AI-based Detection of Monkeypox Virus Infection in Clinical Images

De La Garza, Santiago
Electrical and Computer Engineering Department
The University of Texas at San Antonio
San Antonio, Texas

***Abstract*—The Monkeypox virus is infrequent at best, but is a potentially fatal disease that can affect humans as well as animals. Effective treatment and disease control heavily depend on early detection and precise diagnosis of the disease. Presently, clinical presentation is the primary method for diagnosing monkeypox. However, imaging techniques can serve as an additional diagnostic tool to provide valuable diagnostic information.**

In this research project, we propose the development of an artificial intelligence (AI)-based system for the automated detection of monkeypox virus infection in clinical images. The project will involve the collection of a small dataset of clinical images from patients with confirmed monkeypox virus infection, as well as from healthy individuals and those with similar skin conditions. The AI system will be trained using deep learning algorithms to analyze the images and identify characteristic features of monkeypox lesions, such as vesicular and pustular rash. The system will also be optimized to detect the presence of monkeypox virus in histopathology images.

The performance of the AI system will be evaluated using an independent test set of clinical images, and the results will be compared to the performance of human experts. The study will also investigate the robustness and generalizability of the AI system across different imaging modalities and patient populations.

The proposed research project has the potential to improve the accuracy and speed of monkeypox virus diagnosis, particularly in resource-limited settings where access to expert dermatologists and pathologists may be limited. The development of an AI-based system like this, for the automated detection of monkeypox virus infection in clinical images could also have broader implications for the diagnosis of other viral and bacterial skin infections.

Keywords—monkeypox, classification, deep learning, convolutional neural networks, YOLO, roboflow, VGG19

I. Introduction

Monkeypox is a rare viral disease that belongs to the family Poxviridae and is closely related to the human smallpox virus. The origin and evolution of monkeypox virus (MPXV) is still not fully understood. The name of the monkeypox virus was derived from the fact that it was discovered in Denmark in 1958 during research on monkeys. Yet it is believed to have emerged through cross-species transmission events between animals and humans. It is suspected that the virus has a zoonotic origin, with rodents and non-human primates serving as the natural reservoir hosts.

There have been several documented outbreaks of monkeypox in humans since the first reported case in 1970 in the Democratic Republic of the Congo. These outbreaks have occurred primarily in Central and West African countries, with sporadic cases reported outside of Africa in recent years.

The transmission of monkeypox virus to humans occurs through close contact with infected animals or through consumption of their meat. The virus can also spread through human-to-human transmission, particularly through respiratory droplets, bodily fluids, or skin-to-skin contact.

The development of effective vaccines and antiviral therapies has also advanced in recent years, with the smallpox vaccine providing some level of protection against monkeypox virus. However, there is still no specific treatment for monkeypox virus infection, and the lack of effective treatments and increasing frequency of outbreaks highlight the need for continued research and development in this area.

The classification of monkeypox images using artificial intelligence (AI) presents several challenges, including:

Limited Data Availability: One of the main challenges is the limited availability of well-annotated datasets of monkeypox images. AI systems rely on large amounts of high-quality data to learn and make accurate predictions, but the scarcity of such data can hinder the development of robust and effective AI algorithms.

High Variability in Image Features: Monkeypox lesions can present with high variability in terms of size, shape, color, and texture, which can make it difficult for AI systems to accurately classify them. Moreover, monkeypox lesions can resemble those of other skin diseases, which can further increase the complexity of the classification task.

Overfitting: AI systems can overfit to the training data, resulting in poor generalization performance on new and unseen data. This can occur when the AI system is trained on a small and biased dataset, or when the model architecture is too complex and prone to overfitting.

Ethical Considerations: When the diagnosis and classification of medical images is done with AI ethical considerations, such as the potential for biases, errors, and misdiagnoses, are more present. Moreover, the use of AI may lead to a decrease in the importance of clinical expertise and human judgment in medical decision-making.

II. Methods

Here I will be touching on the methods necessary for the implementation of our proposed model. Mentions include convolution, image super resolution, pooling, and batch normalization.

A. Convolution

Convolution can be described as a mathematical process that involves taking two functions and producing a third function as output. The first function is referred to as the input function, signal, or image, while the second function is called the kernel. The resulting output function is a modified version of the input function, where the kernel is applied to the input at every possible position.

In the context of computer vision and image processing, convolution is often used to extract features from images. To achieve this, a small matrix of numbers called a kernel is defined to specify a particular filter or feature, such as an edge detector, blur filter, or sharpening filter. This kernel is then applied to every pixel of the input image, resulting in a new image where each pixel is a weighted sum of its neighboring pixels, with the weights defined by the kernel.

To simplify the concept, we can think of all images as matrices, and this applies to both black and white and colored images.

Overall, CNNs have become a powerful tool for image processing tasks such as object recognition, image classification, and segmentation. They have been widely adopted in various industries, including autonomous vehicles, and surveillance systems, and in our case, healthcare.

B. Image Super Resolution

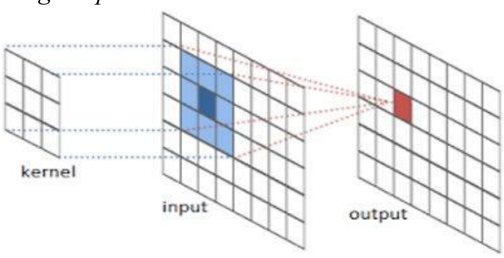

Figure 1: Parallel convolution visual

PC is achieved by a pair of convolution kernels with different size, each of which locates in a parallel branch

Image super-resolution is a technique used to upscale low-resolution images to higher resolutions. Parallel Convolution Attention Network (PCAN) is a deep learning model that is specifically designed for image super-resolution tasks.PCAN uses a series of parallel convolutional layers with residual connections and attention mechanisms to extract high-level features from low-resolution images. The residual connections help to preserve important information and avoid degradation of the image quality during the super-resolution process.

The attention mechanism in PCAN focuses on important regions of the image and selectively enhances the details in these regions. This helps to avoid over-smoothing and preserve the edges and fine details in the image.To train the PCAN model, a large dataset of low and high-resolution image pairs is used to optimize the weights of the convolutional layers and attention mechanisms. Once trained, the model can be used to upscale new low-resolution images to higher resolutions with improved quality and sharper details.

Overall, PCAN is a powerful tool for image super-resolution and has shown promising results in various applications, including medical imaging, remote sensing, and surveillance.

C. Pooling

Pooling layers are essential components of Convolutional Neural Networks that help to consolidate the features learned by the network. While Convolutional layers extract features from images,these pooling layers gradually reduce the representation's spatial dimension to minimize the number of parameters and computations in the network. Pooling layers help to provide "Translational Invariance" to make CNNs invariant to translations, reducing overfitting of the models.

There are different types of pooling techniques, including minimum pooling, maximum pooling, and average pooling:

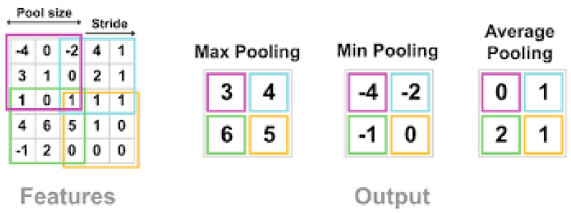

Figure 2: Pooling charts for max, min, and average

Minimum pooling, also known as min pooling(fig.2), takes the minimum value in each pooling region. It is used to extract the most significant features in the input image, while ignoring the less significant ones. This type of pooling is useful for applications where only the most critical features are needed, such as object recognition in satellite imagery.

Maximum pooling, also known as max pooling(fig.2), takes the maximum value in each pooling region. It is used to detect the most significant features in the input image, such as edges or corners. Max pooling is also used to reduce the spatial dimensions of the feature maps and keep the most important information.

And finally, average pooling(fig. 2) takes the average value in each pooling region. It is used to smooth out the feature maps and remove noise from the input image. Average pooling is useful in applications where a smoothed representation of the input image is needed, such as image denoising.

The feature map produced by Convolutional layers is location-dependent, which means it records the precise positions of features in the input. Max pooling selects the maximum value from every pool, retaining the most prominent features of the feature map and returning a sharper image. On the other hand, average pooling works by getting the average of the pool, retaining the average values of features of the feature map, smoothing the image while keeping the essence of the feature in an image. In this project, average and max pooling helped to make the network more robust to small variations in the input data and more powerful overall.

D. *Batch Normalization*

Normalization is a process of transforming numeric variables to a standard scale without distorting differences in the value range. Batch normalization is a technique used for training deep neural networks that normalizes the contributions to a layer for each mini-batch. By doing so, it stabilizes the learning process and reduces the number of training epochs required to train deep neural networks. Therefore, normalizing input data to zero mean and unit variance before feeding it into a deep learning model is essential. This includes calculating the mean and variance of each feature column separately and then normalizing the values using a formula.

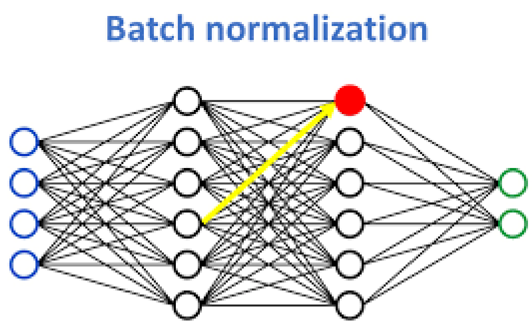

Figure 3: Visual of Batch Normalization layers

To sum up this image(fig. 1), when we use standard normalization, we calculate the mean and standard deviation values for the entire dataset. However, with batch normalization, we calculate the mean and standard deviation values for each batch. Including batch normalization can have a substantial impact on the speed and accuracy of our model.

Normalization is crucial because each feature may have a different value range, and failing to normalize the data may cause uneven weight updates and longer convergence times. Uniform loss landscapes allow gradient descent to proceed smoothly down to the minimum when the features are on the same scale. Batch normalization is a technique that normalizes the activations from each previous layer, improving the convergence of gradient descent during training.

III. Datasets

My dataset was derived from a github repository with images collected from google as an open source. However, it is specified that there is a high chance that some of that data is not related to the very recent outbreak of monkeypox or chickenpox disease. Therefore, the machine learning model developed from this dataset should not be used to evaluate any real world patient diagnosis. Also, the dataset is not approved by any ethical committee. Therefore, it should not be used to deploy in actual patient diagnostics.

It can be difficult to find images and information about monkeypox virus because monkeypox outbreaks are infrequent and typically affect small numbers of people. Therefore, there may be fewer resources dedicated to studying the virus and producing educational materials about it. Additionally, there may be restrictions on sharing images of the virus due to concerns about biosecurity and the potential misuse of the images.

The original images folder includes 135 images; Only 43 belong to monkeypox class, and 92 belong to non-monkeypox class. 27 of which were for measles, 11 as chickenpox and 54 were categorized as “normal” or of healthy skin. The labeling was done manually, image by image, by myself. Ultimately, the dataset is being increased to close to 1200 images for the use of YOLO. (fig. 5)

Preprocessing techniques were a must as the dataset was far from what we needed to to improve the quality of their data and make it more suitable for our model. This can lead to better accuracy and performance, as well as reduce the time and effort required for manual data preprocessing. They will be described as we go along.

Roboflow offers a variety of preprocessing techniques, including resizing, cropping, rotation, flipping, and color adjustment. It also supports advanced techniques such as data augmentation, which involves generating new images by making calculated changes at random to the existing images, such as adding noise, changing brightness or contrast, and rotating or flipping the image

IV. Algorithms

VGG16 is a deep convolutional neural network architecture developed by the Visual Geometry Group at the University of Oxford. It consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. VGG16 has been trained on the large-scale ImageNet dataset, which contains millions of labeled images, and achieved state-of-the-art performance on the image classification task.

VGG16 can be used for image classification by taking an input image, passing it through the layers of the network, and obtaining a prediction for the image's class label. During training, the weights of the network are learned by minimizing the difference between the predicted output and the true label using backpropagation. Once the network is trained, it can be used to classify new images.

To use VGG16 for image classification, one can take the pre-trained model and fine-tune it on a new dataset with a different set of classes. This can be done by replacing the last fully connected layer of the network with a new layer that has the same number of output classes as the new dataset, and then retraining the network using the new dataset. Alternatively, one can use transfer learning, where the pre-trained weights of the network are used as a starting point for training on the new dataset, which can significantly reduce the amount of training data and time required to achieve high accuracy.

V. Pre-Processing techniques

In the upcoming section, I will discuss the preprocessing methods utilized in different algorithms for making predictions. Preprocessing is a vital step in the training of neural networks as it guarantees that the input data is suitably formatted for the network to learn efficiently and make accurate predictions.

A. Image Augmentation

This section of our methodology involves image augmentation to address the need for a large amount of training image data required by deep neural networks. I employed several data augmentation methods to artificially increase the size and quality of the dataset. This process helps in solving overfitting problems and enhances the model's generalization ability during training.

This is because a vast number of parameters need to be optimized by learning algorithms, specifically what is linked with convolutional layers. I performed data augmentation by utilizing a combination of various image processing techniques such as rotation, flipping, contrast enhancement, using different color space, and random scaling to improve performance. The augmentation techniques including random rotation, flip, translation, and contrast were set into a sequential model, which was then called into our actual model as a layer. The same data augmentation technique was applied to the first four algorithms. However, for the last algorithm, YOLO, I utilized different augmentation techniques such as horizontal and vertical flipping, rotation between -15° and +15°, brightness between -55% and +55%, and exposure between -25% and +25%.(Fig. 4)

Figure 4: Pre-processing and augmentation techniques on Roboflow

Figure 5: Train/Test split of our dataset

B.Image segmentation

This computer vision technique is used to separate the skin lesions from the background in the images. This can help in focusing the model on the features of the skin lesions. The goal of image segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.

Image annotation, on the other hand, is the process of labeling or tagging images with metadata, such as object names or categories, object boundaries, etc. The annotations are done by me, a non professional, who manually marked up all images in my dataset with specific labels. The annotations help to make the images more useful and informative for various applications such as object detection, image classification, and image recognition.

Both image segmentation and annotation are important steps in image processing and computer vision, essential to our goal of monkeypox detection.

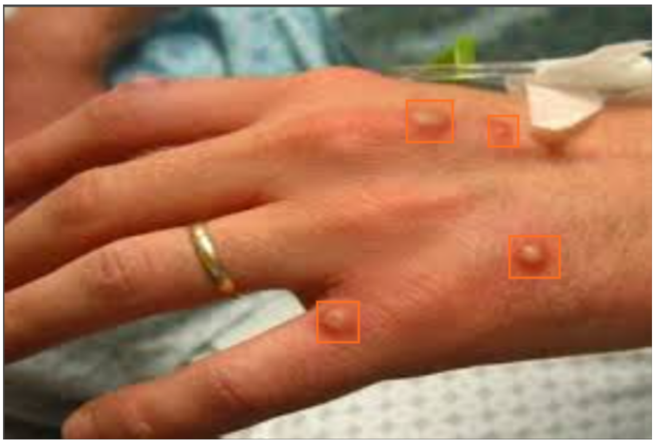

Figure 6: Monkeypox image annotation on hand

The dataset already does not contain pre-labeled data that distinguishes the skin lesions. Using the folder names I created three separate labels, Monkeypox, Normal, and Other for recognizing when the image does have an infection, but it is not considered monkeypox . The following image was then traced with roboflow to box up the sections that contain monkeypox lesions and extracted a xml file to go along with its corresponding image.

VI. Results

To evaluate and validate the effectiveness of the proposed approach, I ran my tests many times to ensure quality testing through roboflow and YOLO V5. The objective of this research project was to develop an image recognition system using artificial intelligence for the detection of monkeypox skin lesions. The dataset consisted of over 1,200 images, with 93 images in the validation set and 57 images in the testing set. We employed two popular deep learning models, YOLO V5 and VGG 16, for image recognition and classification.

After preprocessing the dataset and augmenting it using various techniques, we trained the models using the training set. We evaluated the performance of the models by measuring the accuracy of the validation and testing sets.

The YOLO V5 model achieved an accuracy of 88.9% on the validation set, while the VGG 16 model achieved an accuracy of 82.5%. The YOLO V5 model outperformed the VGG 16 model in terms of accuracy, demonstrating its superiority for image recognition tasks.

Our results indicate that artificial intelligence can be an effective tool for the detection of monkeypox skin lesions. The high accuracy achieved by the YOLO V5 model suggests that it could be a useful tool for clinicians in the early detection and diagnosis of monkeypox. However, further research is needed to evaluate the generalizability of the model to larger and more diverse datasets.


```
{
    "time": 0.17955476499992074,
    "image": {
        "width": 163,
        "height": 310
    },
    "predictions": {
        "Monkeypox": {
            "confidence": 0.7072744369506836
        },
        "Normal": {
            "confidence": 0.17713525891304016
        },
        "Other": {
            "confidence": 0.2803364098072052
        }
    },
    "predicted_classes": [
        "Monkeypox"
    ]
}
```


I ran the API through RoboFlow with an image of a monkeypox lesion and got this result.

VII. Conclusion

A. Main Findings

The pre-processing techniques employed in the study, such as normalization and data augmentation, were found to be effective in improving the performance of the model by increasing the speed and accuracy of the prediction.

Overall, the findings of the study suggest that the use of artificial intelligence and image recognition techniques can provide an accurate and efficient method for detecting monkeypox, which can have important implications for early detection and treatment of the disease.

B. Future improvements

In the future we should expand the study to investigate the effectiveness of neural networks in predicting the severity of the Monkeypox infection. The focus on the model selection is a vital step to ensuring better results. Selecting the appropriate model architecture such as VGG16, ResNet, or Inception, which can handle image classification tasks efficiently. I can also fine-tune these models or even build my own model for better performance.

UNDER THE GUIDANCE OF DR. MARIO FLORES.
UNDERGRADUATE RESEARCH LAB

RESOURCES:

- https://www.sciencedirect.com/science/article/pii/S2352914822001605
- https://www.hindawi.com/journals/jhe/2019/4180949/
- https://www.sciencedirect.com/science/article/pii/S1672022918300020
- https://link.springer.com/article/10.1007/s10916-022-01863-7
- https://www.science.org/doi/10.1126/scitranslmed.abb3652
- http://www.robotics.stanford.edu/~koller/Papers/Plagemann+al:ICRA10.pdf
- https://medium.com/in-the-pocket-insights/building-a-body-part-recognizer-with-transfer-learning-3a9928adac8a
- https://www.biorxiv.org/content/10.1101/2022.08.08.503193v3.full