Hybrid Approach of Classification of Monkeypox Disease: Integrating Transfer Learning with ViT and Explainable AI

MD Abu Bakar Siddick[1], Zhang Yan*[2], Mohammad Tarek Aziz[3], Md Mokshedur Rahman[4], Tanjim Mahmud[5], Sha Md Farid[6], Valisher Sapayev Odilbek Uglu[7], Matchanova Barno Irkinovna[8], Atayev Shokir Kuranbaevich[9] Ulugbek Hajiev[10]

Department of Computer Science and Technology, Beijing Institute of Technology (BIT), Beijing, China[1,2,4]
Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Bangladesh[3]
Department of Computer Science and Engineering, Rangamati Science and Technology University, Bangladesh[5]
Department of Technology, Wilmington University, Delaware, United States[6]
Department of General-Professional Science, Mamun University, Khiva, Uzbekistan[7]
Urgench State Pedagogical Institute, Khorezm, Uzbekistan[8]
Urgench State University, Khorezm, Uzbekistan[9,10]

***Abstract*—Human monkeypox is a persistent global health challenge, ranking among the most common illnesses worldwide. Early and accurate diagnosis is critical to developing effective treatments. This study proposes a comprehensive approach to monkeypox diagnosis using deep learning algorithms, including Vision Transformer, MobileNetV2, EfficientNetV2, ResNet-50, and a hybrid model. The hybrid model combines ResNet-50, MobileNetV2, and EfficientNetV2 to reduce error rates and improve classification accuracy. The models were trained, validated, and tested on a specially curated monkeypox dataset. EfficientNetV2 demonstrated the highest training accuracy (99.94%), validation accuracy (97.80%), and testing accuracy (97.67%). ResNet-50 achieved 99.87% training accuracy, 99.85% validation accuracy, and 97.18% testing accuracy. MobileNetV2 reached 95.47% training accuracy, with validation and testing accuracies of 79.51% and 78.18%, respectively. Designed to mitigate overfitting, the Vision Transformer achieved 100% training accuracy, 87.51% validation accuracy, and 99.41% testing accuracy. Our hybrid model yielded 99.33% training accuracy and 99.09% testing accuracy. The Vision Transformer emerged as the most promising model due to its robust performance and high accuracy, followed closely by the hybrid model. Explainable AI (XAI) techniques, such as Grad-CAM, were applied to enhance the interpretability of predictions, providing visual insights into the classification process. The results underscore the potential of Vision Transformer and hybrid deep learning models for accurate and interpretable monkeypox diagnosis.**

***Keywords*—Monkeypox; vision transformer; hybrid model; transfer learning; explainable artificial intelligence**

I. Introduction

In order to manage outbreaks, it is essential to detect monkeypox accurately and promptly. Vision transformer models, transfer learning[1], [2], and deep learning [3], [4] provide effective ways to improve diagnostic accuracy from image data. Clinical examination and laboratory testing are frequently used in traditional diagnostic techniques, however, they can be time-consuming and less available in remote or underdeveloped places. Because models can be taught to accurately identify the distinctive characteristics of monkeypox lesions, deep learning [5], [6] makes monkeypox detection quicker and more scalable [7]. Transfer learning, which enables models to use pre-trained weights from sizable datasets, is particularly advantageous when it comes to monkeypox because there is a dearth of labeled data. Even with a limited monkeypox dataset [8], accurate models may be deployed thanks to this technique, which builds on previously obtained knowledge to enable faster training and higher accuracy. The detection method is further improved by the employment of vision transformers, which are renowned for their capacity to grasp intricate spatial relationships by processing images as patches. Vision transformers, as opposed to conventional convolutional networks [3], [9], [10], are able to identify global patterns in the image, which enables the model to concentrate on certain lesion features that could otherwise go unnoticed. This capacity is especially crucial for preventing misdiagnosis, improving patient outcomes, and distinguishing monkeypox from other skin disorders that look similar. When combined, these deep learning techniques [11] not only provide a more affordable and easily available diagnostic option, but they also aid in early detection and containment initiatives, which has a big influence on public health by enabling quick action in the event of an outbreak.

The main objective of image-based monkeypox disease detection is to facilitate early, precise, and easily accessible diagnosis, which is crucial for efficient outbreak management and patient treatment. Healthcare professionals can swiftly detect monkeypox lesions by using image-based AI models, which enables the early isolation and treatment of infected people to stop the spread of the disease [12], [13]. Another important goal is to provide diagnostic tools in settings with low resources and remote locations, where access to standard lab-based testing may be difficult [14], [15]. To avoid misdiagnosis and guarantee that patients receive the right care, models that can reliably differentiate monkeypox from other comparable skin disorders, including chickenpox or measles,

must be developed. Since there is a dearth of picture data related to monkeypox, it is crucial to use transfer learning [16] to build trustworthy models from small datasets. This will enable pre-trained models to identify distinctive characteristics of monkeypox. Additionally, by offering a scalable method for monitoring and forecasting outbreaks, AI-based monkeypox detection can help public health initiatives by facilitating prompt reaction and containment [17]. These goals support a strong and workable approach to controlling monkeypox, boosting readiness for infectious disease risks, and improving health outcomes.

The main contributions of this study are:

1) Improved model resilience, a bespoke dataset was created using random images that simulated a variety of real-world situations.
2) We enhanced feature extraction and detection by using ResNet-50's deep residual connections to categorize monkeypox lesions with accuracy.
3) MobileNetV2 was utilized for lightweight, effective detection, making diagnostics accessible on mobile or low-resource devices in remote locations.
4) EfficientNetV2 was used to maximize efficiency by striking a balance between decreased computation and detection accuracy for effective model operation.
5) We enhanced the model's capacity to distinguish monkeypox from related skin disorders by using Vision Transformer to gather global and patch-based picture information.
6) We created a hybrid model with enhanced diagnostic accuracy and dependability by combining ResNet-50, MobileNetV2, and EfficientNetV2.
7) We interpreted model predictions using explainable AI methodologies, offering visual justifications for clear, reliable diagnostics.

In this study, the previous research on existing work is described in Section II, and the proposed working framework is then detailed in Section III. The later Section IV denotes the result analysis, and finally, the conclusion and future plan are explained in the last Section V of this paper.

II. Previous Studies

In recent years, researchers have been trying to prevent the monkeypox disease and they are finding different types of solutions. Hence, some related publications are found online about the solutions to human monkeypox detection. We mentioned and explained some of them.

Bala et al. [18] proposed a deep CNN-based monkeypox disease detection system. Their study summary is that MonkeyNet, a novel deep learning-based model, was created to identify monkeypox from skin images. Its accuracy was 93.19% on the original dataset and 98.91% on an augmented dataset. In order to facilitate model training and testing, this study made the "Monkeypox Skin Images Dataset (MSID)" publicly available. Grad-CAM graphics help doctors diagnose monkeypox early and accurately by highlighting affected areas.

Dahiya et al. [19] explained monkeypox disease detection using a deep learning model. They used CNN and YOLO V5. Using the monkeypox dataset online, they obtain 98.18% accuracy in image classification.

Haque et al. [20] described to find out the monkeypox disease from the images with deep-transfer learning and attention mechanisms. They used online images and obtained a validation accuracy of 83.89%.

Sitaula et al. [21] proposed a method to find out the monkeypox virus detection with seven deep learning model as well as their ensemble method. They used publicly available data and applied seven pre-trained models initially. Later, to improve the performance they used an ensemble method of deep learning. The highest accuracy of their proposed work is 87.13%.

Ali et al. [8] proposed a method to detect human monkeypox from online collected image data. Their study was deep learning-based. Using online data, their classification rate is 82.96%.

Rahman et al. [22] explained federated and deep learning-based monkeypox disease detection from private limited data. Their study summary is-Accurate identification of monkeypox is difficult because it was deemed a global public health emergency after the COVID-19 epidemic. For efficient monkeypox classification, this paper suggests a safe, federated learning system that makes use of deep learning models such as MobileNetV2, Vision Transformer, and ResNet-50. With 97.90% accuracy using the ViT-B32 model, the method improves data security while guaranteeing accurate disease classification.

Azar et al. [23] proposed a deep neural network-based system to detect monkeypox from the images. Their study summary-this study created a deep learning model based on DenseNet201 to identify skin scans as either normal, chickenpox, monkeypox, or measles in response to the 2022 outbreak. The model performed exceptionally well, attaining 95.18% accuracy in a four-class scenario and 97.63% accuracy in a two-class scenario. To enhance model interpretability and help clinicians trust and comprehend the decision-making process, LIME and Grad-CAM were used. This model performs better than previous research, particularly in F1-Score, and provides information on the afflicted skin areas that are essential for diagnosis.

Altun et al. [7] proposed a method to detect monkeypox from sensor-based data with deep-transfer learning. Their work summary is that-in order to target possible pandemic scenarios, this study sought to create a deep learning-based monkeypox detection algorithm that is both quick and accurate. VGG19, DenseNet121, ResNet-50, EfficientNetV2, MobileNetV3, and Xception models were used to create a new CNN model with hyperparameter tuning and transfer learning. With an F1-score of 0.98, AUC of 0.99, accuracy of 0.96, and recall of 0.97, the optimized MobileNetV3 model exhibited the greatest performance, proving the usefulness of deep learning in quick disease classification.

Ahsan et al. [24] demonstrate a deep learning-based monkeypox disease detection from input data. This study evaluated six deep learning models, including Inception ResNetV2 and Mobile NetV 2, for early illness detection utilizing transfer learning in light of widespread worries about monkeypox as a possible pandemic danger. The altered models demonstrated

their diagnostic capabilities with accuracies ranging from 93% to 99%. By identifying important characteristics linked to the development of monkeypox, LIME was used to improve model transparency.

Uysal et al. [25] explained a hybrid deep learning method to identify monkeypox disease from image data. In order to identify monkeypox from skin photos in a multi-class dataset (monkeypox, chickenpox, measles, and normal), this study created a hybrid AI model by merging CSPDarkNet, InceptionV4, MnasNet, MobileNetV3, RepVGG, SE-ResNet, Xception, and LSTM. Following data augmentation, the hybrid model demonstrated strong performance in differentiating monkeypox from related diseases, with 87% test accuracy and a Cohen's kappa value of 0.8222.

Saleh et al. [26] proposed a new approach to detect monkeypox from image data. They used AI, Chimp algorithm etc. The two-phase AI-based Human Monkeypox Detection (HMD) strategy is presented in this article as a means of early monkeypox detection. Weighted Naïve Bayes, Weighted K-Nearest Neighbors, and a deep learning model through weighted voting are all combined in the Detection Phase (DP) to create an Ensemble Diagnosis (ED) model. The first phase, the Selection Phase (SP), uses an Improved Binary Chimp Optimization (IBCO) algorithm for optimal feature selection. With an accuracy of 98.48%, precision of 91.1%, and recall of 88.91%, HMD outperforms contemporary diagnostic techniques.

Almufareh et al. [27] explained how to detect monkeypox from two different datasets with transfer learning. As a safer substitute for conventional PCR testing, this work suggests a non-invasive, computer-vision-based approach for detecting monkeypox by employing deep learning to analyze skin lesion photos. The method's high sensitivity, specificity, and balanced accuracy, as established by tests on the MSLD and MSID datasets, make it an attractive option for general usage, particularly in places with inadequate lab infrastructure. IoT and AI are used in this method to provide safe, contactless diagnostics.

Table I is the summary of the mentioned work that was published in 2022, 2023, and 2024. After analyzing, we see that the previous work has limitations in some cases, such as that they almost used deep learning, transfer learning, and a hybrid model to classify the monkeypox data without explainability. But, we proposed a new method named Vision Transformer and a hybrid model with explainable AI with the best accuracy [see Table I]. So, our work is superior to theirs because our proposed work has the best accuracy from them. Particularly, we reached 100% accuracy using the vision transformer model without any overfitting and 99.33% using deep hybrid learning. So, we can say that our proposed work is the best work till now.

III. Methods

The overall workflow diagram of monkeypox disease detection and classification is illustrated in Fig. 1. In this part, we will discuss data collection, preprocessing, image augmentation, image separation, using of different types of deep and transfer learning models with details and proposed framework, and finally, we will explain explainable AI results applied to input image and predicted image. From, data collection to the result performance of each model, the sequential explanation is included.

A. Data Collection

We collected the image dataset from the online website such as Kaggle and we customized the data for later use https://www.kaggle.com/datasets/mdmokshedurrahman/monkeypox-image-dataset. This dataset has a total of six classes with one directory. We separated the data for training and testing. The total amount of image data is 7,532 and the classes are "Chickenpox", "HFMD", "Measles", "Healthy", "Cowpox", and "Monkeypox". The images are in different color modes. The sample dataset is shown in Fig. 2.

B. Image Preprocessing

Preprocessing images is an essential step in getting data ready for visual transformer models and transfer learning, particularly in applications like the diagnosis of monkeypox disease [28]. To start, pictures are gathered and their sizes are standardized to guarantee consistency, usually shrinking them to 224x224 pixels [29]. Convergence during model training is accelerated by normalization, which involves scaling pixel values to a range between 0 and 1 or standardizing them to have a mean of 0 and a standard deviation of 1.

C. Image Augmentation

In machine learning applications [30], [31] such as transfer learning, ensemble learning [32], and vision transformers, image augmentation is crucial for enhancing model performance by producing a variety of data variants. Rotation, flipping, scaling, cropping, and other techniques expand the amount of the dataset, which improves model generalization and lowers overfitting, particularly in small or unbalanced datasets [33]. Augmentation enables pre-trained models to successfully adjust to new datasets in transfer learning. By encouraging each model to concentrate on unique features, applying different augmentations across models for ensemble learning lowers prediction variance. Augments like patch shuffling and random cropping within vision transformers (ViTs) improve the model's resistance to visual fluctuations by enhancing its capacity to capture global patterns. Augmentation improves generalization overall, allowing models to better handle changes in real-world data. In our dataset, we applied the some rules for image augmentation [34]. We set the parameter as follows:

```
train-datagen= image.ImageDataGenerator(
rescale=1./255,
shear-range=0.2,
zoom-range=0.2,
horizontal-flip=True)
test-dataset=image.ImageDataGenerator(rescale=1./255)
```

D. Image Partitioning

After finishing the preprocessing and augmentation method to the image data, we separate the data for training, validation, and testing[35], [36]. We partitioned the data as follows:
total image for training: 5,273
total image for validation: 2,259 and
total image for validation: 2,259
That is, we separated the total images into three categories.

TABLE I. SUMMARY OF THE RELATED WORK

Reference	Dataset	Used Methods	Accuracy	XAI
[18]	MSID Dataset	Deep-CNN	98.91%	Yes
[19]	Monkeypox detection dataset	Deep Learning, YOLOV5	98.18%	No
[20]	Online image data	Deep learning and Attention Mechanism	83.89%	No
[21]	Online dataset	Deep learning ensembles	87.13%	Yes
[8]	Collected from online portal	Deep learning	82.96%	No
[22]	Their own data	Deep learning and federated learning	97.90%	No
[23]	Kaggle data	Deep neural network	97.63%	Yes
[7]	Real-time data	Deep-transfer learning	99%	No
[24]	Puclic data	Deep learning	99%	Yes
[25]	Puclicly availabe data	Hybrid Deep Learning	87%	No
[26]	Public dataset	AI, Chimp Algorithm and DL	98.48%	No
[27]	MSLD, MSID Online data	Transfer learning	94%	No
Proposed Approach	Online recent data	Vision Transformer, Hybrid Model	100% for ViTs, 99.33% for Hybrid	Yes

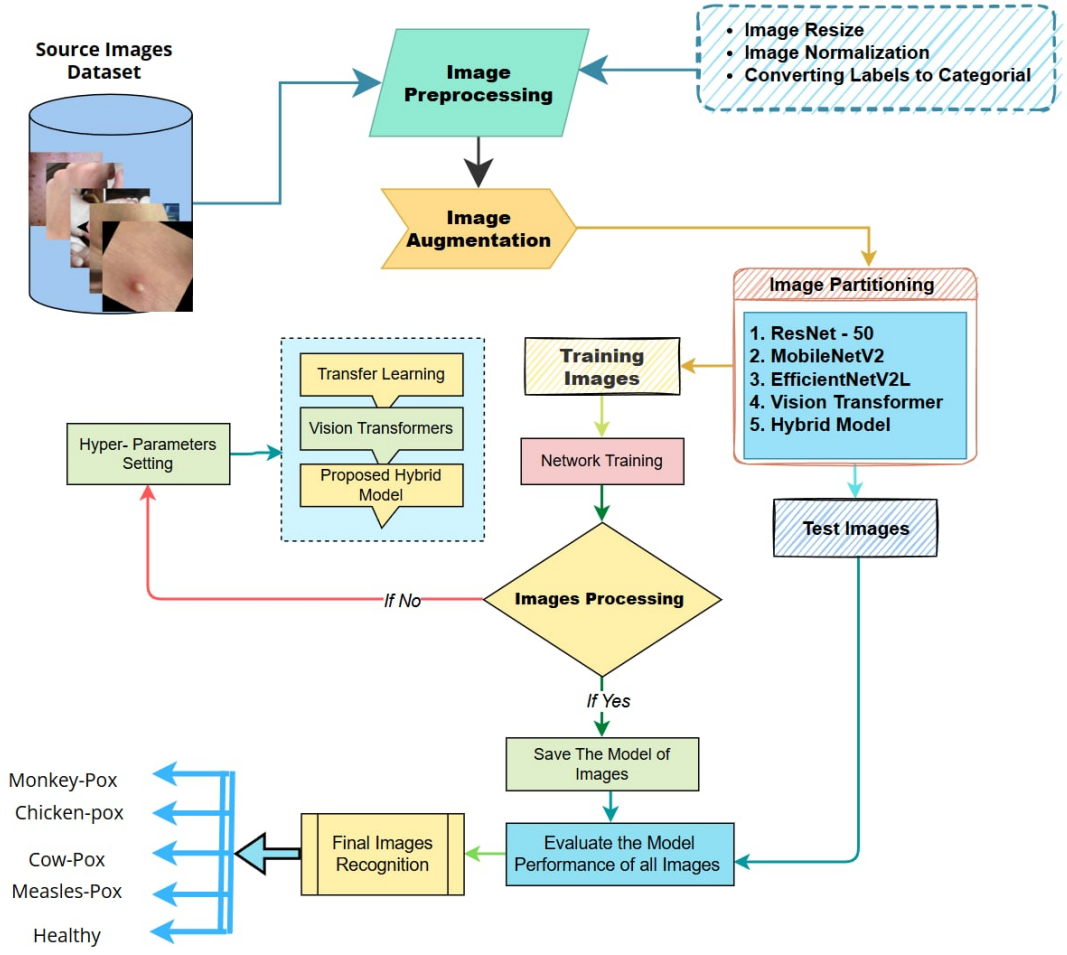

Fig. 1. System architecture.

E. Proposed Neural Network Framework

For image classification, deep learning models are suitable. These models can detect the disease more accurately. In this study, we applied some deep learning models such as EfficientNetV2 and MobileNetV2 [37]. A separate explanation is below.

1) EfficientNetV2: Monkeypox may be successfully detected by picture analysis using EfficientNetV2, a cutting-edge deep-learning model created for image classification applications [38]. By employing a compound scaling technique that consistently increases network depth, width, and resolution, EfficientNetV2's fundamental concept is its capacity to strike a compromise between accuracy and processing efficiency. The model is perfect for processing

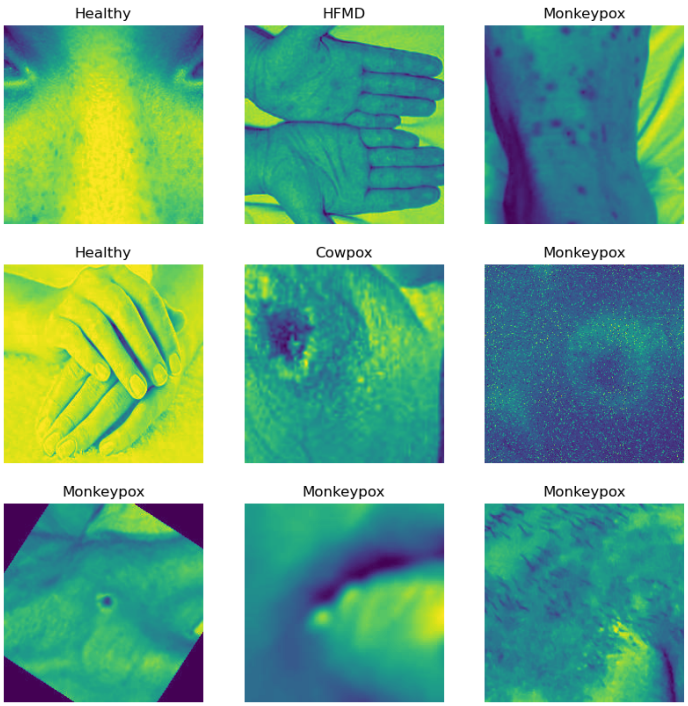

Fig. 2. Dataset sample.

medical images where accuracy is crucial because of its architecture, which allows it to extract complex information from images at a reduced computational cost. EfficientNetV2 may be optimized on a dataset of photos of monkeypox by utilizing transfer learning, which enables it to pick up unique patterns and characteristics that are suggestive of the illness. Its capacity to generalize from sparse data is further improved by its sophisticated training methods, such as progressive training, which begins with images of lower quality and progressively raises the resolution during training. In order to facilitate early diagnosis and prompt action in clinical settings, EfficientNetV2 provides a potent solution for precisely identifying monkeypox lesions in pictures. In this study, we used monkeypox image data to detect and classify the disease from those images. Total images were separated into six classes. In this model, we set the parameters as follows:
batch-size = 32
img-height = 224
img-width = 224
lr-rate = 1e-3
lr-mode = 'cos'
epochs = 30
For the training, we set the parameter as follows"
validation-split=0.2,
subset="training",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size
For the validation, we have,
data0dir,
validation0split=0.2,
subset="validation",
seed=123,
image0size=(img0height, img0width),
batch-size=batch-size

2) MobileNetV2: The lightweight deep learning model MobileNetV2 was created especially for mobile and edge devices, which makes it ideal for real-time applications like identifying monkeypox in medical photos. MobileNetV2's fundamental concept is based on depthwise separable convolutions, which drastically cut down on the number of parameters and calculations needed to analyze data quickly without sacrificing accuracy. Through a sequence of linear bottleneck layers that promote effective information flow and the retention of significant visual details, this architecture improves the model's capacity to learn key elements from images. MobileNetV2 may be optimized on a specific dataset of monkeypox photos by using transfer learning techniques, which will allow it to recognize the distinct patterns and traits linked to the illness. Because of its small size, the model may be used on gadgets with minimal processing power, like smartphones or portable medical imaging equipment, guaranteeing that medical practitioners can make good use of it in a variety of contexts. Consequently, MobileNetV2 offers a quick and easy way to identify monkeypox, which helps with prompt diagnosis and efficient treatment of the illness. In this study, we set the parameter as follows:
batch-size = 32
img-height = 224
img-width = 224
lr-rate = 1e-3
lr-mode = 'cos'
epochs = 15

For the training data, we used,
train-split=0.2,
subset="training",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size

For the validation of data, we difine,
validation-split=0.2,
subset="validation",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size

3) ResNet-50: An excellent option for identifying monkeypox in a six-class image dataset is ResNet-50, a potent transfer learning architecture that performs exceptionally well in image classification tasks. ResNet-50's fundamental concept is its creative use of residual connections, which mitigate the vanishing gradient issue that sometimes arises in very deep networks while enabling the model to learn intricate features [39]. These residual connections make it easier to train deeper networks by allowing gradients to have direct paths during backpropagation, which enhances the model's capacity to recognize complex patterns in images. ResNet-50 can be refined in the context of monkeypox detection using a broad dataset that comprises several classes associated with the disease, such as distinct lesion phases or other skin disorders. This flexibility improves classification accuracy

by enabling the algorithm to pick up subtle visual cues that distinguish monkeypox from related illnesses. ResNet-50 is also well-suited for managing sparse or unbalanced datasets, which are typical in medical imaging, due to its resilience to overfitting. In the end, ResNet-50 offers a dependable method for precisely identifying monkeypox lesions by utilizing its depth and architectural improvements, which helps with prompt diagnosis and efficient patient treatment in clinical settings [40]. In this study, we used the below parameter for ResNet-50 model training and also validation.
batch-size = 32
img-height = 224
img-width = 224
lr-rate = 1e-3
lr-mode = 'cos'
epochs = 30
For the training, we set the parameter as follows:
validation-split=0.2,
subset="training",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size
For the validation, we used the parameter list as follows:
validation-split=0.2,
subset="validation",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size

The basic architecture of the ResNet 50 model for this study is illustrated in Fig. 3.

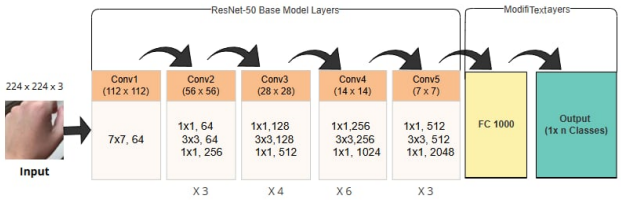

Fig. 3. Proposed ResNet 50 model architecture.

4) Vision Transformer: By adapting transformer-based designs, which were initially created for natural language processing, to visual data, the Vision Transformer (ViT) model(google/vit-base-patch16-224) offers a fresh method for image analysis [41]. ViT interprets images as a series of tiny patches, collecting global dependencies throughout the image, in contrast to convolutional neural networks (CNNs), which rely on local feature extraction [42]. It is especially useful for differentiating intricate visual patterns linked to illnesses like monkeypox because of its capacity to comprehend spatial relationships. ViT can learn the distinct visual indicators of monkeypox lesions, such as shape, texture, and distribution, across different phases and classes by training on a collection of monkeypox images. This method is useful for detecting monkeypox because it enables the model to understand both little details and more general contextual patterns, which helps it distinguish monkeypox from other skin disorders that are similar. Furthermore, ViT can concentrate on pertinent image regions thanks to its attention mechanism, which could improve interpretability in medical diagnostics. All things considered, Vision Transformer offers a potential tool for detecting monkeypox by fusing high accuracy with knowledge of the spatial patterns that characterize the illness [43]. In our study, we follow the working mechanism of the Vision Transformer model shown in Fig. 4.

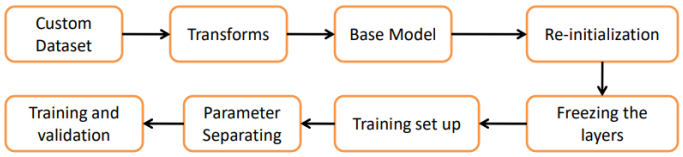

Fig. 4. Working procedure of the vision transformer model.

Initially, we customized the dataset for the transformer. The parameter set as:
transform = transforms.Compose([
transforms.Resize((224, 224)),
transforms.ToTensor(),
transforms.Normalize(mean=[0.5, 0.5, 0.5],
std=[0.5, 0.5, 0.5]),])

After transforming, we defined the base model and re-initialized the features. Freezing the layers, we did training setup and parameter dividing for the training and validation.

5) Hybrid Model: To obtain better performance and more accuracy, we used a hybrid version of three models. The hybrid model is generated from averaging the output from ResNetV2, MobilNetV2, and EfficientNetV2 [44]. This hybrid model technique detects and classifies monkeypox across six different image classes by averaging the outputs of three sophisticated deep-learning architectures: ResNetV2, MobileNetV2, and EfficientNetV2. Every one of these types has special advantages: While MobileNetV2 offers lightweight efficiency, making it highly responsive and appropriate for real-time processing with limited CPU resources, ResNetV2's residual connections enable the capture of intricate image features and deep hierarchical patterns. In contrast, EfficientNetV2 offers a balanced scaling technique that simultaneously modifies network depth, width, and resolution to improve accuracy and efficiency. The hybrid model leverages the combined advantages of each architecture by averaging the predictions from these three networks. This integration lessens the biases present in any one model, producing a more robust and balanced outcome that is particularly helpful for managing the many visual traits of monkeypox lesions in various classes [45]. Even with the variances found in a medical picture dataset, the hybrid model can function effectively thanks to the averaging technique, which may enhance generalization and lower mistakes. In the end, this team approach improves the accuracy and dependability of monkeypox detection, offering a complete tool to assist medical professionals in diagnosing and categorizing the illness. In this combined model, we set the features for the training data and also the same for validation as follows:
validation-split=0.2,
subset="training",

seed=123,
image-size=(224, 224),
batch-size=32)
For the base model, the parameters are:
weights='imagenet',
include-top=False,
input-shape=(224, 224, 3)

The basic organization of the proposed hybrid model is illustrated in Fig. 5.

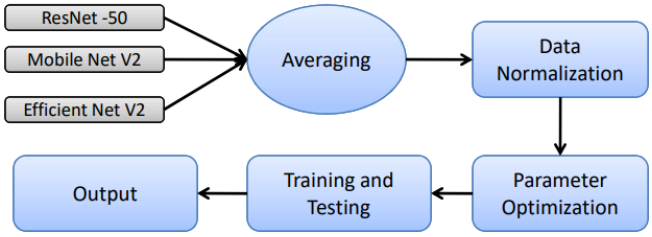

Fig. 5. Working procedure of proposed hybrid model.

In Fig. 5, for the section of Normalization, we set the parameter as:
Rescaling=1./255,
map(lambda x, y: (normalization-layer(x), y)).cache().prefetch(buffer-size=tf.data.AUTOTUNE)
We optimized the some parameters such as:
Input(shape=(224, 224, 3))
optimizers.Adam(learning-rate=1e-3),
losses.SparseCategoricalCrossentropy(from-logits=True),
metrics=['accuracy']
epochs=30

F. Explainable AI Approach

We use Grad-CAM (Gradient-weighted Class Activation Mapping) to show the relevant areas in monkeypox images as part of our Explainable AI feature extraction method for monkeypox detection [46]. The interpretability technique Grad-CAM improves the transparency of deep learning models by assisting in determining which aspects of an image have the greatest influence on a model's prediction. This method highlights the main characteristics in the images used for monkeypox categorization across several classes by implementing Grad-CAM for a ResNet-based model [21]. First, a Grad-CAM class is created, which initializes a gradient model by choosing the network's last convolutional layer and attaching it to the output layer of the model. Grad-CAM computes the gradient of the output class (predicted as monkeypox or another) in relation to the feature mappings in the final convolutional layer after an input picture has been run through the model. The features that are important for the model's categorization are shown by these gradients [47]. A heatmap highlighting the significant regions of the image that influenced the model's prediction is produced after pooling the gradients and appropriately weighting the feature maps. In order to create a superimposed visualization, the calculated heatmap is enlarged to the original image proportions, colored using a "jet" colormap, and then superimposed on the original image. Medical practitioners may more easily validate the model's focus areas and comprehend predictions thanks to this overlay, which shows the portions of the image that the algorithm looks for in order to detect monkeypox. By making the model's decision-making process more clear, these explainable strategies increase confidence in the model's application for medical imaging diagnosis and categorization of monkeypox.

IV. Results

In this section, we discussed the result of the deep learning model in Monkeyfox disease detection [16]. Particularly, we explored the results of EfficientNetV2 and MobileNetV2.

From the above part, we know that EfficientNetV2 is used in the monkeypox image dataset and has a total of six classes. The basic parameter details of EfficientNetV2 is shown in Table II.

TABLE II. EfficientNetV2 Parameter Details

Layer (type)	Output Shape	Param
input-layer-18 (Input-Layer)	(None, 224, 224, 3)	0
efficientnetv2-l (Functional)	(None, 7, 7, 1280)	117,746,848
conv2d-7 (Conv2D)	(None, 7, 7, 512)	5,898,752
global-average-pooling2d-7	(None, 512)	0
dense-14 (Dense)	(None, 256)	131,328
dense-15 (Dense)	(None, 6)	1,542
Total params:	123,778,470	0
Trainable params:	6,031,622	0
Non-trainable params:	117,746,848	

In this model,
Training accuracy is 99.94%
Training loss is 0.38%.
The validation accuracy is 97.80% and
The validation loss is 6.72%.
The testing accuracy is: 97.67%
The testing loss is: 6.94%.
The training accuracy and validation accuracy are shown in Fig. 6 and the training loss and validation loss are shown in Fig. 7.

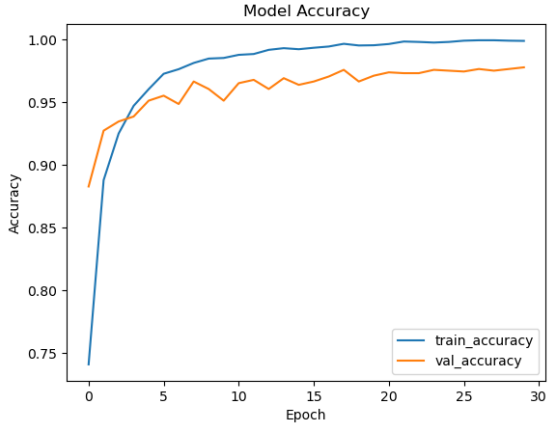

Fig. 6. Train accuracy vs. Validation accuracy of EfficientNetV2.

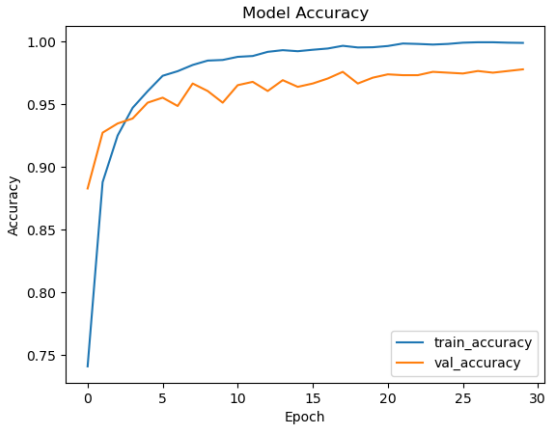

Fig. 7. Train loss vs. Validation loss of EfficientNetV2.

The classification report of the EfficientNetV2 is illustrated in Table III.

TABLE III. CLASSIFICATION REPORT OF EFFICIENTNETV2 MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.95	0.96	0.95	140
Chickenpox	0.95	0.96	0.95	140
Cowpox	1.00	0.97	0.98	120
HFMD	0.99	0.98	0.98	345
Healthy	0.99	0.99	0.99	251
Measles	0.99	0.94	0.96	98
Monkeypox	0.97	0.99	0.98	549
accuracy	0	0	0.98	1503
macro avg	0.98	0.97	0.97	1503
weighted avg	0.98	0.98	0.98	1503

We used MobileNetV2 in image data for detecting monkeypox. After applying the model, we have the following parameter list shown in Table IV.

TABLE IV. MOBILENETV2 PARAMETER DETAILS

Layer (type)	Output Shape	Param
input-layer-14 (Input Layer)	(None, 224, 224, 3)	0
mobilenetv2-1.00-224	(None, 7, 7, 1280)	2,257,984
conv2d-5 (Conv2D)	(None, 7, 7, 512)	5,898,752
global-average-pooling2d-5	(None, 512)	0
dense-10 (Dense)	(None, 256)	131,328
dense-11 (Dense)	(None, 6)	1,542

In this model, we got,
Training accuracy is 95.47%
Training loss is 16.2%.
The validation accuracy is 79.51% and
The validation loss is 64.62%.
The testing accuracy is: 78.18%
and the testing loss is: 20%.
The classification report of MobileNetV2 is shown in Table V.

We used one Transfer Learning model named ResNet-50 to detect the monkeypox from the image data. After applying this model, we have the model parameter summary shown in Table VI.

TABLE V. CLASSIFICATION REPORT OF MOBILENETV2 MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.95	0.96	0.95	140
Chickenpox	1.00	0.97	0.98	140
Cowpox	1.00	0.97	0.98	120
HFMD	0.99	0.98	0.98	345
Healthy	0.99	0.99	0.99	251
Measles	0.99	0.94	0.96	98
Monkeypox	0.97	0.99	0.98	549
accuracy	0	0	0.98	1503
macro avg	0.98	0.97	0.97	1503
weighted avg	0.98	0.98	0.98	1503

TABLE VI. RESNET-50 PARAMETER DETAILS

Layer (type)	Output Shape	Param
input-layer-12 (Input Layer)	(None, 224, 224, 3)	0
ResNet-50 (Functional)	(None, 7, 7, 2048)	23,587,712
conv2d-4 (Conv2D)	(None, 7, 7, 512)	9,437,696
global-average-pooling2d-4	(None, 512)	0
dense-8 (Dense)	(None, 256)	131,328
dense-9 (Dense)	(None, 6)	1,542

Training accuracy is 99.87%%
Training loss is 0.43%.
The validation accuracy is 99.85% and
The validation loss is 0.39%.
The testing accuracy is: 97.18%
and the testing loss is: 4%.
The classification report of this model is shown in Table VII and the ROC Curve of this model is shown in Fig. 8.

TABLE VII. CLASSIFICATION REPORT OF RESNET-50 MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.97	0.86	0.91	140
Cowpox	0.97	0.98	0.98	120
HFMD	0.98	0.97	0.98	345
Healthy	0.97	0.98	0.97	251
Measles	0.96	0.95	0.95	98
Monkeypox	0.96	0.98	0.97	549
accuracy	0	0	0.97	1503
macro avg	0.97	0.95	0.96	1503
weighted avg	0.97	0.97	0.97	1503

The accuracy and loss curve of this model is illustrated in Fig. 9.

Using Vision Transformer (google/vit-base-patch16-224) for detecting monkeypox disease, we have the following results.
The number of epochs: 20
Training accuracy is 100%
Training loss is 0.00%.
The validation accuracy is 87.51% and
The validation loss is 0.37%.

The classification report is shown in Table VIII and the confusion matrices of this model is shown in Fig. 10 where true label vs predicted label and actual label vs. predicted lable is illustrated. The multiclass ROC Curve and Precision-recall curve is explained in Fig. 11 and 12.

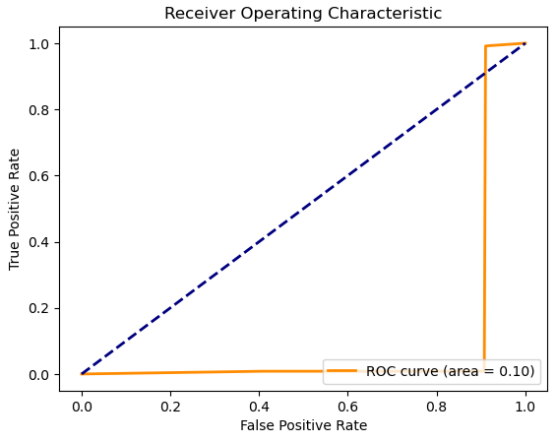

Fig. 8. The ROC curve for ResNet-50 model.

TABLE VIII. CLASSIFICATION REPORT OF VISION TRANSFORMER MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.64	0.86	0.73	110
Cowpox	0.96	0.93	0.94	91
HFMD	0.94	0.97	0.96	229
Healthy	0.95	0.86	0.90	175
Measles	0.96	0.89	0.92	80
Monkeypox	0.88	0.84	0.87	445
accuracy	0	0	0.88	1130
macro avg	0.89	0.89	0.89	1130
weighted avg	0.89	0.89	0.89	1130

We combined three models to improve the accuracy and reduce the error rate as well as loss amount. Hence, ResNet-50, MobileNetV2, and EfficientNetV2 models are averaging into a single unit named as Hybrid model. After applying this technique, we have the following results:
The number of epochs: 30
Training accuracy is 99.33%
Training loss is 2.11%.
The validation accuracy is 90.09% and
The validation loss is 40.62%.

The summary of the parameter is listed in Table IX and the classification report of this model is shown in Table X.

TABLE IX. HYBRID MODEL PARAMETER DETAILS

Layer (type)	Output Shape	Param	Connected to
input-layer-14 (Input Layer)	(None, 224, 224, 3)	0	-
functional-15	(None, 6)	24,113,798	input-layer-32[0...
functional-16	(None, 6)	2,587,462	input-layer-32[0...
functional-17	(None, 6)	118,076,3...	input-layer-32[0...
average-1 (Average)	(None, 6)	0	functional-15[0]... functional-16[0] functional-17[0]...

The accuracy and loss curve is shown in Fig. 13 and the confusion matrix is shown in Fig. 14.

TABLE X. CLASSIFICATION REPORT OF HYBRID MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.94	0.77	0.85	140
Cowpox	0.99	0.87	0.92	120
HFMD	0.98	0.85	0.91	345
Healthy	0.94	0.91	0.93	251
Measles	0.83	0.83	0.83	98
Monkeypox	0.84	0.98	0.91	549
accuracy	0	0	0.90	1503
macro avg	0.92	0.77	0.89	1503
weighted avg	0.91	0.90	0.90	1503

The results summary of the proposed model are shown in Table XI.

TABLE XI. RESULT SUMMARY OF USED METHODS

Model	Training Accuracy	Validation Accuracy	Testing Accuracy
EfficientNetV2	99.94%	97.80%	97.67%
MobileNetV2	95.47%	79.51%	78.18%
ResNet-50	99.87%	99.85%	97.18%
Vision Transformer	100%	87.51%	99.41%
Hybrid Model	99.33%	90.09%	99.09%

A. Exploring Grad-CAM

We applied explainable AT to the predicted image to explain it based on trained images. If we use any monkeypox-positive image for the explanation, then based on the training and predicted value, the machine can explain the image using the heat-map method [48]. Using the Grad-CAM, the system can explain the input image for clearance. One suitable example is shown in Fig. 15. Especially, this image is the monkeypox positive input image, the system will use a heat map to analyze and explain it. Finally, the system is successful, saying that it is the monkeypox positive image [49]. Table XII shows evaluation metrics for Grad-CAM for hybrid model.

B. Comparison with Previous Studies

The comparison presented in Table XIII highlights the effectiveness of the proposed models in achieving state-of-the-art performance for monkeypox diagnosis. Our study demonstrated superior accuracy with Vision Transformers (ViTs) achieving 99.41% and the hybrid model achieving 99.09%. These results outperform many of the existing studies, such as [18] (98.91%) and [7] (99%), showcasing the robustness of our approach.

The incorporation of Vision Transformers proved particularly impactful due to their ability to capture global dependencies within the input data, which is critical for nuanced image classification tasks. The hybrid model further enhanced performance by combining the strengths of ResNet-50, MobileNetV2, and EfficientNetV2, enabling better feature extraction and classification accuracy.

Compared to prior studies, such as [20] and [8], which reported lower accuracies of 83.89% and 82.96%, respectively, our models exhibited a significant improvement. Additionally, while models like [21] and [23] incorporated Explainable AI (XAI) techniques, their accuracies (87.13% and 97.63%) were

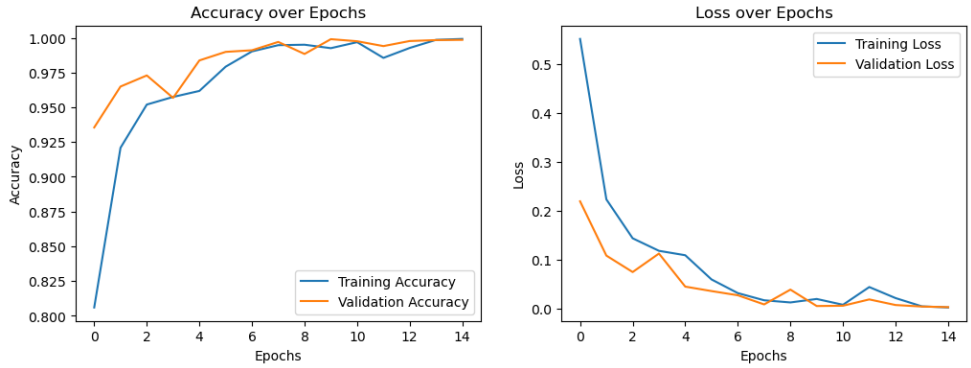

Fig. 9. Accuracy and loss curve of ResNet-50 model.

TABLE XII. EVALUATION METRICS FOR GRAD-CAM FOR HYBRID MODEL

Evaluation Metric	Grad-CAM	Explanation
Ground Truth Mask Overlap	97%	Percentage of overlap between the ground truth mask and the highlighted region.
Feature Coverage	0.97	Proportion of the image covered by the highlighted features in the explanation.
Relevant Activation	96%	Percentage of activation in relevant areas.
Feature Relevance	0.96	Relevance of features in the explanation, corresponding to the model's decision.
Similarity (Mean Absolute Error)	0.89	Mean absolute error between the predicted and actual class.
Consistency Error	0.89	Error in consistency when input is perturbed or modified.

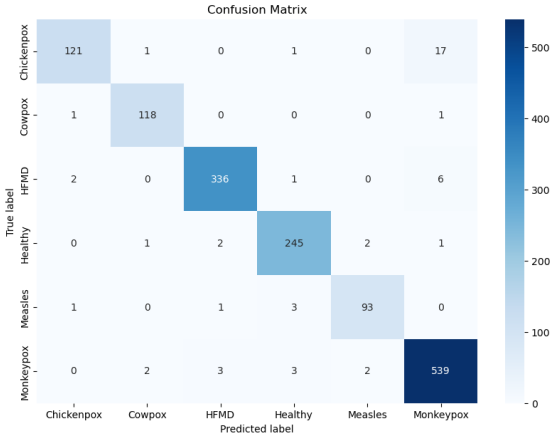

Fig. 10. Confusion matrix of vision transformer in true label vs. Predicted label.

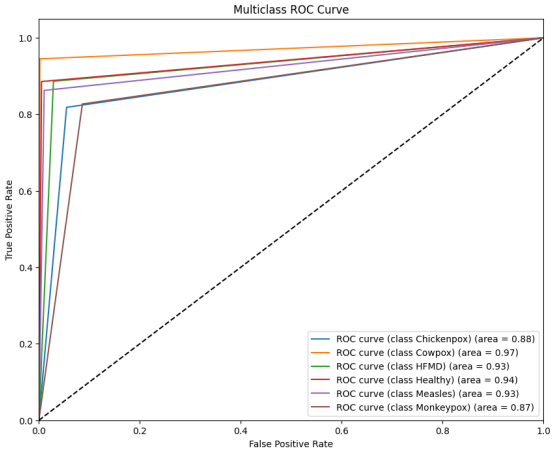

Fig. 11. ROC Curve for vision transformer model.

lower than ours, demonstrating that our integration of Grad-CAM not only enhanced interpretability but also maintained high performance.

Explainability remains a critical aspect of monkeypox diagnosis, as accurate predictions alone are insufficient in sensitive medical applications. Our use of Grad-CAM enabled a detailed understanding of model decisions, providing visual insights into key features contributing to the classification. This is a step forward in building trust and transparency in AI-based healthcare solutions, addressing concerns in studies like [24] and [26], which either did not integrate XAI or lacked detailed visualization.

V. CONCLUSION AND FUTURE RESEARCH

This study demonstrates the potential of deep learning models, particularly Vision Transformer and hybrid approaches, in achieving accurate and interpretable monkeypox diagnosis. Among the models tested, the Vision Transformer emerged as the most effective, achieving high accuracy across training, validation, and testing phases while maintaining robustness against overfitting. The hybrid model, combining ResNet-50, MobileNetV2, and EfficientNetV2, also delivered competitive performance, highlighting the benefits of leveraging diverse architectural strengths. The integration of Grad-CAM enhanced the interpretability of the models, providing valuable insights into their decision-making processes, a critical requirement for clinical applications. These findings highlight the role of AI-driven solutions in enabling early and

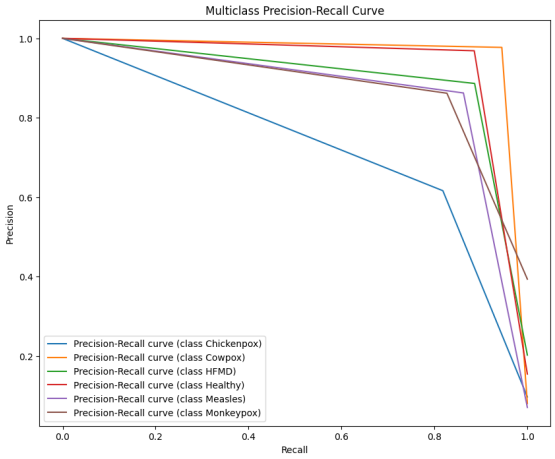

Fig. 12. Precision-recall curve for vision transformer model.

TABLE XIII. COMPARISON WITH PREVIOUS STUDIES

Reference	Accuracy	XAI
[18]	98.91%	Yes
[19]	98.18%	No
[20]	83.89%	No
[21]	87.13%	Yes
[8]	82.96%	No
[22]	97.90%	No
[23]	97.63%	Yes
[7]	99%	No
[24]	99%	Yes
[25]	87%	No
[26]	98.48%	No
[27]	94%	No
Our Study	99.41% for ViTs, 99.09% for Hybrid	Yes

precise monkeypox diagnosis, thereby aiding timely containment and treatment. One of the limitations of this study is that the dataset was taken from an online publication from the clinical sector, and to detect the proper place of monkeypox, the segmentation method can be applied in real-time image data. Future research will cover this technique. Future research should focus on enhancing the generalizability of the proposed models by expanding the dataset to include diverse populations, imaging conditions, and clinical real-time data. Additionally, improving model efficiency for deployment in resource-constrained environments will be crucial for enabling widespread adoption. Incorporating other explainability techniques, such as LIME or SHAP, could provide deeper insights into model predictions, fostering greater trust among clinicians. Exploring federated learning frameworks may further enhance privacy and scalability, allowing collaborative training across institutions without compromising data security. Longitudinal studies spanning various demographic and clinical contexts will help validate model reliability over time. Moreover, integrating multi-modal data, such as clinical biomarkers and patient metadata, could improve diagnostic accuracy and provide a more holistic understanding of monkeypox.

DATA AVAILABILITY

The used datasets are open-access and referenced in this manuscript.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] S. Das, T. Mahmud, D. Islam, M. Begum, A. Barua, M. Tarek Aziz, E. Nur Showan, L. Dey, and E. Chakma, "Deep transfer learning-based foot no-ball detection in live cricket match," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, p. 2398121, 2023.

[2] S. Umme Habiba, F. Tasnim, M. S. Hasan Chowdhury, M. K. Islam, L. Nahar, T. Mahmud, M. S. Kaiser, M. S. Hossain, and K. Andersson, "Early prediction of chronic kidney disease using machine learning algorithms with feature selection techniques," in *International Conference on Applied Intelligence and Informatics*. Springer, 2023, pp. 224–242.

[3] S. U. Habiba, T. Mahmud, S. R. Naher, M. T. Aziz, T. Rahman, N. Datta, M. S. Hossain, K. Andersson, and M. Shamim Kaiser, "Deep learning solutions for detecting bangla fake news: A cnn-based approach," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 107–118.

[4] M. T. Aziz, T. Mahmud, N. Datta, M. Maskat Sharif, N. U. A. Khan, S. Yasmin, M. D. N. Uddin, M. S. Hossain, and K. Andersson, "A state-of-the-art review of machine learning in cybersecurity data science," in *Innovations in Cybersecurity and Data Science*. Singapore: Springer Nature Singapore, 2024, pp. 791–806.

[5] S. R. Naher, S. Sultana, T. Mahmud, M. T. Aziz, M. S. Hossain, and K. Andersson, "Exploring deep learning for chittagonian slang detection in social media texts," in *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET*. IEEE, 2024, pp. 1–6.

[6] T. Mahmud, K. Barua, K. Chakma, R. Chakma, N. Sharmen, M. S. Kaiser, M. S. Hossain, M. S. Hossain, and K. Andersson, "Exploring the effectiveness of region-based cnns in skin cancer diagnosis," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 371–389.

[7] M. Altun, H. Gürüler, O. Özkaraca, F. Khan, J. Khan, and Y. Lee, "Monkeypox detection using cnn with transfer learning," *Sensors*, vol. 23, no. 4, p. 1783, 2023.

[8] S. N. Ali, M. T. Ahmed, J. Paul, T. Jahan, S. Sani, N. Noor, and T. Hasan, "Monkeypox skin lesion detection using deep learning models: A feasibility study," *arXiv preprint arXiv:2207.03342*, 2022.

[9] T. Mahmud, T. Akter, M. T. Aziz, M. K. Uddin, M. S. Hossain, and K. Andersson, "Integration of nlp and deep learning for automated fake news detection," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2024, pp. 398–404.

[10] N. A. Chowdhury, T. Mahmud, A. Barua, N. Basnin, K. Barua, A. Iqbal, M. S. Hossain, K. Andersson, M. S. Kaiser, M. S. Hossain *et al.*, "A novel approach to detect stroke from 2d images using deep learning," in *International Conference on Big Data, IoT and Machine Learning*. Springer, 2023, pp. 239–253.

[11] M. T. Aziz, J. Sikder, T. Rahman, A. D. Del Mundo, S. F. Faisal, and N. U. A. Khan, "Covid-19 detection from chest x-ray images using deep learning," *The Seybold Report*, vol. 17, pp. 706–718, 2022.

[12] M. H. Ali, T. Mahmud, M. T. Aziz, M. F. B. A. Aziz, M. S. Hossain, and K. Andersson, "Leveraging transfer learning for efficient classification of coffee leaf diseases," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024, pp. 1–6.

[13] T. Mahmud, N. Datta, R. Chakma, U. K. Das, M. T. Aziz, M. Islam, A. H. M. Salimullah, M. S. Hossain, and K. Andersson, "An approach for crop prediction in agriculture: Integrating genetic algorithms and machine learning," *IEEE Access*, 2024.

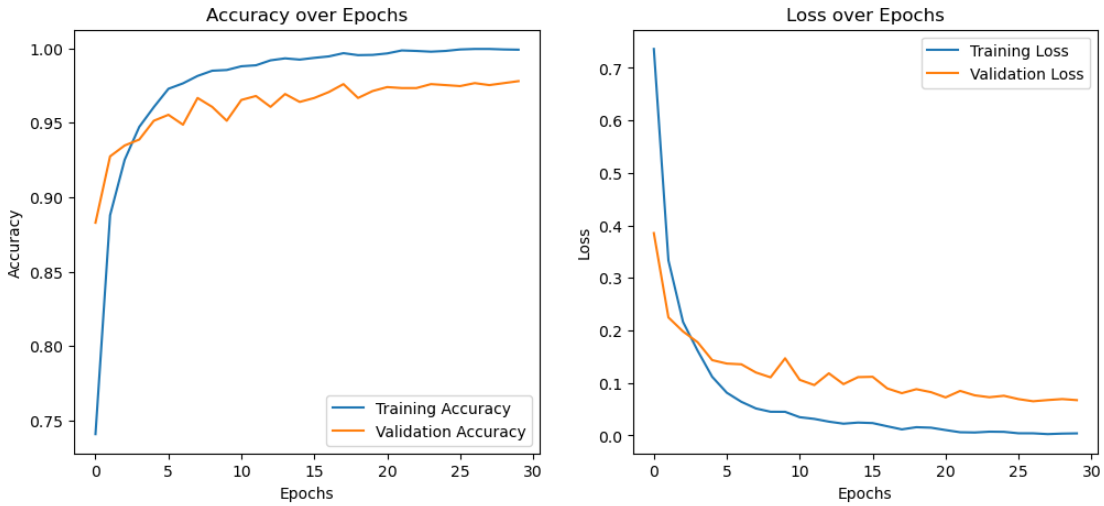

Fig. 13. Accuracy and loss curve for hybrid model.

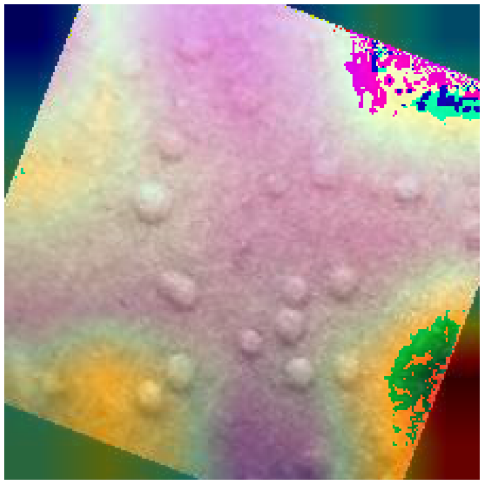

Fig. 15. Explainable AI predicted output image by Grad-CAM for hybrid model.

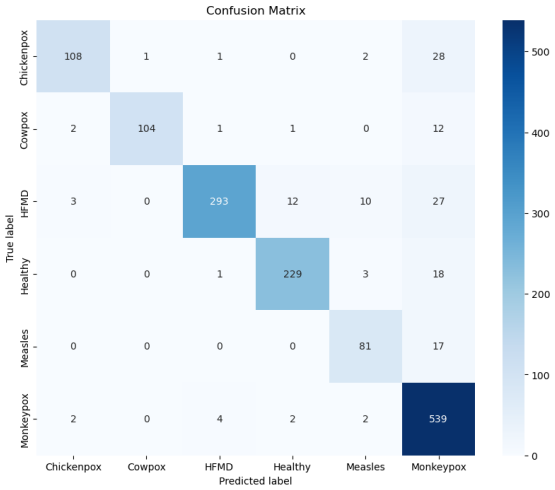

Fig. 14. The confusion matrix of hybrid model.

[14] T. Mahmud, T. Akter, S. Anwar, M. T. Aziz, M. S. Hossain, and K. Andersson, "Predictive modeling in forex trading: A time series analysis approach," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2024, pp. 390–397.

[15] N. Datta, T. Mahmud, M. T. Aziz, R. K. Das, M. S. Hossain, and K. Andersson, "Emerging trends and challenges in cybersecurity data science: A state-of-the-art review," in *2024 Parul International Conference on Engineering and Technology (PICET)*. IEEE, 2024, pp. 1–7.

[16] T. Mahmud, I. Hasan, M. T. Aziz, T. Rahman, M. S. Hossain, and K. Andersson, "Enhanced fake news detection through the fusion of deep learning and repeat vector representations," in *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, 2024, pp. 654–660.

[17] M. T. Aziz, T. Mahmud, N. Datta, M. M. Sharif, N. U. A. Khan, S. Yasmin, M. N. Uddin, M. S. Hossain, and K. Andersson, "A state-of-the-art review of machine learning in cybersecurity data science."

[18] D. Bala, M. S. Hossain, M. A. Hossain, M. I. Abdullah, M. M. Rahman, B. Manavalan, N. Gu, M. S. Islam, and Z. Huang, "Monkeynet: A robust deep convolutional neural network for monkeypox disease detection and classification," *Neural Networks*, vol. 161, pp. 757–775, 2023.

[19] N. Dahiya, Y. K. Sharma, U. Rani, S. Hussain, K. V. Nabilal, A. Mohan, and N. Nuristani, "Hyper-parameter tuned deep learning approach for effective human monkeypox disease detection," *Scientific Reports*, vol. 13, no. 1, p. 15930, 2023.

[20] M. E. Haque, M. R. Ahmed, R. S. Nila, and S. Islam, "Human monkeypox disease detection using deep learning and attention mechanisms," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2022, pp. 1069–1073.

[21] C. Sitaula and T. B. Shahi, "Monkeypox virus detection using pre-trained deep learning-based approaches," *Journal of Medical Systems*, vol. 46, no. 11, p. 78, 2022.

[22] D. Kundu, M. M. Rahman, A. Rahman, D. Das, U. R. Siddiqi, M. G. R. Alam, S. K. Dey, G. Muhammad, and Z. Ali, "Federated deep learning for monkeypox disease detection on gan-augmented dataset," *IEEE Access*, 2024.

[23] A. Sorayaie Azar, A. Naemi, S. Babaei Rikan, J. Bagherzadeh Mohasefi, H. Pirnejad, and U. K. Wiil, "Monkeypox detection using deep neural networks," *BMC Infectious Diseases*, vol. 23, no. 1, p. 438, 2023.

[24] M. M. Ahsan, T. A. Abdullah, M. S. Ali, F. Jahora, M. K. Islam, A. G. Alhashim, and K. D. Gupta, "Transfer learning and local interpretable model agnostic based visual approach in monkeypox disease detection and classification: A deep learning insights," *arXiv preprint arXiv:2211.05633*, 2022.

[25] F. Uysal, "Detection of monkeypox disease from human skin images

with a hybrid deep learning model," *Diagnostics*, vol. 13, no. 10, p. 1772, 2023.

[26] A. I. Saleh and A. H. Rabie, "Human monkeypox diagnose (hmd) strategy based on data mining and artificial intelligence techniques," *Computers in Biology and Medicine*, vol. 152, p. 106383, 2023.

[27] M. F. Almufareh, S. Tehsin, M. Humayun, and S. Kausar, "A transfer learning approach for clinical detection support of monkeypox skin lesions," *Diagnostics*, vol. 13, no. 8, p. 1503, 2023.

[28] Y. Zhang, J. Wang, J. M. Gorriz, and S. Wang, "Deep learning and vision transformer for medical image analysis," p. 147, 2023.

[29] R. Karthik, V. Thalanki, and P. Yadav, "Deep learning-based histopathological analysis for colon cancer diagnosis: A comparative study of cnn and transformer models with image preprocessing techniques," in *International Conference on Intelligent Systems Design and Applications*. Springer, 2023, pp. 90–101.

[30] M. T. Aziz, T. Mahmud, M. K. Uddin, S. N. Hossain, N. Datta, S. Akther, M. S. Hossain, and K. Andersson, "Machine learning-driven job recommendations: Harnessing genetic algorithms," in *International Congress on Information and Communication Technology*. Springer, 2024, pp. 471–480.

[31] T. Mahmud, M. Ptaszynski, and F. Masui, "Leveraging explainable ai and sarcasm features for improved cyberbullying detection in multilingual settings," in *2024 IEEE Digital Platforms and Societal Harms (DPSH)*. IEEE, 2024, pp. 1–8.

[32] S. Barman, M. R. Biswas, S. Marjan, N. Nahar, M. H. Imam, T. Mahmud, M. S. Kaiser, M. S. Hossain, and K. Andersson, "A two-stage stacking ensemble learning for employee attrition prediction," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 119–132.

[33] D. Sehgal and I. Saini, "Gan-based image augmentation and comparative analysis of various cnn models for monkeypox detection," in *2024 First International Conference on Electronics, Communication and Signal Processing (ICECSP)*. IEEE, 2024, pp. 1–7.

[34] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artificial Intelligence Review*, vol. 55, no. 3, pp. 2351–2377, 2022.

[35] P. Dey, T. Mahmud, K. M. Foysol, N. Sharmen, M. S. Hossain, and K. Andersson, "Hybrid deep transfer learning framework for humerus fracture detection and classification from x-ray images," in *2023 4th International Conference on Intelligent Technologies (CONIT)*. IEEE, 2024, pp. 1–6.

[36] P. Dey, T. Mahmud, M. S. Chowdhury, M. S. Hossain, and K. Andersson, "Human age and gender prediction from facial images using deep learning methods," *Procedia Computer Science*, vol. 238, pp. 314–321, 2024.

[37] M. H. Imam, N. Nahar, R. Bhowmik, S. B. S. Omit, T. Mahmud, M. S. Hossain, and K. Andersson, "A transfer learning-based framework: Mobilenet-svm for efficient tomato leaf disease classification," in *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, 2024, pp. 693–698.

[38] S. Vats, J. P. Bhati, A. Singla, V. Kukreja, and R. Sharma, "Advanced image classification on intel datasets using optimized efficientnet and mobilenetv2," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*. IEEE, 2024, pp. 1–4.

[39] M. B. Sahaai, G. Jothilakshmi, D. Ravikumar, R. Prasath, and S. Singh, "Resnet-50 based deep neural network using transfer learning for brain tumor classification," in *AIP Conference Proceedings*, vol. 2463, no. 1. AIP Publishing, 2022.

[40] T. Tian, L. Wang, M. Luo, Y. Sun, and X. Liu, "Resnet-50 based technique for eeg image characterization due to varying environmental stimuli," *Computer Methods and Programs in Biomedicine*, vol. 225, p. 107092, 2022.

[41] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu *et al.*, "A survey on vision transformer," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 87–110, 2022.

[42] M. Aloraini, "An effective human monkeypox classification using vision transformer," *International Journal of Imaging Systems and Technology*, vol. 34, no. 1, p. e22944, 2024.

[43] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, "Pre-trained image processing transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 12 299–12 310.

[44] T. Mahmud, M. T. Aziz, M. K. Uddin, K. Barua, T. Rahman, N. Sharmen, M. Shamim Kaiser, M. Sazzad Hossain, M. S. Hossain, and K. Andersson, "Ensemble learning approaches for alzheimer's disease classification in brain imaging data," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 133–147.

[45] S. Maqsood, R. Damaševičius, S. Shahid, and N. D. Forkert, "Moxnet: Multi-stage deep hybrid feature fusion and selection framework for monkeypox classification," *Expert Systems with Applications*, vol. 255, p. 124584, 2024.

[46] G. M. Idroes, T. R. Noviandy, T. B. Emran, and R. Idroes, "Explainable deep learning approach for mpox skin lesion detection with grad-cam," *Heca Journal of Applied Sciences*, vol. 2, no. 2, pp. 54–63, 2024.

[47] A. Akram, A. A. Jamjoom, N. Innab, N. A. Almujally, M. Umer, S. Alsubai, and G. Fimiani, "Skinmarknet: an automated approach for prediction of monkeypox using image data augmentation with deep ensemble learning models," *Multimedia Tools and Applications*, pp. 1–17, 2024.

[48] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of explainable ai techniques in healthcare," *Sensors*, vol. 23, no. 2, p. 634, 2023.

[49] D. Saraswat, P. Bhattacharya, A. Verma, V. K. Prasad, S. Tanwar, G. Sharma, P. N. Bokoro, and R. Sharma, "Explainable ai for healthcare 5.0: opportunities and challenges," *IEEE Access*, vol. 10, pp. 84 486–84 517, 2022.