and-mouth disease. To guarantee a thorough and reliable analysis, the data was meticulously divided into subsets for a robust model evaluation. Dataset converted 70% of the dataset is the training set. Test set accounts for 30% of the remaining 30%. A stratified sampling technique was used to ensure fair data collection. Each class has the same number of samples in training and test sets. The result is a well-representative class system that prevents bias. The data is split with a random seed for reproducibility. The same steps can be used if someone else wants to do the same experiment in the future.

CoAtNet

Using CoAtNet, combines the strength of convolutional neural networks (CNNs) with the strength of transformers. The image classification models can be enhanced by combining both strengths. With this architecture, local patterns are captured with convolutional operations, while global context is captured with attention mechanisms. In the CoAtNet system, the core principle is combining CNNs with Transformers to perform depthwise convolutions. This combination of approaches is motivated by a number of strengths:

Convolutional Layers: An inductive bias can be captured in a convolutional layer, which can result in better generalization in scenarios with limited data. Local spatial patterns are used as inputs in convolutional layers to capture local spatial patterns in data. An edges and textures are detected by applying filters to regions of the image called receptive fields. By concentrating on localized features, convolutional layers are particularly well suited for extracting essential information from small patches of an image.

Depth Wise Convolution: An aggregated field is formed by applying a fixed kernel. Based on input features, $X \in R^{H \times W \times CX}$, where H is the height, W is the width, and C is the number of channels, each channel c is convolutioned with a different filter K_c, resulting in an output feature map Y. The depthwise convolution at position (i,j) for channel c is given by:

$$Y_{i,j,c} = \sum_{m=1}^{k_H} \sum_{n=1}^{k_W} K_{m,n,c} \cdot X_{i+m,j+n,c} \qquad [1]$$

Where, k_H and k_W are two dimensions to a convolutional kernel: its height and width. $K_{m,n,c\_}$ the kernel value for channel c, at position (m,n).

In this operation, local spatial information is stored by aggregating features within a small, localized area of the image (referred to as the receptive field) in order to capture local spatial information. The reason why depthwise convolutions are efficient is that each channel's filter operates independently, so fewer computations are required than with traditional convolutions. Although depthwise convolutions are useful for capturing local context within an image, they may not be as effective at capturing long-range, image-wide dependencies.

The inductive bias of these layers is also introduced as they are constructed with the assumption of the data's structure (i.e., that local features are important in the data). Convolutional networks have the advantage of being efficient and able to generalize well, especially in instances where there are limited data sets. Essentially, this means that they are able to learn from fewer examples because their structure is ideal for analyzing local patterns, so they can learn from fewer examples.

Self-Attention Layers: The self-attention layer is capable of modeling global dependencies and has a higher capacity than other layers, which makes it a great choice when it comes to dealing with large datasets. Alternatively, self-attention layers are designed to handle global dependencies more intelligently. A spatial relationship is analyzed across all spatial positions rather than just within a single image instead of focusing on local patterns. It detects long-range dependencies using pairwise relationships between any two points. A large and complex dataset requires this to understand better. Furthermore, the model is capable of handling information and modeling complicated patterns, making it particularly useful when scaling up. This combination improves generalization and scalability of CoAtNet by handling various data sizes. In CoAtNet, attention mechanisms are combined with convolutional layers. As a result of this observation, depthwise convolutions and self-attention can be seen as complementary operations that can be used to process spatial information, as follows:

To unify these two processes, CoAtNet uses relative positional embeddings to compute attention, and in addition to that, it also incorporates convolutional kernels into the computation of self-attention, thereby combining both operations. With this approach, the model can benefit from both translation equivariance of convolution and attention based models while maintaining the benefits of both.

Input: Image dataset D= {(x1, y1), (x2, y2)..,(xn,yn)} where xi is an image and yi is the corresponding label, Number of stages S0,S1,S2,S3,S4 for CoAtNet, Number of layers in each stage L0,L1,L2,L3,L4. Model hyper parameters (learning rate, batch size, epochs, etc.).

Output: Trained CoAtNet model with image classification predictions.

Algorithm: CoAtNet for Image Classification

- ■ Initialize the CoAtNet architecture with the following components:
    - A convolutional stem in Stage S0S_0S0 to process input images and extract low-level features.
    - Subsequent stages S1, S2 with depth wise convolutional layers (MBConv) to capture local spatial features.
    - Later stages S3, S4 with self-attention layers to capture long-range dependencies and global context.
    - Employ relative positional encoding in the attention layers to retain translation equivariance.
- ■ Initialize model parameters θ (weights for convolution and attention layers).
- ■ For each input image x∈:
- Stage S0: Convolutional Stem
    - Apply a series of convolutional operations to the input image: f0=ConvStem(x)
    - Down-sample the image to reduce spatial dimensions while increasing the number of channels.
- Stage S1, S2: Depth wise Convolution Layers
    - For each layer l∈{1,2}: fl+1=DepthwiseConv(fl)
    - Use MBConv blocks with squeeze-and-excitation (SE) modules to extract local patterns, reduce spatial size, and increase channel depth.
- Stage S3, S4 : Self-Attention Layers

- For each layer l ∈ {3,4}: fl+1=RelativeAttention(fl)
- In the self-attention mechanism, compute pairwise relationships between all spatial positions: $Attention(q,k,v) = Softmax(qkTdk)$
- where q, k, and v are the query, key, and value matrices from the input features, and dk is the dimension of the key.
- Include relative positional encoding to incorporate spatial information into the attention mechanism.

- **Final Stage: Global Pooling**
  - Apply global average pooling to the final feature map to obtain a fixed-size representation: fglobal=GlobalAveragePooling(f4)
- Feed the globally pooled features into a fully connected layer to produce logits for the classification task: ypred=Softmax(Wfglobal+b) where W and b are the weights and biases of the fully connected layer.
- Compute the loss L(ypred,ytrue) using cross-entropy loss between the predicted class probabilities and the true labels ytrue.
- Compute gradients of the loss with respect to model parameters θ (including both convolution and attention layers) using backpropagation:$\partial L/\partial \theta$
- Update the model parameters θ using an optimizer (e.g., AdamW) based on the computed gradients: θ←θ−η∂L∂ where η is the learning rate.
- Repeat Steps 2-4 for each batch of images in the dataset over multiple epochs.
- Apply data augmentation (e.g., RandAugment, MixUp) and regularization techniques (e.g., stochastic depth, weight decay) to improve generalization.
- Once the model is trained, use the forward pass (Steps 2-3) on unseen test images to generate predictions: ytest=Softmax(Wfglobal)
- Classify the image based on the predicted class with the highest probability.
- Evaluate the trained model on a validation/test set using accuracy or other relevant metrics (e.g., top-1 or top-5 accuracy).
- Fine-tune or further train the model if necessary based on the evaluation results.

End Algorithm

Algorithm 4: CoAtNet for Image Classification

Algorithm 4 shows the COATNET. The CoAtNet network is structured in a similar way to traditional ConvNet networks with multiple stages. An architecture is divided into multiple stages in which the resolution of the input feature maps is gradually reduced over the course of the design:

- The S0 stage includes a convolutional stem, which performs the initial down sampling of the image as well as the extraction of its features.
- During the following stages, S1 to S4, the depth wise convolutions are alternated with attention blocks (in the earlier stages) as well as the attention blocks themselves. The spatial resolution decreases at each stage, as well as the number of channels increases as well.

Based on the MBConv structure, the convolution blocks that will be used are based on depth wise separable convolutions, which works well in terms of efficiency. The attention blocks incorporate a mechanism that allows relative attention to be determined, and this mechanism scales effectively as input size increases. Using relative self-attention, CoAtNet enhances the standard self-attention mechanism by incorporating information about the relative positions of image patches into its attention mechanism, which is an enhanced version of the standard self-attention mechanism. The relative encoding of the image patches enables the attention mechanism to be able to take into account the relative spatial relationships between them, making it more efficient when capturing both local and global dependencies within an image patch. In summary, the following steps can be taken to complete the process:

- Taking each pixel from the feature map into account, CoAtNet computes the pairwise attention between that pixel and every other pixel on the feature map.
- The attention score incorporates relative positional information into the model, which is essential to maintain translation equivariance, one of the desirable properties of convolutional models.

A significant benefit that this mechanism provides is that it improves the ability of the model to generalize, especially in tasks where spatial patterns repeat across multiple parts of an image. A basic description of the process involved in training and using CoAtNet for image classification can be found in this algorithm. The course covers the initialization of the hybrid architecture, a forward pass through both the convolution layer and the attention layer, as well as backpropagation for parameter updates and inference for making predictions based on the data.To ensure reliability in diverse clinical settings, CoAtNet must be evaluated for its robustness to variations in image quality, skin tone, and lesion presentation. Simulate real-world image quality discrepancies with simulated datasets with controlled resolutions, lighting conditions, and noise. Datasets with diverse demographic representations enable differentiating performance across skin tones. The lesion presentations should vary in size, shape, color, and stage. Measure the resilience of the model by measuring its robustness metrics. Additionally, data augmentation methods like adding noise and simulated lighting changes during training, as well as preprocessing methods like Automated White Balance Correction, help ensure the model remains high performing.

In CoAtNet, depthwise convolutions and self-attention combine to create complexity. Due to pairwise token interactions, the convolutional layers contribute complexity proportional to the spatial dimensions and channels of the input. The result is an overall complexity that balances local feature extraction with global context modeling:

$$o_{CoATNet} = \sum \quad K_H * K_W * C_{in} * H * \ W + \sum \quad (N^2 * D) \qquad [2]$$

Where $\mathrm{K_H}$, $\mathrm{K_W}$ are Kernel dimensions, $\mathrm{K_H}$ input channels, H, W Saptial dimensions, N number of tokens, −Embedding Size.

Results and Discussion

The predictive ability of machine learning models to identify monkeypox is typically measured using a variety of metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The following metrics help determine how well a model can distinguish between monkeypox cases and non-cases from input features, which may include clinical information, demographic information, and possibly image data for visual diagnosis (such as skin lesions), to determine whether it can make a successful prediction model for monkeypox.

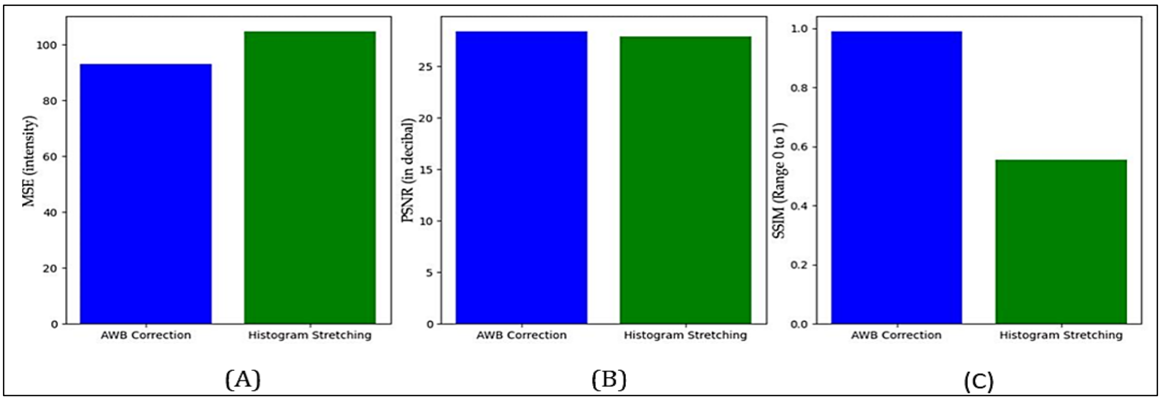

Figure 6: Pre-Processing Metrics for Different Values. (A) Mean Squared Error (MSE), (B) Peak Signal-to-Noise Ratio (PSNR), (C) Structural Similarity Index (SSIM)

With three key image quality metrics: Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM), two image correction techniques are compared as shown in Figure 6, AWB Correction (Auto White Balance Correction) and Histogram Stretching. A lower MSE performs better with AWB Correction than with Histogram Stretching, which exhibits more distortion than a higher MSE. The PSNR for both techniques is nearly identical, with AWB Correction slightly outperforming, reflecting a marginally higher signal-to-noise ratio. In terms of SSIM, which measures perceived image quality, AWB Correction yields a significantly higher score, indicating that it retains structural similarity to the original image better than Histogram Stretching. In terms of maintaining image quality, AWB Correction appears to outperform Histogram Stretching. Figure 6 shows the Pre-Processing Metrics.

Table 1: Performance Evaluation

Classifiers	Accuracy	Error Rate	Precision	Recall	F1-Score
CNN	87.96	12.04	84.87	89.32	87.0381583
Inception V3	91.25	8.75	89.43	92.44	90.9100918
Coat Net	95.42	4.58	92.56	95.74	94.1231482

Table 1 shows the performance evaluation. The Comparison of accuracy, error rate, precision, recall, and F1-Score of CNN, Inception V3, and Coat Net is shown in Table 1. An F1-Score of 87.04 means CNN's accuracy is 87.96%, with an error rate of 12.04%, precision is 84.87%, recall is 89.32%, and recall is 89.32%. As a result, Inception V3's accuracy is 91.25% and error rate is 8.75%. A higher F1-Score of 90.91% can be achieved with greater precision (89.43%) and recall (92.44%). Among all models, Coat Net performs the best, with an accuracy of 95.42% and an error rate of 4.58%. F1-Score of 94.12 indicates superior performance in balancing precision and recall with 92.56% precision and 95.74% recall. The representation of accuracy and error rate is illustrated based on the evaluation as shown in Figure 7.

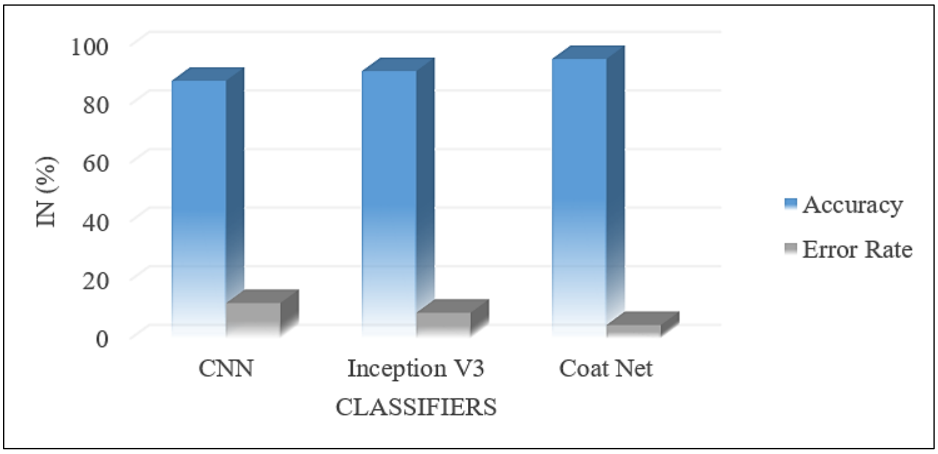

Figure 7: Accuracy and Error Rate

This machine's high level of accuracy and its high F1-score suggest that it offers a good balance between precision and recall as shown in Figure 8, which makes it ideal for medical diagnosis tasks that require a balance between precision and recall, such as those involving positive and negative diagnoses, which can result in serious consequences (21). As a result, Inception V3 shows marked improvements when compared to the basic CNN model, which is in line with the more sophisticated architecture that is used in the software, which makes it possible to achieve such improvements. As a result, the performance difference between Inception V3 and CoatNet is notable, indicating that the hybrid architecture of CoatNet (which combines CNNs with Transformers) would be particularly well suited to this task due to the hybrid architecture of CoatNet.

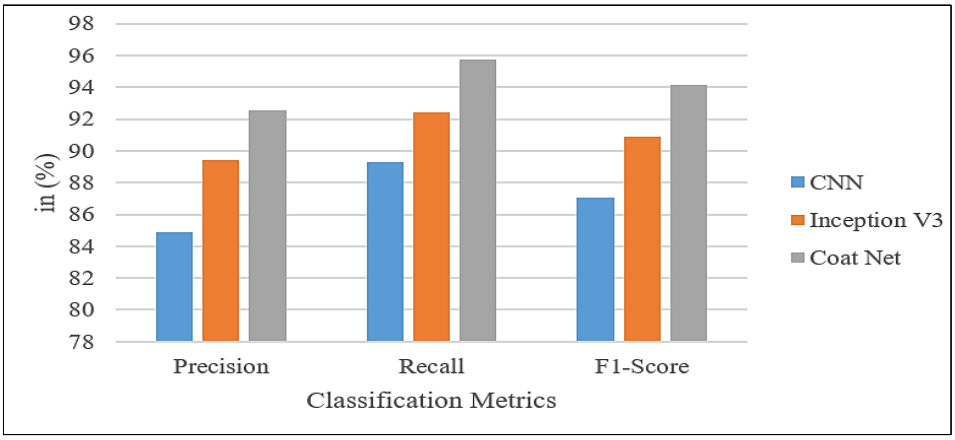

Figure 8: Precision, Recall F1-Score

As good as the basic CNN is, it has a significant lag behind the more advanced architectures, even though it performs reasonably well. The results of this study indicate the importance of using state-of-the-art models for complex image classification tasks, such as the detection of monkeypox lesions, which are exceedingly complex (22). There is no question that the results of this study support the choice of CoatNet as the primary model for this study, as it shows superior performance when it comes to identifying monkeypox lesions while minimizing errors and producing accurate results. Compared to other well-established models such as Inception V3, hybrid architectures are significantly better in terms of medical image analysis tasks than other well-established models. Figure 9 shows the ROC curve.

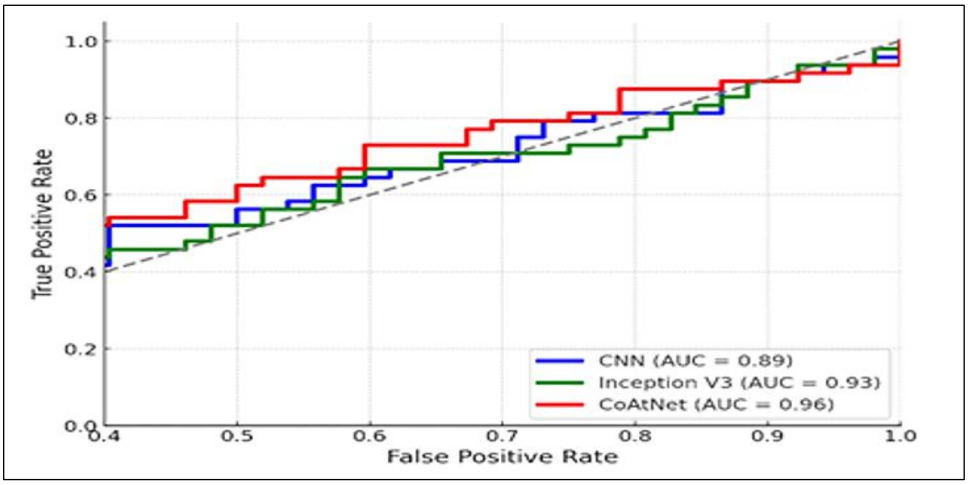

Figure 9: ROC Curve

A transparent and understandable decision-making process is essential for CoAtNet's adoption in clinical settings. A technique called Grad-CAM (Gradient-weighted Class Activation Mapping) can be used to visualize the regions of an image that the model focuses on when making predictions. Additionally, saliency maps and Layer-wise Relevance Propagation (LRP) can highlight how specific patterns in skin lesions contribute to diagnosis. Further understanding can be gained by revealing global dependencies in the transformer layers. SHAP (SHapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) can also elucidate input features' contribution. Integrating models into clinical workflows can be enhanced through explainability dashboards or real-time decision trails.

Models using AI for disease diagnosis, such as CoAtNet, can be biased due to underrepresentation of certain skin tones, lesions, age groups, or geographic regions. This bias can disproportionately impact marginalized communities and reduce the model's generalizability. Relying on limited datasets may worsen health disparities by favoring more prominently represented conditions or demographics. Ethical concerns include misdiagnosis, loss of trust in AI systems, and increased healthcare inequalities. To address these issues, it is crucial to use diverse, representative datasets and incorporate fairness measures during model training. Transparent reporting of model limitations and ongoing clinical oversight are also essential to ensure ethical standards and equitable healthcare delivery.

The proposed multi-scale attention-guided lesion segmentation technique and the CoatNet-based classification approach that the work propose for monkeypox detection represent significant advances over the existing systems in the literature that have been used so far. First of all, with FMR (Flip, Mirror, Rotate) Random Image Augmentation, the work is addressing limitations in the dataset size that have been seen in previous studies. As reported (1), augmented the sample size using data augmentation, but did not specify their methodology. An improved generalization of the model may result from the process of providing a more robust augmentation strategy.

The preprocessing pipeline has been enhanced with AWBC (Automated White Balance Correction), a new addition not found in most existing studies. This technique standardizes image quality across diverse datasets by addressing the white lens problem. The work (14) demonstrated, variability in image quality was a limiting factor in their studies. This applies particularly to studies with variable image quality. As far as detecting lesion boundaries is concerned, the multi-scale attention-guided lesion segmentation algorithm is clearly a significant advance over traditional segmentation methods. The work (8), which focused solely on classification without explicitly segmenting lesion areas, this approach isolates lesion areas with greater precision and accuracy. During the segmentation step, the model is enhanced in its ability to focus on relevant features, which could lead to an increase in the accuracy of the classification.

As far as performance is concerned, the CoatNet model performs superior to traditional CNNs and even more advanced architectures like Inception V3 as far as classification is concerned. The study was (2) reported an accuracy of 98% using EfficientNetB3, while the work was able to achieve 95.42 % accuracy using CoatNet. There is no question that the raw accuracy figure of the model is slightly lower, but given the complexity of the multi-class classification task in the study.

It is due to its hybrid nature, which combines CNN and transformer architectures, that CoatNet is capable of capturing both local as well as global features more effectively. The work (10) used a single-architecture approach, the MobileNetV2, which achieved 99% accuracy on a simpler dataset, but this approach was disadvantaged by the fact that it was a single-architecture approach. A further benefit of the approach is that it addresses the need for models that can be deployed in resource-limited settings, an issue that (2) who raised the need for such models. Despite maintaining high accuracy, the work has designed the model with computational efficiency in mind, so it can be used in a variety of healthcare environments, as it is designed with high accuracy in mind. In addition, the study has been distinguished from many other studies that have used binary classification in order to differentiate monkeypox from non-monkeypox, due to the fact that it utilized the MSLD v2.0 dataset, which includes a wide range of skin conditions besides monkeypox. By using advanced segmentation and classification techniques, this work utilizes a multi-class approach (12), allowing for a more realistic and applicable model. Table 2 shows the comparison with existing work.

Table 2: Comparison with Existing Work

Feature	Existing Work	Proposed System (Novel-Contribution)
Multi-Disease Dataset	Concentrates primarily on single-disease datasets, which makes co-detection difficult. This work has a single study diseases (1,12)	It includes multi-disease datasets (Monkey pox, Chickenpox, Measles, Cowpox, HFMD).
Data Augmentation	Using flipping and rotating techniques does not address significant data scarcity. The work (2) used basic augmentation	FMR (Flip, Mirror, Rotate) Random Image Augmentation increases dataset diversity. Additionally, it improves model generalization.

Handling Small Data Size	In many methods, data augmentation techniques are all that is needed to expand a small dataset. The work (3) use limited datasets.	A robust model uses FMR Augmentation coupled with advanced generative methods.
White Balance Correction	Lighting conditions vary, causing inconsistencies in image quality. Lighting variations are rarely addressed in studies (5).	Automated White Balance Correction (AWBC) improves image clarity and normalizes lighting conditions.
Segmentation of Multi-Disease Images	Segmentation limited to diseases without considering multiple diseases. For multi-disease datasets, This work (12) address segmentation.	Proposes Multi-scale Attention-guided Lesion Segmentation, enhancing diagnostic accuracy across multiple diseases.
Model Architecture	CNN-based models (e.g., ResNet, MobileNet) are used individually, which limits local/global feature extraction.	CoAtNet combines CNN and transformers for better local and global feature extraction.

The hybrid architecture's unique strengths are shown by skin lesions. Models such as EfficientNet, known for its scalable efficiency, demonstrate high accuracy in skin lesion classification, but often lack the ability to capture both local and global features. ViTs, on the other hand, excel at modeling global dependencies, but are computationally expensive and require large datasets. MobileNet variants provide lightweight architectures for resource-constrained deployments, but they may compromise precision. With CoAtNet, convolutional layers extract localized patterns while transformers analyze global features. Compared to EfficientNet-B3 and Inception V3, its accuracy (95.42%) and F1-score are higher. Both diagnostic accuracy and computational efficiency can be improved by CoAtNet.

CoAtNet-based approaches, although promising, have a few limitations. The use of high-quality, labeled datasets may limit generalizability, especially in resource-constrained regions. Because of its hybrid CNN-transformer architecture, the model may be difficult to deploy on low-power or edge devices. Due to potential biases in the dataset, CoAtNet's performance may vary across demographic groups. Additionally, image-based diagnosis overlooks clinical context or symptoms that are not visible visually. A hybrid model also presents interpretability challenges due to its complexity. Finally, the approach's reliance on preprocessing techniques may introduce variability if not standardized across clinical settings, affecting reproducibility. For robust and equitable deployment, these limitations must be addressed.

A real-world clinical comparison with expert dermatologists is crucial to evaluating CoAtNet's diagnostic performance. The model is evaluated for its accuracy, precision, and recall under variable conditions, ensuring its practical utility. Cohen's kappa or concordance correlation coefficient should be used to assess agreement, and any discrepancies should be investigated to identify gaps in the algorithm's generalization or clinical nuances. Comparison of this model validates its robustness and potential as a clinical decision aid.

Conclusion

Advanced deep learning techniques are capable of accurately detecting monkeypox from skin lesion images, according to the proposed research. CoAtNet-based models provide superior performance over traditional CNN and Inception V3 models because they incorporate a Random Image Augmentation method and an Automated White Balance Correction method. By using the model, monkeypox detection accuracy, precision, and recall have been improved by 95.42 %, 92.56%, and 95.74%, respectively. A number of factors have contributed to the successful outcome of the work. FMR Random Image Augmentation overcame the limitations of limited training data and enhanced the model's ability to generalize across different types of lesions. Furthermore, Automated White Balance Correction improved image quality by reducing lighting variations and color casts that could potentially confound diagnosis. Due to its hybrid architecture that combines the strengths of CNNs and transformers, CoAtNet is particularly effective at detecting both local and global features of skin lesions. Infectious disease outbreaks can be controlled by AI-assisted diagnostics, according to these studies. The available models for detecting monkeypox that are rapid, noninvasive, and accurate. By improving early detection efforts, it can reduce disease spread and improve patient outcomes.

Including multi-modal imaging, such as dermoscopy and radiography, is planned. An in-depth look at the infection could improve the diagnostic process. Advanced generative models, such as GANs, address the scarcity of data. A further aspect of optimizing models for real-time deployment in remote healthcare setting is the need to make them suitable for mobile and low-resource devices so that they can be used in real-time. It is also likely that by integrating noise-reduction algorithms and denoising algorithms into the AI pipeline, performance would be improved when dealing with data of low quality or that is noisy due to their lack of noise reduction techniques. To conclude, it is crucial that datasets are expanded so they include a wide range of racial and geographic diversity in order to ensure the models are capable of performing effectively across a wide range of populations, which will enhance the generalizability and fairness of the proposed solutions.

Abbreviation

Nil.

Acknowledgement

Nil.

Author Contributions

The corresponding author confirm sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Conflict of Interest

Authors declared there is no conflict of interest.

Ethics Approval

Not applicable.

Funding

No Funding.

References

1. Agrawal S, Castelino K, Mehta J, Bhavathankar P. EfficientNet-B3 and Image Processing for Monkeypox Detection using Skin Lesion Images. Int Conf Smart Gen Comput Commun Netw (SMART GENCON), Bangalore, India. 2022; 1:1–5.
2. Bogar SM, Deshmukh P, Reddy CVR, Muvva S. Monkeypox Detection using CNN-Based Pretrained Models. Int Conf Augment Intell Sustain Syst (ICAISS), Trichy, India. 2023; 1:173–178.
3. Dwivedi M, Tiwari RG, Ujjwal N. Deep Learning Methods for Early Detection of Monkeypox Skin Lesion. Int Conf Signal Process Commun (ICSC), Noida, India. 2022; 2:343–348.
4. Gupta P, Mittal U, Jha T, Agarwal M, Tiwari A. Efficient Prediction and Analysis of Monkeypox Skin Lesion: A Comparative Study for Web-based Application. IEEE Delhi Sect Flagship Conf (DELCON), Rajpura, India. 2023; 1:1–4.
5. Shah A. Monkeypox Skin Lesion Classification Using Transfer Learning Approach. IEEE Bombay Sect Signat Conf (IBSSC), Mumbai, India. 2022; 3: 1–5.
6. Gürbüz S and Aydin G. Monkeypox Skin Lesion Detection Using Deep Learning Models. Int Conf Comput Artif Intell Technol (CAIT), Quzhou, China. 2022; 2:66–70.
7. Irmak MC, Aydin T, Yağanoğlu M. Monkeypox Skin Lesion Detection with MobileNetV2 and VGGNet Models. Med Technol Congr (TIPTEKNO), Antalya, Turkey. 2022; 1:1–4.
8. Eliwa EH, El Koshiry AM, Abd El-Hafeez T, Farghaly HM. Utilizing convolutional neural networks to classify monkeypox skin lesions. Scientific reports. 2023 Sep 3;13(1):14495.
9. Arshed MA, Rehman HA, Ahmed S, Dewi C, Christanto HJ. A 16 × 16 Patch-Based Deep Learning Model for the Early Prognosis of Monkeypox from Skin Color Images. Computation. 2024; 12:1–14.
10. Alhasson HF, Almozainy E, Alharbi M, Almansour N, Shuaa S. A Deep Learning-Based Mobile Application for Monkeypox Detection. Appl Sci. 2023; 13:12589–12600.
11. Alharbi AH and Saber M. Diagnosis of Monkeypox Disease Using Transfer Learning and Binary Advanced Dipper Throated Optimization Algorithm. Biomimetics. 2023; 8:313–320.
12. Uysal F. Detection of Monkeypox Disease from Human Skin Images with a Hybrid Deep Learning Model. Diagnostics. 2023; 13:1772–1780.
13. Almufareh MF, Tehsin S, Humayun M, Kausar S. A Transfer Learning Approach for Clinical Detection Support of Monkeypox Skin Lesions. Diagnostics. 2023; 13:1503–1520.
14. Jaradat A, Al Mamlook RE, Almakayeel N, Alharbe N, Almuflih AS, Nasayreh A, Gharaibeh H. Automated Monkeypox Skin Lesion Detection Using Deep Learning and Transfer Learning Techniques. Int J Environ Res Public Health. 2023; 20:4422–4432.
15. Altun M, Gürüler H, Özkaraca O, Khan F, Khan J. Monkeypox Detection Using CNN with Transfer Learning. Sensors. 2023; 23:1783–1800.
16. Altindis M, Puca E, Shapo L. Diagnosis of monkeypox virus – An overview. Travel Med Infect Dis. 2022;50(10):1-6.
17. Yadav S and Qidwai T. Machine learning-based monkeypox virus image prognosis with feature selection and advanced statistical loss function. Med Microecol. 2024; 19:1-12.
18. Maqsood S, Damaševičius R, Shahid S, Forkert ND. MOX-NET: Multi-stage deep hybrid feature fusion

and selection framework for monkeypox classification. Expert Syst Appl. 2024; 255:(1245):1-16.

19. Nayak T, Chadaga K, Sampathila N, Mayrose H, Gokulkrishnan N, Bairy MG, Prabhu S, Swathi KS, Umakanth S. Deep learning based detection of monkeypox virus using skin lesion images. Med Novel Technol Devices. 2023; 18:1-13.
20. Ali SN, Ahmed MT, Jahan T. A Web-based Mpox Skin Lesion Detection System Using State-of-the-art Deep Learning Models Considering Racial Diversity. Biomed Signal Process Control. 2024; 98:1-11.
21. Govindaprabhu GB and Sumathi M. Ethno medicine of Indigenous Communities: Tamil Traditional Medicinal Plants Leaf detection using Deep Learning Models. Procedia Comput Sci. 2024; 23(5):1135–1144.
22. Govindaprabhu GB and Sumathi M. Safeguarding Humans from Attacks Using AI-Enabled (DQN) Wild Animal Identification System. Int Res J Multidiscip Scope (IRJMS).2024; 5(3):285–302.