Comparative Analysis of Deep CNN, Transfer Learning, and Proposed Ensemble Architecture for Monkeypox Detection

Md Rofiqul Bari[1], Jannatul Ferdaus[1], Farjana Akter Tonny[1], Tanzina Bithi[1], Talha Zubaer[1] and Amir Sohel[1*]

[1] Department of Computer Science and Engineering,
Daffodil International University
Dhaka, Bangladesh.
{bari15-5598, ferdaus15-4982, tonny15-5965}@diu.edu.bd, bithi9@gmail.com, {zubaer15-5705, amir.cse*}@diu.edu.bd

Abstract. Monkeypox detection and prevention depend on proper and timely diagnosis of the disease. Manual clinical examination is risky for healthcare staff, time-consuming, and costly. Hence, computer-aided diagnosis of monkeypox is highly valuable. In this work, multiple deep learning algorithms are employed to distinguish monkeypox using skin lesion images. The Monkeypox Skin Images Dataset (MSID) contains six classes: Chickenpox, Cowpox, Measles, Hand-Foot-Mouth Disease (HFMD), Monkeypox, and healthy skin. To classify monkeypox, six independent CNNs, six transfer learning models, and a ranked-based ensemble model were used. Among the CNNs, DenseNet201 achieved the highest accuracy of 99.68%, while SeResNet152 obtained the lowest accuracy of 96.67%. For transfer learning models, DenseNet201 again achieved the best performance with 94.53% accuracy, whereas VGG19 yielded the lowest at 63.35%. Finally, the ranked-based ensemble model (DVX), employing DenseNet201, VGG19, and Xception, achieved 100% accuracy, outperforming all individual CNNs and transfer learning approaches. These findings indicate that ensemble deep learning is a highly promising method for automated monkeypox detection on skin images, with strong potential for clinical applications and early disease diagnosis.

Keywords: Monkey Pox, Skin disease, Skin lesion, Comparative Analysis

1 Introduction

Monkeypox (Mpox) is a viral disease that can occur in humans through contact with infected persons. An orthopox virus named monkeypox virus(MPXV) is the cause of monkeypox [1]. This disease has been found in 104 countries worldwide since 2022 [2]. The World Health Organization declared monkeypox a global pandemic on 14 August 2024[3]. The manual detection of monkeypox is time-consuming, risky, costly, and labor-intensive. Computer-aided Diagnosis is used to automatically optimize the detection of monkeypox from images [4]. Computer-Aided Diagnosis (CAD) uses skin

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_4

lesion detection, and Convolutional Neural Network (CNN) based models have become a standard technique for fixing large amounts of images and interpreting problems [5]. Cad depends on deep learning, such as deep CNN, which can effectively find the issue and get reliable outcomes. It gathers the features of deep learning of every CNN model and pooling, and connected layers [6].

Deep Convolutional Neural Networks (D-CNNs) have demonstrated great success in detecting skin lesion diseases, and they allow the effective analysis of large images with well-optimized methods. The most recent works point to the application of space to efficiently train pre-trained CNN models [7], examples of which have succeeded in detecting monkeypox and chickenpox on human skin with high accuracy [8]. CNNs deal with the issues of monkeypox visual diagnosis, and the strategies to optimize the performance, like the Grey Wolf Optimizer (GWO), improve the performance [9]. Hyperparameter optimization and transfer learning have been used to enhance efficiency in hybrid learning models are tailored to achieve superior outcomes in the CNN frameworks [10]. The data in the images are patched, normalized, and passed through convolutional layers with varying filters in order to enhance classification [11]. Particularly, it has been trained, validated, and tested using the “Monkeypox Skin Lesion Dataset"; transfer learning-based techniques allowed the identification of monkeypox on skin lesion pictures accurately [12], which demonstrates that the approaches are effective in timely detection [13].

The ensemble process combines multiple trained models for better accuracy. For detecting monkeypox, using various models gives an accuracy, but using the ensemble technique gives more accurate and more accuracy. It provides a sustainable improvement in the classification of monkeypox [14]. Ensemble learning yields promising results by combining multiple trained models and gives more accuracy than before. It improves the potential of deep learning architectures and improves predictive accuracy. In monkeypox lesion classification, ensemble learning shows a promising result and outperforms other machine learning and deep learning algorithms [15]. Used ensemble classification to analyze new cases, which combines images from three different classifiers they are (a) Layered K-Nearest Neighbors (LKNN), (b) Statistical Naïve Bayes (SNB), and (c) Deep Learning Classifier (DLC) [16].

By understanding the gaps, this work’s goal is to detect monkeypox using six convolutional neural networks (CNN) models: DenseNet201, ResNet152V2, Se-ResNet152, Xception, InceptionV3, VGG19, and to use transfer learning techniques on DenseNet201, ResNet152V2, Se-ResNet152, Xception, InceptionV3, VGG19 to see that transfer learning can improve accuracy or not, and lastly for developing a hybrid ensemble model called DVX(DenseNet-201, VGG-19, Xception) aiming to increase the monkeypox skin disease detection accuracy. However, these are the key contributions and unique aspects of this study:

1. This study evaluates the performance of six CNN architectures—DenseNet201, ResNet152V2, Se-ResNet152, Xception, InceptionV3, and VGG19.

2. Transfer learning is applied to these CNN models to leverage pre-trained knowledge, enhancing feature extraction and improving detection efficiency.
3. A novel ensemble model, DVX, is proposed to address the limitations of individual CNN networks, demonstrating superior classification performance and comparing this with other models. Here, the ensemble models achieve a high detection accuracy of 100% with DVX—thereby contributing valuable insights into effective diagnostic model development.

2 Literature Review

Proposed a model DenseNet201 for the best performance, they found the Accuracy of 97.63%, F1-Score = 90.51%, and Area Under Curve (AUC) is 94.27% in the two-class scenario; and the Accuracy of 95.18%, F1-Score = 89.61% and AUC = 92.06% for the four-class scenario[17]. In a study of Monkeypox detection, VGG(VGG-16, VGG-19), inceptionV3, ResNet(Resnet-50, ResNet101), MobileNet(MobileNetV1, MobileNetV2), DenseNet121, and Xception for identify monkeypox skin disease and get the average Precision, Recall, F1-score, and Accuracy of 85.44%, 85.47%, 85.40%, and 87.13%[18]. In a Nobel monkeypox classification used CNN models named MobilenetV3-s, ResNet50, VGG-19, DenseNet121, and Xception. Here, it gets the highest F1-score, accuracy, and recall simultaneously 98%,99%, and 97% [11]. The Xception CNN model gets the accuracy without space at 97.66%. It enhances the performance of the VGG-19 model and gets an accuracy of 99.87% by using the SpasA optimizations [8]. Introduced a convolutional neural network (CNN) along with a transfer learning model inception (xception), and combined them to get an accuracy ranging from 77% to 88% with or without monkeypox classifications. When they used the Resident Network (ResNet)-101, they got an accuracy range of 84% to 99% [27].

A robust ensemble of CNN models was used by [15]. They used ResNet50, EfficientNetB0, and MobileNetV2 pre-trained CNN models and got an average accuracy of 93%, and by using an ensemble over these models, they got an accuracy of 98%. The top four CNN architectures, EfficientNetB0, ResNet50, Mobile-Net, and Xception, are used for feature extraction. And applied ensemble operation over these models. Here, the proposed ensemble model got an accuracy of 95.45%. Also, the model achieves high precision (95.51%), recall (95.45%), and F1 score (95.46%)[19]. Five pre-trained models: DenseNet121, ResNet152V2, ResNet50, InceptionV3, and EfficientNetV2B3. The gained accuracy scores for the MD-2022 dataset were 89.4%, 84.2%, 89.4%, 84.2%, and 84.2%, respectively, and for the MSID dataset, the accuracy of DenseNet121, ResNet50, InceptionV3, EfficientNetV2B3, and ResNet152V2, respectively 97.4%, 96.2%, 93.6%, 93.6%, and 95%. Here, after using the ensemble, the "Monkeypox-dataset-2022" was 89.4%, and for "Monkeypox Skin Images Dataset (MSID)," it was 98.7% [20]. A customized CNN model named MpoxSLDNet. This customized CNN model (MpoxSLDNet) got a validation accuracy of 94.56%, compared to pre-

trained CNN models VGG16, DenseNet121, and ResNet50, respectively 86.25%, 84.38%, and 67.19% [21]. A custom deep convolutional neural network named the MonkeyPoxNet model was introduced by [22]. The MonkeyPoxNet model was used for prediction, comparing the MonkeyPox Virus and normal disease classes by comparing test features with the trained model. Remarkably, the proposed MonkeypoxNet demonstrated an accuracy of 99.06%, sensitivity of 98.66%, specificity of 99.11%, and an F-measure of 99.67%. And it gives better accuracy than the pre-trained Deep CNN models. BinaryDnet53 is a customized lightweight Binarized CNN model introduced by [23] for detecting MonkeyPox Skin problems. They assume that their model was 20 times more computationally efficient and 2 times more efficient than the SOTA (State of the Art) CNN.

Author	Dataset	Model &Accuracy	Limitations
Sitaula et al. 2022	Monkeypox Skin Lesion Dataset (MSLD)	VGG-16,VGG-19, inceptionV3, Resnet-50, ResNet101, MobileNetV1, MobileNetV2, DenseNet121and,Xception Avg accuracy is 87.13%	Accuracy is not on target.
Sorayaie Azar et al., 2023	Monkeypox, Chickenpox, Measles, and Normal cases	DenseNet201 (97.63%,94.27, and 95.18%)	Accuracy decreases when changing classes.
Ahsan et al. (2023)	Monkeypox dataset	Transfer learning based on Inception (Xception) (77% to 88%)	Limited data class
Saavedra et al. (2023)	Public Dataset of monkeypox	Ensemble (98%)	The number of images is low
Haque et al. (2023)	MD-2022, MSID	Ensemble (89.4%)	Accuracy is not good.
Abdelrahim et al. (2024)	Monkeypox Skin Lesion Dataset (MSLD)	Ensemble (95%)	Lack of multimodal data.
Dihan et al. (2024	Monkeypox Skin Lesion Dataset	MpoxSLDNet (94.56%)	A diverse range of lesion variations.
Pratama et al. 2025	Monkeypox Skin Lesion Dataset (MSLD)	Feature Fusion with Albumentation (85.96%)	Accuracy is low.

Table 1 describes the key summary of some recent work.

Despite high accuracies reported by existing studies, most approaches face limitations such as reduced performance when class distributions change, a lack of multimodal data, limited dataset validity, and an inability to handle diverse lesion variations. Additionally, many models rely on only a few classes or datasets, indicating the need for more robust, generalized solutions for monkeypox detection.

3. Methodology

The testing for this study was carried out using the Keras Library and Google Colab. Since TensorFlow is one of the best deep learning libraries for Python, it was used to deal with machine learning techniques. The working flowchart for our work is displayed in Figure 1.

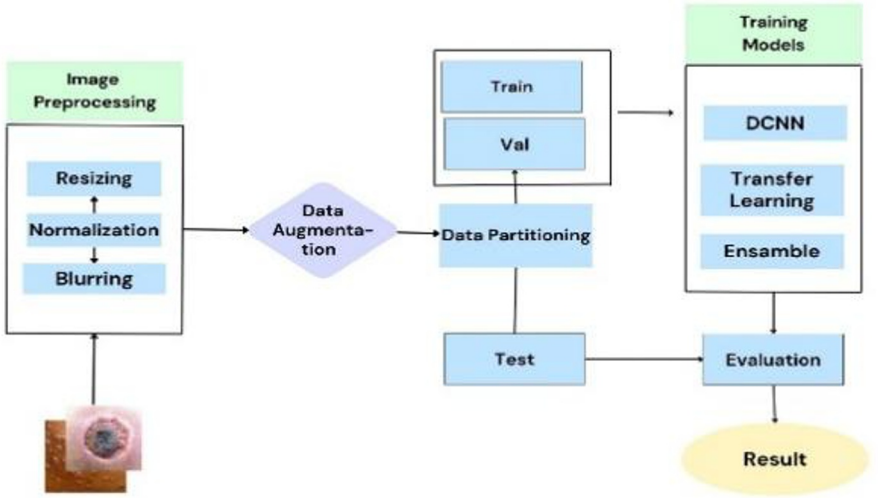

Figure 1: Process of the study

The experiment begins with image preprocessing. After preprocessing, data augmentation is applied to generate additional image variations, which helps improve the robustness and generalization of the models. The dataset is then partitioned into training, validation, and testing subsets for model development and performance assessment. Next, the training phase involves applying different approaches such as Deep Convolutional Neural Networks (DCNN), Transfer Learning, and Ensemble methods. Finally, the trained models are evaluated on the test set, and the best-performing model is selected to provide the final result.

3.1.Datasets

For this research, the dataset [32] was collected from a public shared repository. The data set has six classes (Chicken Pox, Cow Pox, healthy, HFMD, Measles, Monkey Pox). The images were captured with a camera and stored in JPG files and RGB format. Figure 1 displays samples of images used in the study. In the dataset, the classes Chickenpox have 77 images, Cowpox have 66 images, Healthy have 115 images, Measles have 169 images, and Monkeypox have 284 real images of human infected images. Figure 2 visualizes the sample image of our dataset.

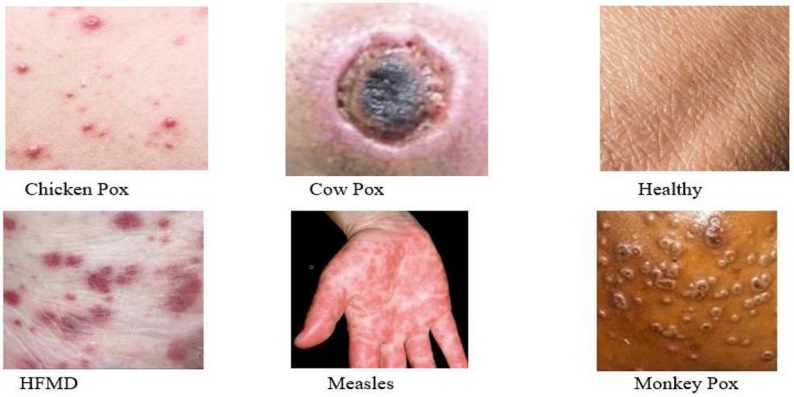

Figure 2: Samples of the image dataset

3.2.Image Acquisition and Pre-processing

At this step, the downloaded image dataset, which was downloaded from Kaggle, was checked manually to see if any irrelevant images existed. we also checked if there were any white background images. If found, these types of images were removed from the data set. And carefully named the classes in short form. Also checked that all images are in same size of 256 x 256. Then, from the images, unnecessary portions were removed by cropping, and some of them needed to be rotated. Here, checked images of skin spots are clearly visible. If spots of the images were not clearly visible, that image was removed from the data set. Here, we make the data set balanced by using SMOTE.

3.3.Image Augmentations and Data Partitioning

In this phase of work, we used position augmentation, like scaling, rotation, flipping, color augmentation, and pixel intensity. In color augmentation, change brightness, contrast, and saturation. In horizontal flipping, it randomly flips images with a 50% rotation. Using a Gaussian blur for better performance of the model. This used a standard deviation of 0 to 0.5, and the value was chosen randomly. Additive Gaussian noise was used with a scale between 0 to 5%. It also randomly changes the contrast of the images in the range of 0.75 to 1.5. This way, augmented images from the original image have been created. Figure 2 shows the augmented images of the data set used in this study. After the augmentation of data, we get a balanced data set. This balanced data set is divided into Train, Test, and Val for analyzing the data. In this procedure, it is divided into 70% for training, 15% for testing, and 15% for the validation data folder. Here, a total of 18000 images in 6 classes in every class, there were 6000 images. After partitioning the data into three folders, it gave 12600 in train, 2700 in test, and 2700 in the val folder for the next steps.

3.4.Classification Techniques

In this step, DenseNet201, ResNet152V2, Se-ResNet152, Xception, InceptionV3, and VGG19 were used to automatically detect monkeypox Skin disease. After that, we used transfer learning models. In this step, we used DenseNet201, ResNet152V2, Se-

ResNet152, Xception, InceptionV3, and VGG19 models for the transfer learning approach. The models were chosen as classification tools due to their well-known technique of being a perfect, accurate classifier for any real applications. Here, after training the model, the evaluation model was built for monkeypox skin disease detection based on their height accuracy of its existence. After that, the images were classified into different disease classes using a SoftMax output layer.

Our main objective in this work was to create ensemble models that could identify and classify monkeypox skin lesions accurately. The key motivation for establishing the ensemble model is that even if a weak classifier makes an incorrect prediction, the entire ensemble classifier (robust classifier) may rectify the error.

In addition, the ensemble technique may minimize variance. The ensemble model employed in this work aggregates the Sum of Probability data from three CNNs (DenseNet201, VGG-19, and Xception) and estimates the sum of probabilities for each class based on the individual CNN design. The final forecast is based on the class with the highest normalized total.

Our proposed framework uses the ensemble function $f(n)$, where $n = 6$. Each class receives n confidence values for a given image I. The final categorization choice is made using the classes' greatest probability. The confidence values S_{ij} , where $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$, are aggregated using the Sum of Probabilities. A normalization factor $\sum_{i=1}^{m}\sum_{j=1}^{n} s_{ij}$ is used to normalize values after the summation of corresponding class values $\sum_{j=1}^{n} s_{ij}$

$$x = \frac{\sum_{j=1}^{n} s_{ij}}{\sum_{j=1}^{m}\sum_{j=1}^{n} s_{ij}}$$

4. *Results and discussion of experiments*

This study applied three different approaches for detecting and identifying skin lesion images using pre-trained deep learning models and discovering the best model (Transfer Learning and Ensemble) that shows results.

4.1. Performance metrics

The experimental outcomes are measured using the following machine learning classification model performance metrics:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision}*\text{Recall}}{\text{Precision+Recall}} \quad (4)$$

Data loss curves and confusion matrices have also been used to measure the performance of the models.

4.1.1. Experimental Setup

This experiment trained deep learning models for 250 epochs with the Adam optimizer at a learning rate of 0.0001 and categorical cross-entropy loss. It uses ReLU activations with softmax in the final layer.

Table 2. Experimental setup of the experimented algorithms.

Parameters	Value
Epoch	250(patients = 10)
Optimizers	Adam
Learning rate	0.0001
Activation function	ReLU, softmax (last layer)
Entropy	Categorical cross-entropy

4.2. Experiment 1- Performance of the Original CNN network for monkeypox Detection

This section presents the results of six original CNN networks: DenseNet201, ResNet152V2, Se-ResNet152, Xception, InceptionV3, and VGG19. The DenseNet-201 model achieved the maximum accuracy of 99.68%, while the SeresNet152 model had the lowest accuracy of 96.67%. Table 3 shows the overall performance of CNN models.

Table 3: Performance of pre-trained CNN models.

Model	Precision	Recall	F1 Score	**Accuracy**
DenseNet201	100%	100%	100%	99.68%
Xception	99%	99%	99%	98.90%
InceptionV3	100%	100%	100%	99.60%
SeResNet152	97%	97%	97%	96.67%
ResNet152V2	99%	99%	99%	99.17%
VGG19	99%	99%	99%	98.43%

According to Table 3, the precision values for each model on the test dataset are considered; the DenseNet-201 architectures provide the best performance. According to the above table, the VGG-19, Xception, ResNet152V2, and InceptionV3 models correctly detected and classified monkeypox skin disease compared to other models. However, Se-ResNet152 performed poorly, with the lowest identification.

Also, from the confusion matrix of the original CNNs. Densenet201 provides a better result, as expected. A total of 428,428,421,425,421 and 427 images were correctly classified using Densenet201. Also, in ResNet152V2 provides a good performance with 429, 433, 423, 431,425, and 419.

In the first approach, a set of six deep CNN networks (DenseNet201, ResNet152V2, Se-ResNet152, Xception, InceptionV3, and VGG19) was applied on 18000 skin lesion

images of six different classes (Chicken Pox, Cow Pox, Healthy, HFMD, Measles, Monkey Pox). From these models performance, DeseNet-201 performs the best with the accuracy of 99.68% and Se-resNet152 gives the lowest accuracy of 96.67%. The models are aligned with [24], [17], [25] that DenseNet201 had a high accuracy percentage. DenseNet gave the best accuracy because its design allows each layer to draw collective knowledge of all used layers. And when it receives inputs from every layer, then it passes the inputs forward. This dense connectivity fosters features for reuse, then it improves the gradient flow and enhances the parameter and efficiency. Also, its boosts the accuracy in the classification of images. Moreover, DenseNet is best at capturing fine-grained details and representing objects hierarchically across various parameters [26].

4.3. Experiment 2: Transfer Learning CNN Network Accuracy in Detecting Monkeypox Skin Lesions

In this experiment, different transfer learning models were evaluated to compare their performance. Six transfer learning CNN architectures' performance is presented in this section. SE-ResNet152, InceptionV3, VGG19, Resnet152V2, Xception, and DenseNet-201 models all had high accuracies in the test sets, as shown in Table 4. Among them, DenseNet201 achieved the best results with the highest accuracy and F1 Score, showing its strength in feature reuse and efficient learning. With an accuracy of 94.53%, the DenseNet-201 was the most accurate. And with the accuracy of 63.35% VGG-19 was the least accurate transfer Learning Model. The decrease from the original CNNs is significant for transfer learning.

Xception, SeResNet152, and ResNet152V2 also performed competitively with balanced precision and recall, making them reliable choices for image classification. InceptionV3 showed moderate performance, while VGG19 performed the worst due to its older architecture and limited feature extraction ability. Table 4 shows the precision, Recall, and F1-score findings from CNN networks incorporating transfer learning.

Table 4: Performance of CNN networks incorporating transfer learning.

Model	Precision	Recall	F1 Score	**Accuracy**
DenseNet201	96%	94%	95%	94.5%
Xception	94%	93%	93%	93.35%
InceptionV3	92%	92%	92%	92.46%
SeResNet152	94%	94%	94%	94.02%
ResNet152V2	94%	94%	94%	93.37%
VGG19	67%	64%	64%	63.35%

Generally, a model with high Precision, Recall, and support is superior. With a 63%, the trial results show that VGG19 has a low precision in Monkeypox disease. Resnet152V2 has the highest precision.

Here, the second approach was transfer learning, which offers less accuracy in monkeypox detection than advanced deep learning and machine learning models. In

general, transfer learning increases the accuracy of the proposed models, but this study found negative results. Here are all the models that agree with [27], [13], [26], and [11] studies. Transfer learning (TL) gives negative results because of the noisy accumulation of learning steps. The negative result of Transfer Learning (TL) affects

the performance when a large amount of training data is used in the models. The image data set differs from the train data set this also a reason for the negative transfer learning result, [28].

4.4. Experiment 3: Ensemble model development

Ranked-Based Ensemble Model DVX (DenseNet201, VGG19, Xception)

Table 5: Performance of CNN networks with ensemble techniques (n=6)

Ensemble Model (DenseNet201, VGG19, Xception)						
	Chicken Pox	Cow Pox	Healthy	HFMD	Measles	Monkey Pox
Precision	100%	100%	100%	100%	100%	100%
Recall	100%	100%	100%	100%	100%	100%
F1-score	100%	100%	100%	100%	100%	100%
Support (N)	420	420	420	420	420	420

In this experiment, DenseNet201 (99.68%), VGG-19 (98.4%), and Xception (98.91%) were selected as the candidates for the ensemble model 'DVX' (see Table 2). For getting a substantial improvement in classification accuracy and potential of advanced deep learning architecture, these Deep CNN models are combined [16]. Figure 3 represents the confusion matrix of proposed ensemble model.

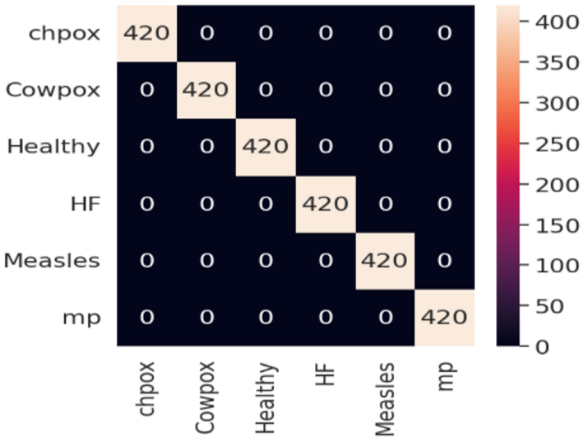

Figure 3: Confusion Matrix of Ensemble model.

The confusion matrix suggests that 420 (TP) Chicken Pox and 420 (TP) Cow Pox, 420 (TP) healthy, 420 (TP) HFMD, 420 (TP) Measles, and 420 (TP) Monkey Pox were correctly classified. Moreover, the False positives and False negatives were less, resulting the higher model accuracy. This model gave the best accuracy of 100%.

In the third approach, a ranked-based ensemble model DVX (DenseNet-201, VGG-19, and Xception) was applied, and it yielded highly promising results. The DVX ensemble achieved 100% accuracy, which clearly demonstrates the effectiveness of combining diverse architectures through a ranking mechanism. DenseNet-201 contributed deep feature reuse, VGG-19 captured spatial hierarchies, and Xception enhanced representation with depthwise separable convolutions. [29], [30], and [31] show ensemble of CNN models gives a promising accuracy on monkeypox skin lesion classification.

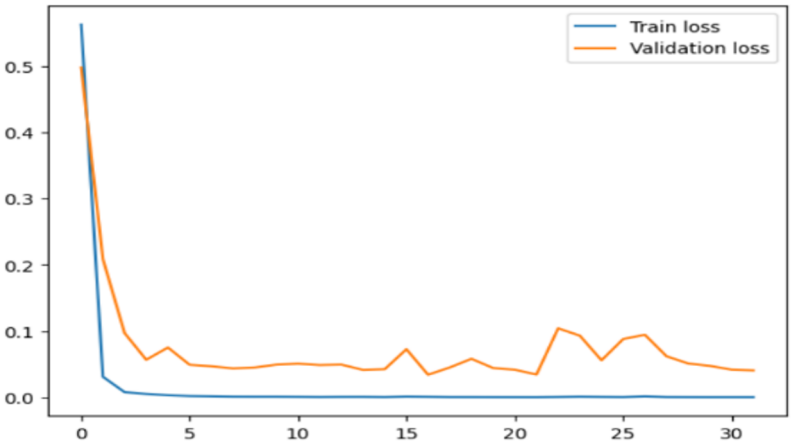

Figure 4: Train and Validation Loss Curve

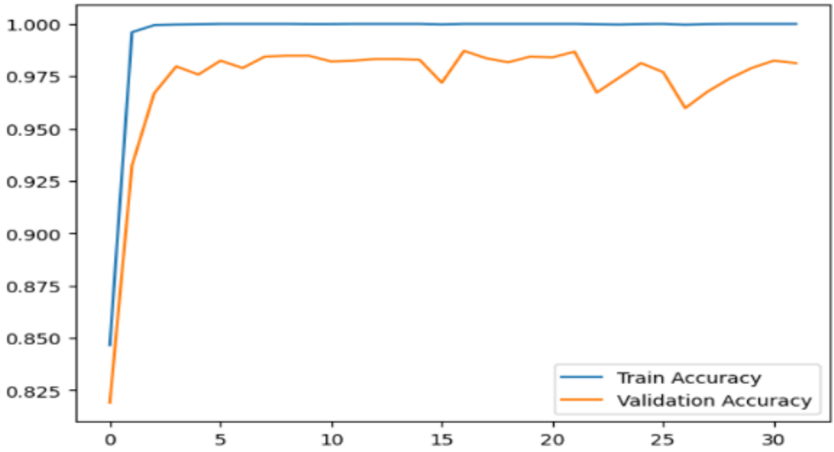

Figure 5: Train and Validation Accuracy Curve

Overall, the three approaches highlight clear differences in performance and applicability. Among the individual pre-trained models, DenseNet-201 stood out with 99.68% accuracy, confirming its strong feature reuse and gradient flow advantages, while SeResNet152 performed the weakest. In contrast, transfer learning unexpectedly produced negative results, likely due to dataset mismatch and the large size of the available training set, which reduced the benefits of knowledge transfer and even harmed performance. The ranked-based ensemble (DVX) outperformed both, achieving a perfect 100% accuracy by combining the complementary strengths of DenseNet-201, VGG-19, and Xception through a ranking mechanism that minimized misclassifications. Taken together, these results suggest that while transfer learning may not always be effective for large domain-specific datasets, DenseNet-201 and ensemble learning—particularly ranked-based ensembles—offer the most reliable and accurate solutions for skin lesion classification.

4.5. K-fold Cross-validation

We use k-fold cross-validation to make sure the model is tested on multiple different splits of the data, instead of relying on a single train-test split that may give misleading results. By training on $k - 1$ k−1 parts and testing on the remaining part repeatedly, we can check whether the model performs consistently on unseen data. If the performance varies widely between folds, it indicates overfitting. This method also ensures that the entire dataset is used for both training and validation, improving model generalization. Overall, k-fold helps detect and reduce overfitting by giving a more reliable estimate of real-world performance.

Table 6- Fold Wise Classification report.

Fold	Precision	Recall	F1 Score	Accuracy
Fold-1	100%	100%	100%	100%
Fold-2	100%	99%	100%	99.67%
Fold-3	100%	100%	100%	100%
Fold-4	99%	100%	99%	99.33%
Fold-5	100%	100%	100%	100%

The table presents the fold-wise performance metrics obtained through k-fold cross-validation, where k =5, including Precision, Recall, F1-Score, and Accuracy. All folds achieved exceptionally high and consistent results, with most folds reaching 100% across all metrics, and in only two folds, we obtained accuracy in between (99–100%) in Fold-2 and Fold-4. This consistency across multiple independent validation splits demonstrates that the model performs reliably on different subsets of the data rather than memorizing the training samples. Therefore, the results indicate that the model is not overfitting and generalizes well to unseen data.

4.6. Explainable AI

In our study, two popular explainable AI techniques, LIME and SHAP, were used to generate local and global explanations for the deep learning model's predictions on the validation and test data sets. LIME creates an explainable model by applying a deep

learning model around the prediction point. And, in this process, SHAP provides a unified framework for feature importance estimation.

4.6.1. Lime

LIME is a known technique for generating an interpretation for every prediction generated by a trained model. It creates a local linear model around the prediction point and weights the input features to estimate their importance in the prediction. We used the Lime package in Python to generate explanations for our model's predictions on the validation and test sets.

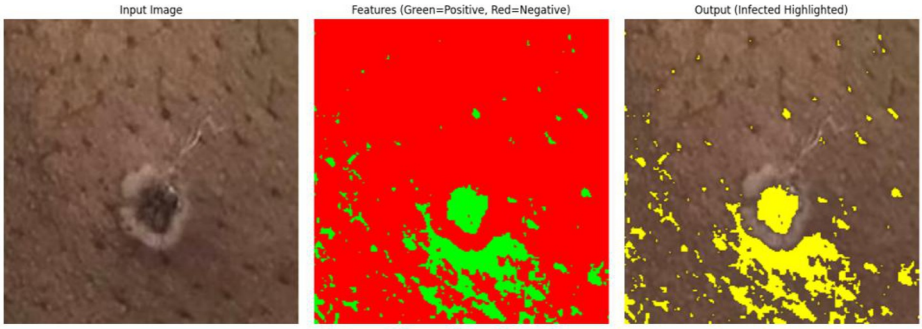

Figure 6. A lime explanation of the DVX ensemble model

As such, LIME is used for generating an explanation for each prediction. Figure 6 shows the explanation of LIME. Here, includes two features using color Green and Red. "Green regions" represent areas that positively contributed to the predicted class, and "Red regions" represent areas that negatively contributed to the predicted class.

4.6.2. SHAP

Here, in Figure 7, SHAP visualization, red regions indicate areas of the image that positively influenced the model's predicted class, while blue regions represent areas that pushed the prediction in the opposite direction. The brightness or intensity of each color reflects how strongly that region contributed to the final decision. This helps identify which parts of the image had the greatest impact on the model's output.

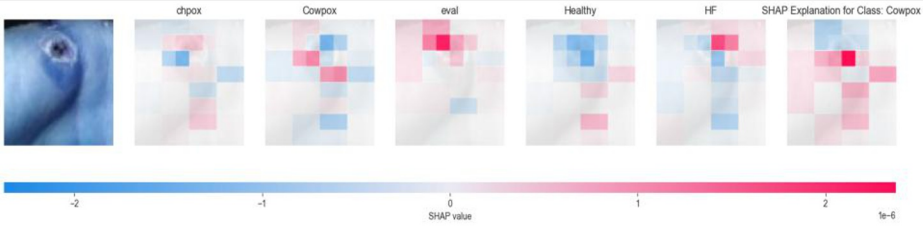

Figure 7. SHAP Explanation of the DVX model

5. Conclusion

In this study, we applied different deep learning methods for detecting skin lesions in the diagnosis of monkeypox disease, where ensemble approaches achieved higher

accuracy, recall, precision, and F1 scores compared to individual models. The ranked-based ensemble model, combining DenseNet201, VGG19, and Xception, achieved the best accuracy, proving the effectiveness of integrating complementary features from diverse architectures. However, limitations remain, such as the use of a non-clinically certified dataset, which may not capture the diversity of real-world cases, including variations in skin tone, lighting conditions, lesion severity, and image quality across devices, potentially affecting generalizability. Moreover, advanced fine-tuning or domain adaptation was not explored, and the lack of explainable AI methods reduces interpretability and clinical confidence. The models have also not yet been validated with dermatologist feedback in real medical settings, and the high computational demand of deep CNNs could be a barrier in low-resource healthcare systems. Despite these challenges, our findings demonstrate that computer-aided diagnosis using ensemble deep learning models is a promising solution for early-stage monkeypox detection, crucial for preventing widespread outbreaks. Future work should focus on collecting clinically certified, large-scale, and diverse datasets, integrating more explainable AI to enhance clinical trust, applying advanced fine-tuning techniques, and optimizing lightweight models to enable deployment in low-resource environments, making the system more reliable and practically applicable in healthcare.

References

1. Kameli, N., Algaissi, A., Taha, M. M. E., Alamer, E., Alhazmi, A., Hakami, W., ... & Abdelwahab, S. I. (2024). Monkeypox Global Research: A Comprehensive Analysis from Emergence to Present (1961-2023) for innovative prevention and control approaches. *Journal of Infection and Public Health*, 102593.
2. Harapan, H., Ophinni, Y., Megawati, D., Frediansyah, A., Mamada, S. S., Salampe, M., ... & Sallam, M. (2022). Monkeypox: a comprehensive review. *Viruses*, *14*(10), 2155.
3. www.who.int/news/item/14-08-2024-who-director-general-declares-mpox-outbreak-a-public-health-emergency-of-international-concern (14 august 2024).
4. Khafaga, D. S., Ibrahim, A., El-Kenawy, E. S. M., Abdelhamid, A. A., Karim, F. K., Mirjalili, S., ... & Ghoneim, M. E. (2022). An Al-Biruni earth radius optimization-based deep convolutional neural network for classifying monkeypox disease. *Diagnostics*, *12*(11), 2892.
5. Bakkouri, I., & Afdel, K. (2020). Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images. Multimedia Tools and Applications, 79(29), 20483-20518
6. Attallah, O. (2024). Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning. *Computers in Biology and Medicine*, *178*, 108798.
7. Bamaqa, A., Bahgat, W. M., AbdulAzeem, Y., Balaha, H. M., Badawy, M., & Elhosseini, M. A. (2024). Early detection of monkeypox: Analysis and optimization of pretrained deep learning models using the Sparrow Search Algorithm. *Results in Engineering*, *24*, 102985.
8. Uzun Ozsahin, D., Mustapha, M. T., Uzun, B., Duwa, B., & Ozsahin, I. (2023). Computer-aided detection and classification of monkeypox and chickenpox lesion in human subjects using deep learning framework. Diagnostics, 13(2), 292.
9. Eliwa, E. H. I., El Koshiry, A. M., Abd El-Hafeez, T., & Farghaly, H. M. (2023). Utilizing convolutional neural networks to classify monkeypox skin lesions. *Scientific reports*, *13*(1), 14495.

10. Altun, M., Gürüler, H., Özkaraca, O., Khan, F., Khan, J., & Lee, Y. (2023). Monkeypox Detection Using CNN with Transfer Learning. Sensors (Basel, Switzerland), 23(4), 1783. https://doi.org/10.3390/s23041783
11. Chakraborty, C., Achar, U., Nayek, S., Achar, A., & Mukherjee, R. (2024). CAD-PsorNet: deep transfer learning for computer-assisted diagnosis of skin psoriasis. *Scientific Reports*, *14*(1), 26557.
12. Thorat, R., & Gupta, A. (2024). Transfer learning-enabled skin disease classification: the case of monkeypox detection. *Multimedia Tools and Applications*, 1-19.
13. Jaradat, A. S., Al Mamlook, R. E., Almakayeel, N., Alharbe, N., Almuflih, A. S., Nasayreh, A., ... & Bzizi, H. (2023). Automated monkeypox skin lesion detection using deep learning and transfer learning techniques. *International Journal of Environmental Research and Public Health*, *20*(5), 4422.
14. Muñoz-Saavedra, L., Escobar-Linero, E., Civit-Masot, J., Luna-Perejón, F., Civit, A., & Domínguez-Morales, M. (2023). A robust ensemble of convolutional neural networks for the detection of monkeypox disease from skin images. *Sensors*, *23*(16), 7134.
15. Akram, A., Jamjoom, A. A., Innab, N., Almujally, N. A., Umer, M., Alsubai, S., & Fimiani, G. (2024). SkinMarkNet: an automated approach for prediction of monkeyPox using image data augmentation with deep ensemble learning models. *Multimedia Tools and Applications*, 1-17.
16. Sorayaie Azar, A., Naemi, A., Babaei Rikan, S., Bagherzadeh Mohasefi, J., Pirnejad, H., & Wiil, U. K. (2023). Monkeypox detection using deep neural networks. *BMC Infectious Diseases*, *23*(1), 438.
17. Sitaula, C., & Shahi, T. B. (2022). Monkeypox virus detection using pre-trained deep learning-based approaches. *Journal of Medical Systems*, *46*(11), 78.
18. Abdelrahim, E. M., Hashim, H., Atlam, E. S., Osman, R. A., & Gad, I. (2024). TMS: Ensemble Deep Learning Model for Accurate Classification of Monkeypox Lesions Based on Transformer Models with SVM. *Diagnostics*, *14*(23), 2638.
19. Haque, R., Sultana, A., & Haque, P. (2023, May). Ensemble of fine-tuned deep learning models for monkeypox detection: A comparative study. In *2023 4th International Conference for Emerging Technology (INCET)* (pp. 1-8). IEEE.
20. Dihan, F. J., Murad, S. A., Muzahid, A. J. M., Uddin, K. M., Alenazi, M. J., Bairagi, A. K., & Biswas, S. (2024). MpoxSLDNet: A Novel CNN Model for Detecting Monkeypox Lesions and Performance Comparison with Pre-trained Models. *arXiv preprint arXiv:2405.21016*.
21. Chintamaneni, V., Krishna, B. H., Suresh, M., Bukkaptanm, K., Sujatha, C. N., Swaraja, K., & Kumar, P. M. (2024). Deep Learning-Based Diagnostic Model for Automated Detection of Monkeypox: Introducing MonkeypoxNet. *Traitement du Signal*, *41*(1).
22. Biswas, D., & Tešić, J. (2024). Binarydnet53: a lightweight binarized CNN for monkeypox virus image classification. *Signal, Image and Video Processing*, 1-12.
23. Demir, F. B., Baygin, M., Tuncer, I., Barua, P. D., Dogan, S., Tuncer, T., ... & Acharya, U. R. (2024). MNPDenseNet: automated monkeypox detection using multiple nested patch division and pretrained densenet201. *Multimedia Tools and Applications*, 1-23.
24. Bala, D., Hossain, M. S., Hossain, M. A., Abdullah, M. I., Rahman, M. M., Manavalan, B., ... & Huang, Z. (2023). MonkeyNet: A robust deep convolutional neural network for monkeypox disease detection and classification. *Neural Networks* , *161*, 757-775.
25. Li, Y., Zhang, L., Wang, X., & Liu, Q. (2021). Leveraging DenseNet for Microscopic Image Classification. Journal of Advanced Microscopy Techniques, 23(2), 112-125.
26. Ahsan, M. M., Alam, T. E., Haque, M. A., Ali, M. S., Rifat, R. H., Nafi, A. A. N., ... & Islam, M. K. (2024). Enhancing monkeypox diagnosis and explanation through modified transfer learning, vision transformers, and federated learning. *Informatics in Medicine Unlocked*, *45*, 101449.
27. Gui, L., Xu, R., Lu, Q., Du, J., & Zhou, Y. (2018). Negative transfer detection in transductive transfer learning. *International Journal of Machine Learning and Cybernetics*, *9*, 185-197.
28. Pratama, N. R., Setiadi, D. R. I. M., Harkespan, I., & Ojugo, A. A. (2025). Feature Fusion with Albumentation for Enhancing Monkeypox Detection Using Deep Learning Models. *Journal of Computing Theories and Applications*, *2*(3), 427-440.
29. Haque, R., Sultana, A., & Haque, P. (2023, May). Ensemble of fine-tuned deep learning models for monkeypox detection: A comparative study. In *2023 4th International Conference for Emerging Technology (INCET)* (pp. 1-8). IEEE.
30. Üzen, H., & Fırat, H. (2024). Deep learning-based adaptive ensemble learning model for classification of monkeypox disease. Konya Journal of Engineering Sciences, 12(4), 822-837.

31. Kumar Saha, D., Rafi, S., Mridha, M. F., Alfarhood, S., Safran, M., Kabir, M. M., & Dey, N. (2025). Mpox-XDE: an ensemble model utilizing deep CNN and explainable AI for monkeypox detection and classification. BMC Infectious Diseases, 25(1), 403.
32. www.kaggle.com/datasets/nafin59/monkeypox-skin-lesion-dataset