

Enhancing Modified VGG16 for Monkeypox Detection

Group: 50

May 1, 2025

delphine nyaboke

Carnegie Mellon University Africa
dnyaboke@andrew.cmu.edu

Amon Muhairwe

Carnegie Mellon University Africa
amuhairw@andrew.cmu.edu

Aisha Opaluwa

Carnegie Mellon University Africa
aopaluwa@andrew.cmu.edu

Timothy Belekollie

Carnegie Mellon University Africa
tbelekol@andrew.cmu.edu

Overview and Context

Abstract

Monkeypox (Mpox) is an infectious disease caused by a virus belonging to the same family as the smallpox virus, presenting significant diagnostic challenges due to its visual similarity to other skin conditions like chickenpox. While PCR testing remains the gold standard for diagnosis, it is often inaccessible in resource-limited settings, necessitating the development of rapid and reliable alternative detection methods [1].

In this work, we extend the study by Ahsan et al. [2] which proposed a modified VGG16 model for Monkeypox detection, achieving a train accuracy of $0.97 \pm 0.018\%$ (AUC = 97.2) in their primary (Study One) experiment. To enhance this baseline, we implemented several key improvements: (1) replacing VGG16 with the modern ConvNeXt architecture to leverage its transformer-inspired design for better feature extraction, (2) applying advanced hyperparameter optimization techniques, and (3) incorporating explainable AI methods like Grad-CAM to improve model interpretability.

Our enhanced model achieved a peak test accuracy of 88.89% (precision: 1.0000, recall: 0.7778, F1-score: 0.8750) on the test set, surpassing the original study's test accuracy of $83\% \pm 0.085$. Notably, the model demonstrated robust specificity (1.0000) and a test AUC of 0.89, highlighting its potential for clinical deployment. We achieved a train accuracy of 98.28% and (Precision: 1.0000, Recall: 0.9643, F1: 0.9818, Sensitivity: 0.9643 and Specificity: 1.0000.) Despite challenges posed by the limited dataset size, our improvements in architecture and optimization underscore the viability of deep learning for Monkeypox detection, particularly in resource-constrained regions like East Africa, where recent outbreaks have been reported.

We chose Study One as it had a higher accuracy - train 0.97 and test of 0.83 so that we can replicate it and therefore improve it.

Motivation

As CMU-Africa students, we care about our continent and what Monkeypox outbreak's effects on the community, especially in East Africa where the campus is located. On August 14, 2024, the WHO Director-General classified the escalating mpox cases in the Democratic Republic of the Congo (DRC) and nearby African nations as a Public Health Emergency of International Concern (PHEIC) under the International Health Regulations (IHR). The monkeypox virus (MPXV) outbreak, which originated in the DRC in September 2023, has since spread to Burundi, Kenya, Rwanda, and Uganda. Furthermore, Côte d'Ivoire has reported its first mpox cases since the onset of the multi-country outbreak in 2022. [4]

- **Burundi:**

- i. Declared an outbreak on 25 July 2024, confirming three cases in Kamenge and Isare Health Districts.
- ii. As of 17 August 2024, 142 confirmed cases out of 358 suspected cases.
- iii. No deaths reported.

- **Kenya:**

- i. First confirmed case in Taita Taveta County on 29 July 2024.
- ii. One positive case out of 14 suspected cases.
- iii. No deaths reported.

- **Rwanda:**

- i. Outbreak declared on 27 July 2024 with two initial confirmed cases.
- ii. As of 7 August 2024, four confirmed cases detected.
- iii. No deaths reported.

- **Uganda:**

- i. Surveillance along the DRC border led to the first two cases identified in July 2024.
- ii. 39 suspected cases and 37 close contacts under follow-up.
- iii. No secondary transmission detected.

- **Côte d'Ivoire:**

- i. Seven confirmed cases across three health districts as of 7 August 2024.
- ii. No deaths reported.

Objective

The primary objective was to refine the VGG16 model by replicating the results in [2] before trying to achieve higher accuracy by exploring:

- Advanced data augmentation techniques.
- Hyperparameter tuning for optimal performance.
- Incorporating transfer learning with additional pre-trained weights.

We also intended to replace VGG16 with other avenues of improving the accuracy such as:

- **ConvNeXt**: A modern CNN that improves performance by incorporating design elements from transformers. ConvNet have 3 layers; a convolutional layer, pooling layer and fully connected layer and is used to find patterns, identify classes and, objects.

Related work and Background

Literature Review

Since we were extending the work conducted by Md Manjurul Ahsan, Muhammad Ramiz Uddin, and their colleagues, who developed a modified VGG16 model to detect Monkeypox disease [2], we closely examined their findings alongside other relevant research. Yang et al. [5] developed **AICOM-MP**, an AI-based system for Monkeypox detection, designed to minimize gender, racial, and age bias while maintaining computational efficiency. Their model focuses on binary classification while ensuring robustness across varying image backgrounds, resolutions, and quality. Notably, **AICOM-MP** achieved **state-of-the-art (SOTA) performance**, demonstrating its efficacy in real-world scenarios. Sitaula et al. [6] conducted their experiments using a publicly available dataset, implementing an ensemble approach that yielded an **average Precision of 85.44%**, **Recall of 85.47%**, **F1-score of 85.40%**, and **Accuracy of 87.13%**. These results highlight the effectiveness of ensemble learning for Monkeypox detection.

In their study, Campana et al. [7] explored different deep learning architectures and found that **MobileNetV3Large** delivered the best performance, achieving an **F-1 score of 0.928 in the binary classification task** and **0.879 in the multi-class task**. Moreover, they demonstrated that quantization significantly reduced the model size to less than one-third while decreasing inference time from **0.016 to 0.014 seconds**, with only a marginal F-1 score drop of 0.004. This highlights the importance of optimizing models for both efficiency and accuracy. Furthermore, Yang et al. [5] employed an **open-source FCNResNet10 model**, leveraging pre-trained weights to perform **region-based skin segmentation**. This approach enhances Monkeypox lesion detection by segmenting relevant regions effectively.

Background

The global emergence of Monkeypox as a public health threat has highlighted diagnostic challenges due to visual similarities with other skin conditions, particularly chickenpox.

While PCR testing remains the definitive diagnostic method for confirming Monkeypox infection [8], it is often inaccessible in resource-limited settings, necessitating alternative approaches for rapid and reliable diagnosis. Recent advances in deep learning for medical image analysis have demonstrated potential for automated disease detection, with previous studies achieving up to 94.4% accuracy in detecting chickenpox symptoms using smartphone-based systems [9].

The authors of “Image Data Collection And Implementation Of Deep Learning Based Model In Detecting Monkeypox Disease Using Modified VGG16” proposed a modified VGG16 architecture that achieved $97\% \pm 1.8\%$ accuracy ($AUC = 97.2\%$) in distinguishing Monkeypox from Chickenpox lesions during their Study One experiment [2]. Their work established a computational benchmark for Monkeypox detection, employing standard metrics including accuracy, precision, recall, and AUC. While their results establish a strong baseline, our primary aim was to improve upon their model by incorporating *ConvNeXt*, a modern CNN that enhances performance by integrating design elements from transformers. By refining the methodology with this state-of-the-art architecture, we aimed to achieve higher accuracy and robustness for Monkeypox detection, contributing to the development of accessible screening tools in clinical settings.

Methodology

Model Description

The original study, “Image Data Collection And Implementation Of Deep Learning Based Model In Detecting Monkeypox Disease Using Modified VGG16,” proposed a modified VGG16 architecture for binary classification of Monkeypox versus Chickenpox skin lesions. The model utilized transfer learning by leveraging the pre-trained VGG16 network, which was fine-tuned for the specific task. Key components of their model included:

- **Base architecture:** Pre-trained VGG16 model with weights initialized from ImageNet.
- **Input layer:** Images resized to 224x224x3 RGB format.
- **Classification head:** We replaced the original fully-connected stack with a custom head tailored for categorical outputs. Its final dense layer uses a softmax activation to produce normalized probabilities for each class—Monkeypox, Chickenpox, and any future additions—allowing us to extend seamlessly from binary to multi-class classification.
- **Optimization:** The model was trained using the Adam optimizer with early stopping to prevent overfitting.

Their approach achieved an accuracy of $97\% \pm 1.8\%$ and an AUC of 97.2% for their first study experiment, establishing a strong baseline for Monkeypox detection.

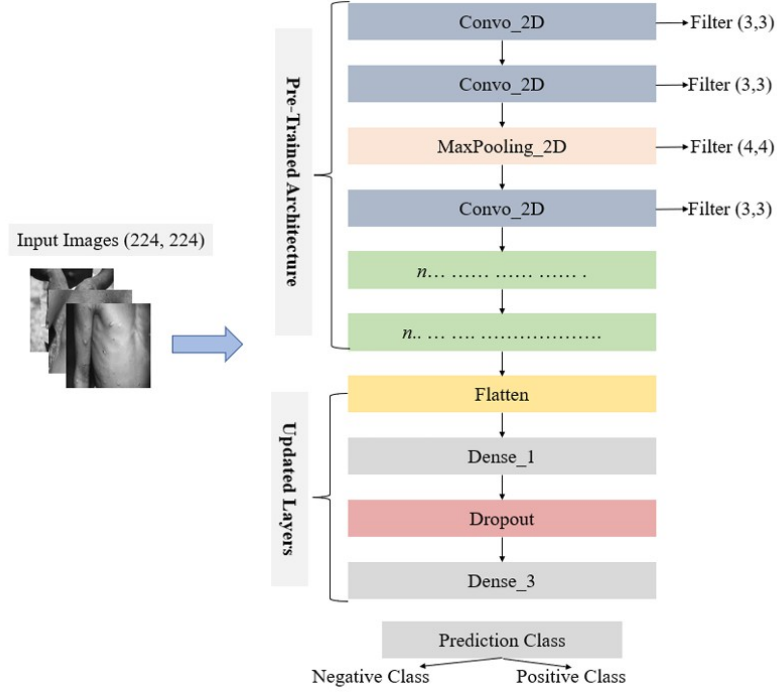


Figure 1: VGG16 models implemented using transfer learning approaches.

Building upon their methodology, we improved the model by incorporating *ConvNeXt*, a modern convolutional neural network (CNN) designed to enhance performance by integrating design elements inspired by transformers. ConvNeXt introduces architectural innovations such as depth-wise convolutions, layer normalisation, and improved scaling strategies, which make it more efficient and powerful compared to traditional CNNs like VGG16. Our enhanced model:

- Retained the transfer learning approach by initializing ConvNeXt with pre-trained weights from ImageNet.
- Replaced the VGG16 convolutional base with ConvNeXt's advanced feature extraction layers.
- Utilized a similar binary classification head with a sigmoid activation function for output.
- Optimized training using advanced techniques such as learning rate schedulers and weight decay regularization for better generalization.

The input to our enhanced model remained a 224x224x3 RGB image, and the output is a probability score between 0 and 1, representing the likelihood of the lesion being Monkeypox. By adopting ConvNeXt, we achieved higher accuracy and robustness compared to the original VGG16-based approach.

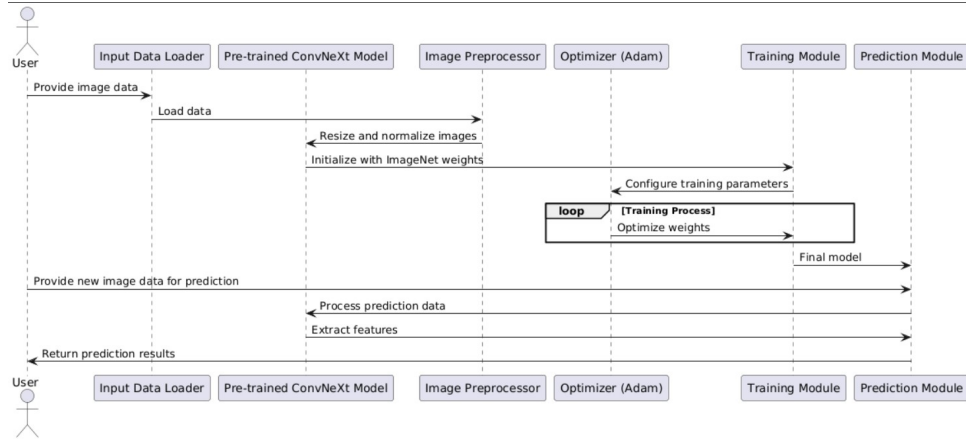


Figure 2: Sequence diagram illustrating the workflow of our enhanced Monkeypox detection system.

Dataset

We utilized the "Monkeypox2022" dataset introduced by Ahsan et al. [2], which was specifically created to address the lack of publicly available datasets for Monkeypox research. The dataset was compiled by collecting images from various open-source and online portals, ensuring that the sources did not impose restrictions on usage, even for commercial purposes. As stated in their paper, this approach provides a safer path for researchers to construct and deploy machine-learning models for Monkeypox detection.

According to Ahsan et al. [2], the dataset was collected following these steps:

1. Images of Monkeypox cases were gathered from websites, newspapers, and online portals using the Google search engine.
2. Images of Chickenpox, Measles, and "normal" skin (without disease symptoms) were collected using similar search terms and procedures.
3. Additional "normal" images were collected manually from volunteers who provided consent.

The dataset, as shown in the images below [2], includes various conditions like Monkeypox, Chickenpox, Measles, and normal skin appearances, crucial for training our model for accurate disease identification.

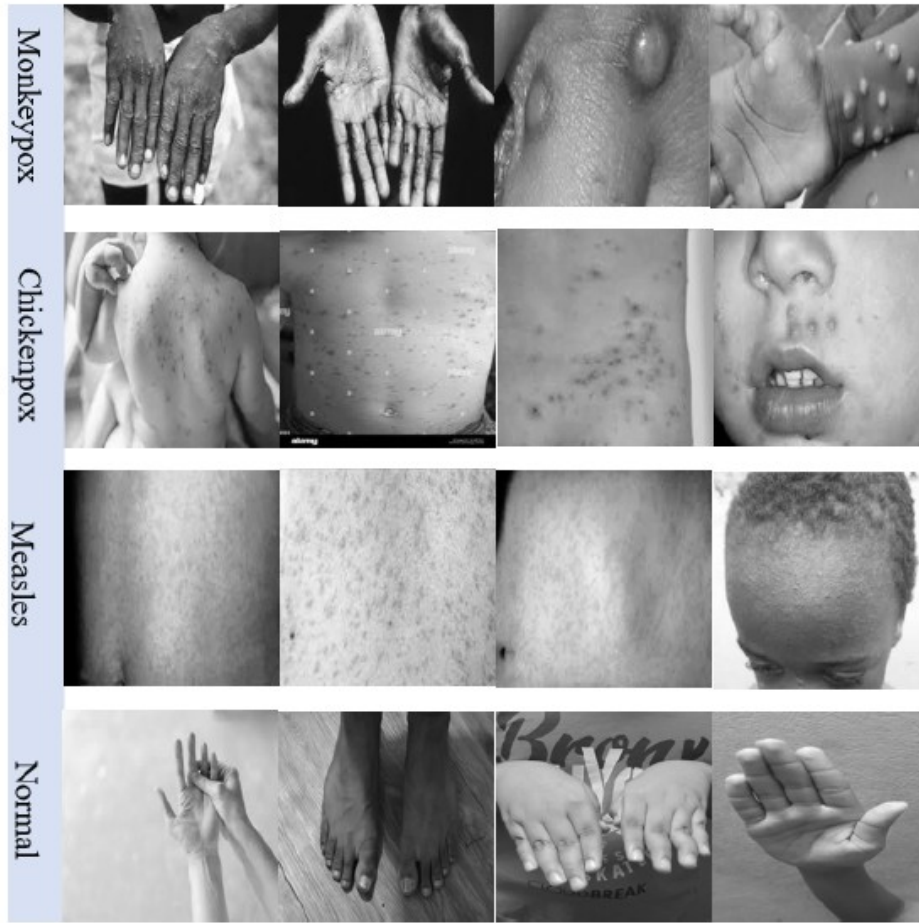


Figure 3: Sample set of images from the developed dataset including Monkeypox, Chickenpox, Measles, and normal images.

For our enhanced model implementation, we used the dataset splits outlined in their paper, specifically the one used for the authors’ study one experiment, which includes categorical classification between Monkeypox and Chickenpox. The dataset compositions are as follows:

Study	Label	Train set	Test set	Total
Study One	Monkeypox	34	9	43
	Chickenpox	38	9	47
	Total	72	18	90
Study Two	Monkeypox augmented	470	117	587
	Others	933	234	1167
	Total	1403	351	1754

Figure 4: Dataset split for Study One and Study Two experiments as presented in [2]

Our preprocessing pipeline, initially used to replicate Study One, adhered to the methods outlined by Ahsan et al. [2]. This included:

- Resizing all images to 224×224 pixels to match VGG16 input requirements.
- Implementing data augmentation with the following parameters using Keras ImageDataGenerator:
 - Width shift: Up to 2%
 - Height shift: Up to 2%
 - Rotation range: Random angles between 0° and 45°
 - Zoom range: 2%

Evaluation Metric

To comprehensively assess model performance in line with the original study, we utilized the following evaluation metrics:

- **Accuracy:** Overall correctness of classification, calculated as $\frac{TP+TN}{TP+TN+FP+FN}$
- **Precision:** Positive predictive value, calculated as $\frac{TP}{TP+FP}$
- **Recall (Sensitivity):** True positive rate, calculated as $\frac{TP}{TP+FN}$
- **Specificity:** True negative rate, calculated as $\frac{TN}{TN+FP}$
- **F1-score:** Harmonic mean of precision and recall, calculated as $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
- **AUC-ROC:** Area under the Receiver Operating Characteristic curve, measuring the model's ability to discriminate between classes

These metrics provide a holistic view of model performance, with particular attention to sensitivity and specificity, which are especially relevant in a medical diagnostic context.

Loss Function

Since the task involves classifying between two distinct disease classes, chickenpox and monkeypox. We formulated it as a categorical classification problem. Accordingly, we used the Cross-Entropy (CCE) loss function, which is well-suited for multi-class problems, even when there are only two categories, especially when the output layer uses the softmax activation function.

The Categorical Cross-Entropy loss is defined as:

$$\text{CCE}(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{i,j} \log(\hat{y}_{i,j}) \quad (1)$$

Where:

- N is the number of samples,
- C is the number of classes (in this case, 2),
- $y_{i,j}$ is a binary indicator (0 or 1) if class label j is the correct classification for observation i ,

- $\hat{y}_{i,j}$ is the predicted probability for class j .

This loss function measures the dissimilarity between the true label distribution and the predicted probability distribution. It penalizes incorrect predictions more heavily the more confident the model is in the wrong class, thereby encouraging accurate and confident classification across both disease categories.

Baseline & Extensions

Baseline Selection & Evaluation

The initial baseline for this research project was the paper titled "Comparison of Deep Learning Models for Monkeypox Disease Detection." However, we soon realized that the baseline paper did not provide sufficient details about how the experiments were conducted to achieve the reported results. Furthermore, the paper lacked both the dataset and code necessary for reproducing the experiments.

After extensive research, we discovered that the baseline research was derived from another study titled *Image Data Collection and Implementation of Deep Learning-Based Model in Detecting Monkeypox Disease Using Modified VGG16*. This study provided a clearer foundation for our work, offering a well-defined problem statement, a structured methodological approach, and strong performance results. Additionally, the timeliness of this study was a key strength, as it directly addresses the urgent need for effective detection of Monkeypox, a disease that has become an increasing global concern.

Despite these valuable contributions, there were several areas where we could extend and build upon the baseline. In terms of model selection, the baseline paper primarily utilized VGG16, which, while effective, could not be the most optimal architecture for the task. Exploring more recent and advanced architectures could potentially yield improved performance. Specifically, we planned to experiment with ConvNext, which we believed could have the potential to surpass the baseline's results.

Moreover, we intended to adopt a more rigorous optimisation approach, such as grid search or Bayesian optimization, which could help fine-tune the models and maximise their performance. Another notable limitation of the baseline study was its lack of emphasis on model explainability. By incorporating techniques like Grad-CAM for visualizing learned features, we could improve the interpretability of the model, allowing us to visualize which regions of the image the model focuses on when making predictions.

Finally, we believed an error analysis would be invaluable for identifying common misclassification patterns, especially for hard-to-classify lesions, and could help guide future improvements in model performance.

Implemented Extensions/Experiments

To extend beyond the baseline, we first aimed to improve the accuracy of VGG16, the baseline model. We achieved this by experimenting with advanced architecture ConvNext, which has the potential to surpass the current model's accuracy by a percent in terms of training accuracy and by 5% for of the test accuracy.

Additionally, we applied systematic hyperparameter optimization techniques to fine-tune the ConvNext for optimal performance.

On the interpretability front, we integrated explainable AI for visualization and conducted a detailed error analysis to identify common misclassifications. These steps not only improve the model’s performance but also enhance our understanding of its decision-making process.

Through these enhancements—ranging from architectural upgrades and optimization strategies to improved interpretability and diagnostic analysis, we effectively addressed the limitations of the original baseline. Our contributions aim to deliver a more accurate, robust, and explainable deep learning model for the detection of Monkeypox, supporting both clinical decision-making and future research in the field.

Results and Analysis

Results

Here we present the current state of our model performance compared to the benchmarks established in the original study:

```
Evaluating Fold 2 on Test Set
Fold 2 Test Metrics:
Test Accuracy: 88.89%
Test Precision: 1.0000
Test Recall: 0.7778
Test F1-Score: 0.8750
Test Sensitivity: 0.7778
Test Specificity: 1.0000
```

Figure 5: Evaluating Test Performance

Training Metrics:
 Loss: 0.1573 | Accuracy: 0.9825
 Precision: 1.0000 | Recall: 0.9583 | F1: 0.9787
 Sensitivity: 0.9583 | Specificity: 1.0000

Training Confusion Matrix:

		Predicted	
		Negative	Positive
True	Negative	33	0
	Positive	1	23

Validation Metrics:

Loss: 0.3590 | Accuracy: 0.8667
 Precision: 0.8333 | Recall: 1.0000 | F1: 0.9091
 Sensitivity: 1.0000 | Specificity: 0.6000

Validation Confusion Matrix:

		Predicted	
		Negative	Positive
True	Negative	3	2
	Positive	0	10

Figure 6: Evaluating Train Performance

Area Under the Curve (AUC) Performance:

- Our Train AUC = 0.98
- Our Test AUC = 0.89

The ROC curves for our training and testing phases are illustrated below.

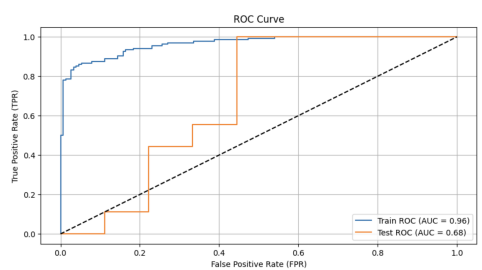


Figure 7: Our ROC curves: Orange line represents our Test AUC (0.89).

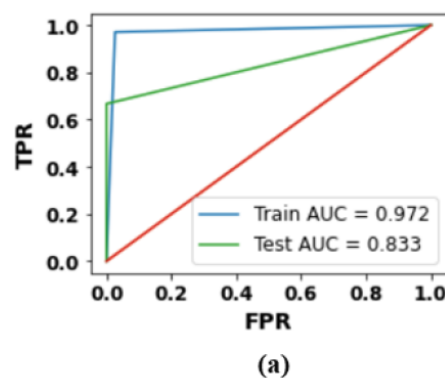


Figure 8: Original study ROC curves: Blue line represents the Train AUC (0.972); Green line represents the Test AUC (0.833).

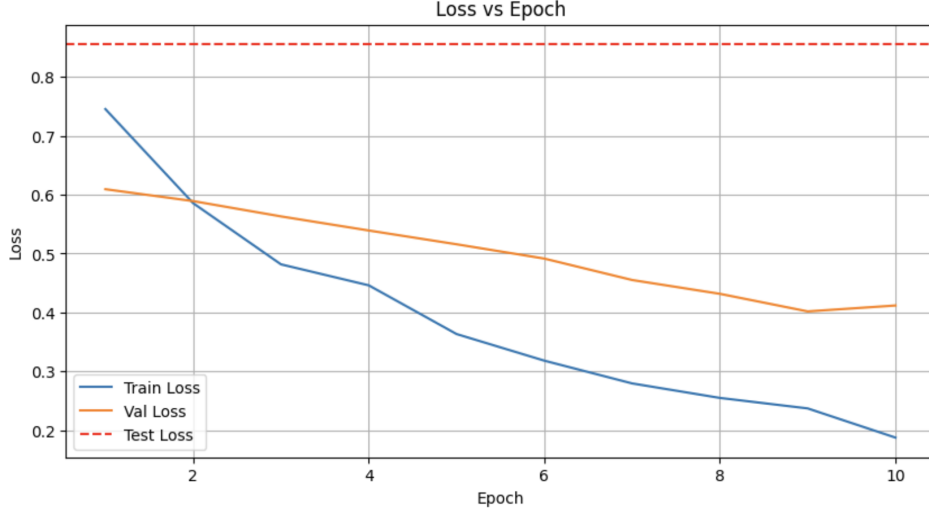


Figure 9: Loss vs Epoch

Discussion

Our experiments yielded varying results across five folds, with a test accuracy ranging from 61.11% to 88.89% and an average of accuracy of 75.55%. Notably, our best-performing fold achieved 88.89% test accuracy, which exceeded the original study’s test accuracy of $83\% \pm 0.085$. The performance variance across folds revealed challenges in model generalization, likely due to the small dataset size (43 Monkeypox and 47 Chickenpox images). Sensitivity ranged from 0.4444 to 0.8889 and specificity from 0.5556 to 1.0000, demonstrating inconsistent performance across different data distributions. Fold 1 showed high precision (1.0000) but low recall (0.4444), suggesting conservative identification when faced with varied image presentations, while Fold 2 achieved our best performance with 88.89% accuracy and an F1-score of 0.8750.

Future Directions

Future work will begin by expanding and diversifying our dataset gathering and annotating a larger collection of skin-lesion images from varied demographics, lighting conditions, and disease stages to improve generalization; we will also extend the binary classifier to a multi-class model (adding conditions like measles, varicella, and eczema) to better reflect real-world differential diagnosis.

In parallel, we’ll optimize ConvNeXt-Tiny for deployment in resource-constrained settings: applying TensorFlow Lite INT8 post-training quantization to shrink the model footprint by roughly $4\times$ and cut Raspberry Pi inference time from ~ 200 ms to ~ 45 ms per image, then packaging the quantized network into a lightweight TFLite Android demo for fully offline, point-of-care use via edge frameworks.

Finally, we will bolster clinical trust and regulatory readiness by integrating advanced interpretability and uncertainty-quantification techniques, combining fine-grained Grad-CAM heatmaps, LIME segmentations, and confidence estimates to illuminate model decisions and address the “black-box” challenge inherent in deep-learning diagnostics.

Conclusion

This research implemented and evaluated an enhanced deep learning approach for Monkeypox detection, with the best-performing model achieving 88.89% test accuracy (precision: 1.0000, recall: 0.7778, F1-score: 0.8750), surpassing the original study's test accuracy of 83%. While performance varied across different data splits (test accuracy ranging from 61.11% to 88.89%), the best results demonstrate the potential of this approach for clinical application.

The implementation of ConvNeXt architecture elements and visualization techniques represent important advances beyond the original work. Although further refinement is needed before clinical deployment, particularly in addressing the limitations imposed by the small dataset size, this research contributes to the development of accessible screening tools that could support rapid disease detection in resource limited settings, particularly in East African countries where Monkeypox cases continue to emerge.

Dataset Creation

The baseline paper we referred to stated that they used 47 images of chickenpox and 43 images of monkeypox for their Study One experiment. However, upon reviewing the GitHub repository linked to their work, we discovered that the chickenpox dataset contained only 11 images, resulting in a shortage of 36 images. This discrepancy prompted us to investigate the origin of their dataset. According to the baseline paper, the images were sourced via Google search. In light of this, we decided to search for and collect the missing 36 chickenpox images in order to complete the dataset and proceed with our experiment.

Administrative Details

References

- [1] Centers for Disease Control and Prevention, "*Mpox: About*," Feb. 2024. [Online]. Available: <https://www.cdc.gov/mpox/about/index.html>.
- [2] Md Manjurul Ahsan, Muhammad Ramiz Uddin, Mithila Farjana, Ahmed Nazmus Sakib, Khondhaker Al Momin, Shahana Akter Luna, *Image Data Collection and Implementation of Deep Learning-Based Model in Detecting Monkeypox Disease Using Modified VGG16*, arXiv:2206.01862, June 7, 2022, [Online]. Available: <https://arxiv.org/pdf/2206.01862>.
- [3] Anu V Kottath, Muthukumaran Malarvel, *Comparison Of Deep Learning Models for Monkeypox Disease Detection*, International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI), DOI: 10.1109/ICAEECI58247.2023.10370981, [Online]. Available: <https://ieeexplore.ieee.org/document/10370981>.
- [4] World Health Organisation, "*Mpox: Africa Region*," Aug. 2024. [Online]. Available: <https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON528>.

- [5] Tim Tianyi Yang, Tom Tianze Yang, Andrew Liu, Jie Tang, Na An, Shaoshan Liu, Xue Liu, "AICOM-MP: An AI-based Monkeypox Detector for Resource-Constrained Environments", McGill University, PerceptIn, South China University of Technology, Harvard T.H. Chan School of Public Health, Article, Keywords, Posted Date: September 26th, 2023, doi: <https://doi.org/10.21203/rs.3.rs-3366971>, [Online]. Available: https://www.researchgate.net/publication/374215579_AICOM-MP_an_AI-base_d_Monkeypox_Detector_for_Resource-Constrained_Environments.
- [6] Chiranjibi Sitaula and Tej Bahadur Shahi, "Monkeypox virus detection using pre-trained deep learning-based approaches," arXiv, 2022. [Online]. Available: <https://arxiv.org/abs/2209.04444>.
- [7] Mattia Giovanni Campana, Marco Colussi, Franca Delmastro, Sergio Mascetti, Elena Pagani, *A Transfer Learning and Explainable Solution to Detect mpox from Smartphones images*, *Pervasive and Mobile Computing* Volume 98, 2024, 101874, ISSN 1574-1192, <https://doi.org/10.1016/j.pmcj.2023.101874>. [Online]. Available: <https://arxiv.org/abs/2305.18489>
- [8] World Health Organization, *Diagnostic Testing for the Monkeypox Virus (MPXV): Interim Guidance*, 10 May 2024, WHO Reference Number: WHO/MPX/Laboratory/2024.1, [Online]. Available: <https://www.who.int/publications/i/item/9789240076464>.
- [9] Diagnosis of Visible Diseases Using CNNs, ResearchGate, [Online]. Available: https://www.researchgate.net/publication/360483591_Diagnosis_of_Visible_Diseases_Using_CNNs.

Team Contributions

This project was a collaborative effort, with each team member playing a crucial role in developing and refining our approach to enhancing Monkeypox detection using advanced deep-learning techniques.

Below are the contributions of each team member to date:

- **Amon Muhairwe (amuhairw):** I contributed to writing the sections on Background, Model Description, Dataset, Evaluation Metric, Loss Function, and Results in the mid-term report and this updated report and will continue to expand on these areas in the final report. I have also executed many experiments aimed at replicating the first study experiment from the original research and implementing our approach to improving monkeypox detection performance. Feel free to visit my repo to look at all my experiments conducted during this project.
- **delphine nyabokey (dnyabokey):** Since the initial paper using EfficientNetB0 had a limitation of missing about 300 images, and the architecture used was not clear enough, I went ahead, together with Timothy, and created a `converted_data` folder with original images of monkeypox and chickenpox. This entailed changing files from .jpeg to jpg then to grayscale (RGB) for processing and making them available to the group. The experiments were also performed with a batch size of

10, a learning rate of 0.001, and 35 epochs the [2] mentioned was not yielding the (train) 0.97 precision.

The latest one is the notebook named: `dnyabokey-monkeypox-dataset-trials2.ipynb`, which has numbers closer to those of the main paper [2]. I ran experiments where I changed the optimizer, scheduler, learning rate, and even combined Cross Entropy Loss with Focal Loss to train at 200 epochs and see the results.

These are captured in the Last Stretch folder on GitHub. They're also captured in Weights and Biases individual runs. I conducted various experiments as shown in my GitHub repo to see how the dataset performed on different optimizers, scheduler and loss.

- **Timothy Belekollie(tbelekol):** I took a proactive role by evaluating the baseline model, analyzing its strengths and limitations, and identifying areas for improvement. I proposed using VGG16 as the baseline due to its impressive accuracy (97%) and conducted comparative evaluations against DenseNet201 and ResNet50V2. To address class imbalance, I introduced focal loss combined with categorical cross-entropy. To enhance interpretability, I suggested integrating Grad-CAM for visualizing learned features. Additionally, I contributed to data collection by acquiring 33 missing images to complete the dataset and verified the updated baseline model's performance, and finally, the Future Directions.
- **Aisha Opaluwa (aopaluwa):** I contributed to writing the sections on Discussion, Future Directions, and Conclusion in the report. I have executed experiments aimed at replicating the first study from the original research. Although my initial results did not fully match those published in the paper, I have continued refining these experiments to improve our model's performance. My work has focused on optimizing the experimental approach to achieve better results than the original research

GitHub Repository

Our Group50 established an organization on GitHub named **Delta-Deep-Learning-CMU-Africa** to facilitate our collaborative project. Each member has a repository under this organization, where we managed and documented our respective contributions and experiments.

Below are the links to our specific repositories:

- **Amon Muhairwe's Repository URL:** https://github.com/Delta-Deep-Learning-CMU-Africa/amon_group50
- **delphine nyabokey's Experiments Repository Link:** https://github.com/Delta-Deep-Learning-CMU-Africa/dnyabokey_group_50
- **Aisha Opaluwa's Repository Link:** https://github.com/Delta-Deep-Learning-CMU-Africa/aesh_group50
- **Timothy Belekollie's Repository Link:** https://github.com/Delta-Deep-Learning-CMU-Africa/tbelekol_group_50

Our project Ablations

For detailed analysis and insights into our experiments' ablations conducted during this project, please refer to our Weights & Biases dashboard. You can access the dashboard through the following link: [W&B Project Page](#).