Identification of Monkeypox Disease Based on MpoxNet and Swin Transformer Models Using Mobile Application

S. Sadesh[1], Rajasekaran Thangaraj[2]*, P. Pandiyan[3], R. Devi Priya[4], Palanichamy Naveen[5]*

[1] Department of Artificial Intelligence and Data Science, Velalar College of Engineering and Technology, Erode (India)
[2] Department of Computer Science and Engineering, Nandha Engineering College, Erode (India)
[3] Department of Electrical and Electronic Engineering, KPR Institute of Engineering and Technology, Coimbatore (India)
[4] Department of Computer Science and Engineering, Centre for IoT and AI (CITI), KPR Institute of Engineering and Technology, Coimbatore (India)
[5] Department of Electrical and Electronic Engineering, Centre for IoT and AI (CITI), KPR Institute of Engineering and Technology, Coimbatore (India)

* Corresponding author: naveenamp88@gmail.com (P. Naveen), rajasekaran30@gmail.com (R. Thangaraj)

Received 12 October 2023 | Accepted 13 August 2024 | Early Access 15 November 2024

Abstract

Humankind is still reeling from the devastating impact of the Covid-19 pandemic, yet another looming threat is the potential global spread of the monkeypox virus. While monkeypox may not pose the same level of lethality or contagion as COVID-19, its significant spread across countries is cause for concern. Already, outbreaks have been reported in 75 nations worldwide. Despite sharing clinical characteristics with smallpox, including lesions and rashes, monkeypox symptoms are frequently mistaken for those of other poxviruses such as chickenpox and cowpox. Consequently, accurate early diagnosis of monkeypox by healthcare professionals remains challenging. Automated monkeypox identification using Deep Learning (DL) techniques presents a promising avenue for addressing this challenge. In this study, a modified deep convolutional neural network (DCNN) model named MpoxNet is proposed for the identification of monkeypox disease. The performance of MpoxNet is evaluated against established DCNN models, including ResNet50, VGG16, VGG19, DenseNet121, DenseNet169, Xception, InceptionResNetV2, and MobileNetV2. This study addresses the pressing challenge of monkeypox identification by proposing MpoxNet. With the aim of enhancing early detection and containment efforts, MpoxNet's performance is evaluated against established DCNN models across two distinct datasets: MSLD and MSID Dataset. Results reveal MpoxNet's superior test accuracy of 94.82% on the MSLD Dataset, surpassing other models. However, evaluation on the MSID Dataset highlights variations in performance, emphasizing the influence of dataset characteristics. Additionally, the introduction of the Swin Transformer model demonstrates exceptional performance on the MSLD and the MSID Dataset and, achieving an accuracy of 98%. These findings underscore the importance of considering diverse datasets and leveraging advanced techniques for robust monkeypox detection systems. Integration of MpoxNet with a mobile application offers a promising solution for rapid and precise monkeypox disease detection, providing valuable insights for future research and real-world deployment strategies to effectively combat the global spread of monkeypox.

Keywords

Mobile Application, Monkeypox, Swin Transformer, Skin Lesion Images, Transfer Learning.

DOI: 10.9781/ijimai.2024.11.001

I. Introduction

Recently, the global monkeypox epidemic has been threatening humankind as the world begins to recover from the COVID-19 pandemic. In 2022, many countries stated that they had experienced an outbreak of monkeypox, illustrating yet another global concern following the effects of COVID-19 in 2020. A report from the World Health Organization (WHO) highlights that the epidemic affects global public health seriously but has refrained from classifying it as a public health emergency currently [1],[2]. Monkeypox is a disease that can spread from animals to humans. It comes from the genus Orthopoxvirus. In terms of its clinical manifestations, it is clinically quite like chickenpox, measles, and smallpox [3]. However, human-to-human transmission is also a prevalent transmission mode [4]. In 1958, in a laboratory in Copenhagen, Denmark, the virus was discovered for the very first time in the body of a monkey [5]. In 1970, during an

Please cite this article as:
S. Sadesh, R. Thangaraj, P. Pandiyan, R. Devi Priya, P. Naveen. Identification of Monkeypox Disease Based on MpoxNet and Swin Transformer Models Using Mobile Application, International Journal of Interactive Multimedia and Artificial Intelligence, (2024), http://dx.doi.org/10.9781/ijimai.2024.11.001

enhanced campaign to eradicate smallpox, the Democratic Republic of the Congo recorded the first case of monkeypox, which was identified in humans and occurred throughout the campaign [6]. Many people who live close to tropical rainforests are susceptible to contracting monkeypox, which is often spread throughout the central and western regions of Africa. The virus spreads through close communication with an infected person, animal, or object. This disease is spread via bodily contact, animal bites, respiratory droplets, and mucous membranes of the eyes, nose, and mouth [7]. Fever, bodily aches, and exhaustion are some of the early-stage symptoms that people who have been infected with monkeypox may experience. The long-term impact of monkeypox is a red bump on the skin [8]. As a result of its rarity and the similarity of its rash to other diseases, early identification of monkeypox has proven to be extremely difficult for medical specialists. On the other hand, the confirmatory PCR test is not used very often either. Even though 3–6% of people who got monkeypox died in the recent outbreak [2], it is crucial to find the virus early, find out who has been in contact with it, and keep that person from spreading it to other people. Although the mortality rate from monkeypox infection is modest ranging between 1% to 10% [7], early diagnosis of monkeypox will help in patient quarantine and origin tracing for efficient limitation of monkeypox within the neighborhood. In the past decade, various artificial intelligence (AI) tools, particularly deep learning approaches, have seen widespread use in various medical image analyses [9]. Recently, AI techniques significantly contributed to diagnosing COVID-19 and ranked the severity from multimodal medical images such as computed tomography (CT), chest X-ray, and chest ultrasound [10]–[14]. This accomplishment encourages researchers to try artificial intelligence methods for diagnosing monkeypox from digitized skin images. This research work presents a modified MobileNetV2 called MpoxNet and Swin transformer model to discover monkeypox disease. Furthermore, this research analyzes the presented model's effectiveness by benchmarking it with various DCNN models such as ResNet50, VGG16, DenseNet121, DenseNet169, Xception, VGG19, InceptionResNetV2, and MobileNetV2. The main focus of this paper is as follows:

- A lightweight DCNN model named MpoxNet and Swin transformer is developed to detect monkeypox disease using image data.
- The MpoxNet and Swin transformer model performance are evaluated with different transfer learning CNN models, including ResNet50, VGG16, DenseNet121, DenseNet169, Xception, VGG19, InceptionResNetV2, and MobileNetV2.
- A comprehensive evaluation is conducted on the models considering the performance metrics such as accuracy, precision, recall, and F1-score.
- A mobile application is developed integrating the Swin transformer model specifically for rapid assessment of monkeypox disease.

The following sections of this paper illustrate the following: Section II discusses the related work carried out in skin lesion disease identification. Section III presents the materials and proposed method for accomplishing the task. It also illustrates some of the existing DCNN models in this section. Section IV presents the findings and analysis, while Section V concludes the work.

II. Related Works

The optimal solutions have yet to be fully realized for the limited challenge of monkeypox detection utilizing skin lesion images. The difficulty of locating image datasets to conduct research in this specialized sector has substantially decreased due to the publication of a few research articles on the detection and classification of monkeypox. Another area for improvement is the shortage of exhaustively annotated datasets with specific features for novel and distinct object categories. Gessert et al. [15] developed a patch-based attention method for classifying skin lesions from high-resolution images using three networks that had already been trained. To address the problem of asymmetrical class distributions, they utilized techniques such as class-specific loss weighting, oversampling, and balanced batch sampling. Additionally, they suggested adopting a diagnosis-guided loss technique, which considers the algorithm used for ground-truth annotation.

Similarly, Kawahara et al. [16] used the DERMOFIT dataset to construct a skin lesion identification system based on AlexNet transfer learning. Convolutional layers have been added on top of the fully connected layers present in the model before it was pre-trained. This was done so that the trained weights of the fully connected layers could be used as filters. Its 81.8% accuracy rate was higher than other competing algorithms using the same dataset.

Mohamed et al. [17] provided a strategy for the categorization of skin lesions, utilizing a transfer learning technique by training all the layers of the MobileNet and DenseNet models. The imbalanced dataset was addressed using data augmentation and down-sampling, resulting in increased performance and ultimately achieving 92.7% accuracy in MobileNet. While the study demonstrates improved accuracy on the specific dataset used (HAM10000), it's unclear how well these models generalize to unseen data from different sources or populations. Without external validation on diverse datasets, it's challenging to ascertain the robustness of these models. In [18], a Full Resolution Convolutional Network (FRCN) carries out image segmentation, and DL models such as Inception-ResNet-V2, ResNet-50, InceptionV3, and DenseNet-201 perform skin lesion classification. To address the imbalance issue, the framework carried out a stratified five-fold cross-validation and employed a weighted class technique. The size of the dataset was increased by inverting and rotating the data. When compared to all other models on the ISIC 2018 dataset, ResNet's diagnostic accuracy was the highest at 89.28 %. The evaluation primarily focuses on datasets from the International Skin Imaging Collaboration (ISIC), which do not fully represent the diversity of skin lesions encountered in clinical practice globally. Also, dermoscopy images can vary significantly in quality, lighting conditions, and presence of artifacts. Assessing the model's performance on images with varying quality and artifacts would provide insights into its robustness and potential limitations in real-world scenarios. To recognize six distinct types of skin lesions, Jinnai et al. [19] proposed a faster region-based CNN (FRCNN) model built on top of VGG-16. The use of digital cameras resulted in the collection of 5846 photos from 3551 patients. After that, annotations and enhancements were added to the photographs. A total of 20 dermatologists' predictions were compared to the study's findings. The FRCNN achieved a higher accuracy rate of 86.2 % than the dermatologists. The study mentions that the neural network was trained using clinical images from only one institution, which introduce biases.

The MobileNet architecture is altered to facilitate the development of Sae-Lim et al. [20] models for identifying skin lesions. The MobileNet's layers from the fifth to the tenth were dropped and balanced with a fully connected dropout layer that includes the Softmax activation function. This both decreases the total number of factors and accelerates the process. The authors researched to determine the effectiveness of data augmentation and upsampling, and concluded that both techniques were beneficial. The evaluation of the proposed model is solely based on the HAM10000 dataset. External validation on different datasets would provide more robustness to the findings and demonstrate the generalizability of the model. The comparison is primarily between the proposed modified MobileNet and the traditional MobileNet. It would be valuable to compare the proposed model with other state-

of-the-art architectures for skin lesion classification to assess its competitiveness. Transfer learning from the MobileNet model was used in [21] to categorize seven different skin lesions taken from the HAM10000 dataset. Several distinct data augmentation methods were utilized, and the result was a classification accuracy of 83.1 %. Harangi et al. in [22] reported a weighted ensemble-based network which achieves 83.8 % accuracy on the ISBI 2017 dataset and the base models for this network are AlexNet, VGG, and Inception. The base CNNs' final fully connected and classification layers were removed and replaced with a joint fully connected layer with softmax activation to increase the prediction accuracy. The dataset mentioned consists of 2,000 training images, with a significant class imbalance (e.g., 1,372 nevus, 374 melanoma, 254 seborrheic keratosis). Class imbalance can significantly affect the performance of machine learning models. Also, it doesn't delve into the generalization capability of the proposed model. Steppan et al. [23] created a transfer-learning-based ensemble model for categorizing nine types of skin lesions. Six different datasets were employed, and data augmentation methods is used to enhance the dataset size. The images are pre-processed to remove any unwanted dark regions. Additionally, three techniques were employed to equalize the classes. EfficientNet outperformed other base models, with an accuracy of 63.4% reported. A technique is proposed for classifying seven distinct types of skin lesions employing deep neural networks. This method employs five state-of-the-art models including ResNet, DenseNet, Xception, ResNeXt, and SeResNeXt. The classification is carried out using the HAM10000 [22] and ISIC datasets [24]–[26]. Data augmentation methods were used to address the issue of large discrepancies in the collected data.

In Hu et al. [27], the authors design a DCNN architecture in which LeNet-5 is used as the backbone for learning the X-ray image features. This architecture includes a powerful single-layer learning machine that acts as the classifier to distinguish standard cases, viral pneumonia, and COVID-19. Similarly, by employing different DL models, Sharma et al. [28] extract features of X-ray images from the pneumonia dataset. Besides, it illustrated that data augmentation enhances the model's accuracy. In addition, they investigated the influence of dropouts in the model. The result confirms that the model with augmentation and dropout achieves an accuracy of 90.68%, whereas the model without augmentation and dropout achieves an accuracy of 74.98%. Heidari et al. [29] detected pneumonia caused by COVID-19 by devising a convolutional neural network-based model which leverages X-ray images and achieves an accuracy of 98.80% and overall accuracy achieved by the model of 94%. Madhavan et al. [30] implemented Res-COvNet to detect the COVID-19 virus. The transfer learning approach is the cornerstone of this DL model, which includes ResNet-50. The ResNet-50 extracts significant X-ray image features and expands the architecture with a classification layer. The proposed model achieves 96.20% accuracy for determining typical COVID-19, bacterial pneumonia, and viral pneumonia diseases.

In addition to the COVID-19 virus detection, periodic investigations employed the DL model for different disease detection like herpes, chicken pox, etc., For instance, Sandeep et al. [31] analyzed various skin infection detection models like Psoriasis, Melanoma, Chicken Pox, Vitiligo, Acne, Ringworm, Herpes, and Lupus employing DL models. Furthermore, they classified eight different skin lesion disease classes through a CNN model and analyzed the results with the support of the pre-trained VGG-16 model. Similarly, Glock et al. [32] use the transfer learning method to detect measles disease. They harnessed the ResNet-50 model which had the input of a diverse rash image dataset, and accomplished the results of sensitivity, specificity, and accuracy at 81.7%, 97.1%, and 95.2%, respectively. Ahsan et al. [33] gathered images of Measles, Chickenpox, Monkeypox, and normal by employing web data mining approaches, and the specialists demonstrated its performance. Subsequently, the authors assessed a transfer learning technique using the VGG-16 model, considering two approaches. The first approach evaluated the identification of images into different disease classes, such as chickenpox and monkeypox. Whereas the second approach employs different data augmentation methods to enhance the dataset size. The model obtained an accuracy of 97% with data augmentation in identifying the monkeypox. In contrast, without applying data augmentation the accuracy of the model dropped to 78%. Moreover, Chiranjibi Sitaula and Tej Bahadur Shahi [34] investigated 13 distinct pre-trained DL models to detect monkeypox disease via images. The authors analyzed the comparison results and devised an ensemble method that hybrids Xception and DenseNet169.This hybrid method produces an accuracy of 87.13%. The author employed 13 pretrained models and enhanced their performance by incorporating custom layers, resulting in an average accuracy rate of 85.44%. However, they acknowledged that the relatively low accuracy could be attributed to the absence of a validation step, which is crucial for mitigating overfitting and ensuring the attainment of satisfactory outcomes.

The author utilized a combination of CNN models and machine learning algorithms for diagnosing Monkeypox disease based on skin images. Three CNN models, namely VGG16Net, GoogleNet, and AlexNet, were employed to extract features from the images. Classification algorithms including Naïve Bayes (NB), Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest were then utilized as classifiers. The findings revealed that when combined with the VGG16Net CNN model, the naïve Bayes algorithm outperforms with the highest accuracy level of 91.11% [35].

S. Maqsood and R. Damaševičius propose a deep learning-based approach for monkeypox detection, achieving high accuracy and performance. However, it relies solely on public datasets for evaluation, potentially limiting its generalizability to real-world scenarios with diverse populations and contexts [36]. The proposed framework for multiclass skin lesion classification combines deep learning and optimization techniques to achieve high accuracy by Hussain et al. [37]. While augmentation improves accuracy, it also leads to a significant increase in redundant features, impacting efficiency. The use of KNN classifiers results in a drop in classification accuracy, requiring further analysis and optimization. The fusion process enhances accuracy but at the cost of increased computational time due to the larger number of predictors.

The above discussion indicates that there is limited research in the field of Monkeypox disease diagnosis, particularly in the development of a refined transfer learning (TL) based diagnostic model. There are notable gaps identified in the existing research that should be addressed. One significant limitation is the lack of large datasets and memory constraints. Previous studies have often relied on small and constrained datasets that may not encompass the full range of variations observed in monkeypox cases. This insufficiency in dataset diversity can lead to overfitting issues and ultimately reduce the overall accuracy of the models. Moreover, a significant aspect that is lacking in most prior studies is the absence of a validation set. The failure to include a validation set prevents the assessment of whether the models are excessively fitted to the training data, potentially leading to reduced accuracy when applied to new and unseen data. Therefore, it is crucial to address this gap by creating a model that is computationally efficient, provides comprehensive model interpretation, incorporates techniques for generalization and regularization to mitigate overfitting. Our proposed model is the modified version of DCNN (MpoxNet) and Swin transformer, which provides better accuracy compared to the other models used in disease detection, and it is a lightweight model which can be easily deployed. The proposed work seeks to contribute to the field by providing a novel and efficient approach to accurately identify and

Fig. 1. (a). Sample Images of the MSLD dataset.

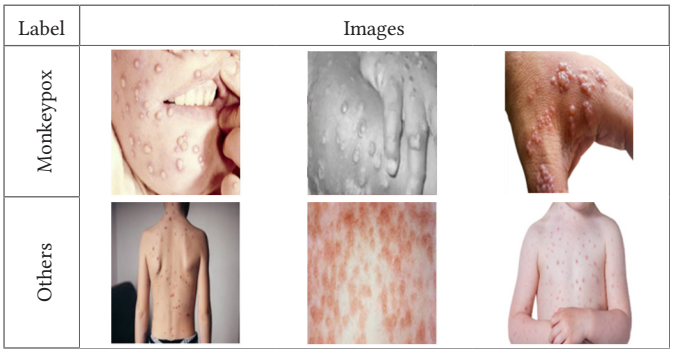


Fig. 1. (b). Sample Images of the MSID dataset.

classify monkeypox based on image data, potentially improving the diagnosis and treatment. Our research leverages the capabilities of modern smartphones and their cameras to design a self-contained diagnostic solution. Emphasizing efficiency, the selected model is custom-designed to ensure both reliable and optimal performance on low-performance devices, such as smartphones.

III. Materials and Methodology

A. Dataset

Monkeypox Skin Lesion Dataset (MSLD) [38], a freely available dataset, is used in this research. The dataset contains images of patients with monkeypox and non-monkeypox conditions (measles, chickenpox) focusing on various body parts. The dataset consists of 228 images, among which 102 belong to the 'Monkeypox' class, and the remaining 126 represent the 'Others' class (i.e., chickenpox and measles). Since the dataset has a limited number of images, it is inadequate for the classification task. Therefore, several data augmentation methods have been applied to increase the dataset size, and augmented images are provided in a separate folder in the same dataset. For the class of Monkeypox, there are 102 original images and 1428 augmented images. For other diseases such as Chickenpox and measles, there are 126 original images and 1764 augmented images. Introducing the Monkeypox Skin Images Dataset (MSID), a valuable resource in the battle against the latest monkeypox outbreak. This dataset, meticulously compiled by Diponkor Bala and Md. Shamim Hossain, offers a crucial tool for healthcare professionals worldwide. With early diagnosis being paramount in mitigating the rapid spread of monkeypox, MSID [38] provides a diverse collection of skin images aimed at aiding in the disease's detection. Comprising four distinct classes - Monkeypox, Chickenpox, Measles, and Normal - this dataset offers a comprehensive array of visual data sourced from reputable internet-based health websites. Diponkor Bala and Md. Shamim Hossain, have meticulously curated this dataset to facilitate research and development in the field of disease detection and diagnosis. A few representative samples of the two datasets are shown in Fig. 1.a. and Fig. 1.b.

B. Training and Testing Dataset

In this study, the entire image dataset is split into an 80:20 ratio, with 80% of the data used for training and 20% for testing. Additionally, the training dataset is further divided into two categories: 80% of the images are exclusively used for training purposes, while the remaining 20% are allocated for validation to assess model overfitting. During the training process, each image is resized to meet the dimensional requirements of the model, with the images resized to 224 x 224 for models such as ResNet50, VGG16, VGG19, MobileNetV2, MpoxNet, and Swin Transformer. Furthermore, hyperparameters such as a learning rate of 0.001, a batch size of 32, 25 epochs, and the Adam optimizer are used for model training.

C. Transfer Learning

Transfer learning is an approach that enables us to leverage the knowledge gained when solving certain problems and apply it to other related but different problems. Medical image categorization for rare or emergent disorders is a prime example of a problem where there is insufficient training data to train a model from scratch; hence, transfer learning is highly effective in such cases. There are primarily two methods by which a pre-trained model can be utilized for a specific task. The first method involves treating the previously-trained model as a feature extractor, without adjusting its internal weights, and then training a classifier on top of it. In the second method, the entire network or a portion of it is fine-tuned to perform optimally on the newly assigned task. Consequently, the model's weights are adjusted during training, with the pre-trained values serving as initial values for the current task. In this instance, fine-tuning the final layer of the CNN and using the pre-trained models as feature extractors is preferred due to the limited number of images in the monkeypox category.

D. MpoxNet

A Depthwise Separable Convolution (DSC) network serves as the foundation of the proposed work, integrated into MobileNet [39]. DSCs are factorized convolutions that combine point-wise and depthwise convolutions to generate output channels. Point-wise convolution convolves the kernels of each filter with each input channel. This research employs depthwise separable convolution layers to develop a modified MobileNetV2 (MpoxNet) model from scratch. The building block of the MobileNetV2 model is presented in Fig. 2(a), while the modified MobileNetV2 (MpoxNet) model's building block is depicted in Fig. 2(b). The modified MobileNetV2 model incorporates Global Average Pooling (GAP) layers, batch normalization (BN), and the ReLU activation function. Node dropout is achieved using the dropout layer. Finally, a new dense layer is added for disease classification. During training, the weights of the newly added layer are updated using the monkeypox dataset, and in the fine-tuning stage, the proposed model utilizes pre-trained weights from the ImageNet dataset. The new dense layer classifies monkeypox disease using the sigmoid activation algorithm.

In the MobileNet model, convolutions factorize a conventional convolution into a depthwise convolution and a 1x1 pointwise convolution. DSCs are a form of factorized convolutions that form the foundation of this model. Each input channel in a MobileNet is subjected to a singular filter in the context of depthwise convolution. To merge the results of the depthwise convolution with the pointwise convolution, a 1x1 convolution is performed on each set of results. A typical convolution takes the inputs in a single step, filters them, and combines them into a new set of outputs. This is separated into two layers using depth-separable convolution where the first layer is responsible for filtering, and the second layer is involved in combining

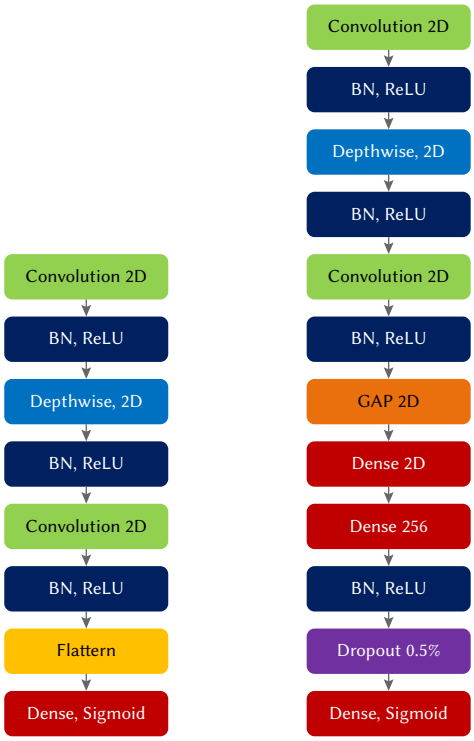


Fig. 2. (a) Core layers of MobileNetV2 (b) Core layers of Modified MobileNetV2 (MpoxNet).

the results. This factorization can considerably decrease the size of the calculation and the model. Fig.3 illustrates how a standard convolution (Fig.3 (a)) can be broken down into a depthwise convolution (Fig.3 (b)) and a pointwise convolution (Fig.3(c)) using factorization.

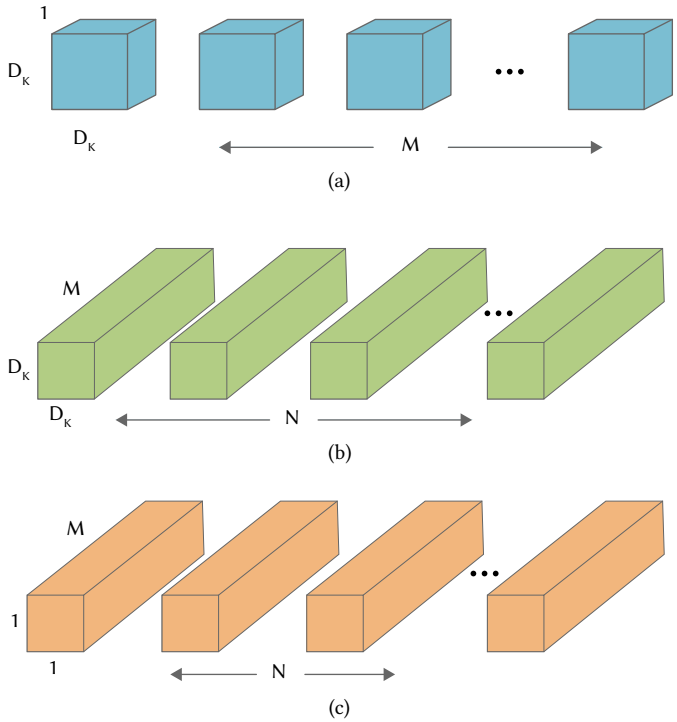


Fig. 3. (a) Standard Convolution Filters, (b) Depthwise Convolutional Filters, (c) Pointwise Convolution filters.

A typical convolutional layer receives as input a square feature map with dimensions (D_F, D_F, M) and outputs a square feature map with dimensions (D_G, D_G, N), where D_F denotes the spatial width and height of the input feature map, the number of input channels is represented by M. Whereas the output feature map's spatial width and height is denoted by DG and N is the number of output channels. Parameterizing the typical convolutional layer is a convolution kernel K with dimensions $D_K \times D_K \times M \times N$, where D_K is the presumed square spatial dimension of the kernel.

Eq. (1) represents the output feature map for standard convolution produced when stride one and padding are considered:

$$G_{K,l,,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \tag{1}$$

The computing cost of conventional convolutions is given in Eq. (2).

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \tag{2}$$

The computational cost multiplies input channels M, output channels N, kernel size $D_k \times D_k$, and feature map size $D_F \times D_F$. All these concepts and their interplay are considered by MobileNet models. First, it employs DSCs which separate the number of output channels from the size of the kernel. The typical convolution process filters the feature using convolutional kernels and combines features to generate a new representation. By using factorized convolutions called DSCs, the filtering and combining steps can be split into two steps. This reduces the amount of work that needs to be done on the computer by a large amount.

Convolutions performed depthwise and pointwise are the building blocks of depthwise separable convolution, which consists of two layers. The depth of the convolutional filter determines how much weight is given to each input channel (input depth). After that, a pointwise convolution, which is a straightforward 1×1 convolution, is applied to the output of the depthwise layer to produce a linear combination. For both layers, MobileNet employs nonlinearities known as batchnorm and ReLU. Convolution in the depthwise direction using one filter for each channel of the input (the depth) can be represented by Eq. (3):

$$\hat{G}_{K,l,,n} = \sum_{i,j,m} \hat{K}_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \tag{3}$$

where K is the depthwise convolutional kernel with the dimensions $D_K \times D_K \times M$, and the m^{th} filter in K is applied to the m^{th} channel in F to form the m^{th} channel of the filtered output feature map G. where G is the filtered output feature map. Eq. (4) is the computing costs associated with depthwise convolution:

$$D_K \cdot D_K \cdot M.. D_F \cdot D_F \tag{4}$$

In comparison to ordinary convolution, depthwise convolution is incredibly efficient. On the other hand, it does not integrate the input channels to generate new features; instead, it filters them. DSCs are expensive and are represented by Eq. (5), equivalent to adding together the convolutions performed in depth and one dimension.

$$D_K \cdot D_K \cdot M.. D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \tag{5}$$

By recasting convolution as a two-stage process consisting of filtering and combining, it can reduce the amount of computing required, as given in Eq. (6):

$$\frac{D_K \cdot D_K \cdot M.. D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2} \tag{6}$$

MobileNet uses 3x3 depthwise separable convolutions, which require 8 to 9 times less processing than normal convolutions. Due to the low computational cost of depthwise convolutions, additional factorization in spatial dimensions discussed in [37], [38], does not yield significant computational savings.

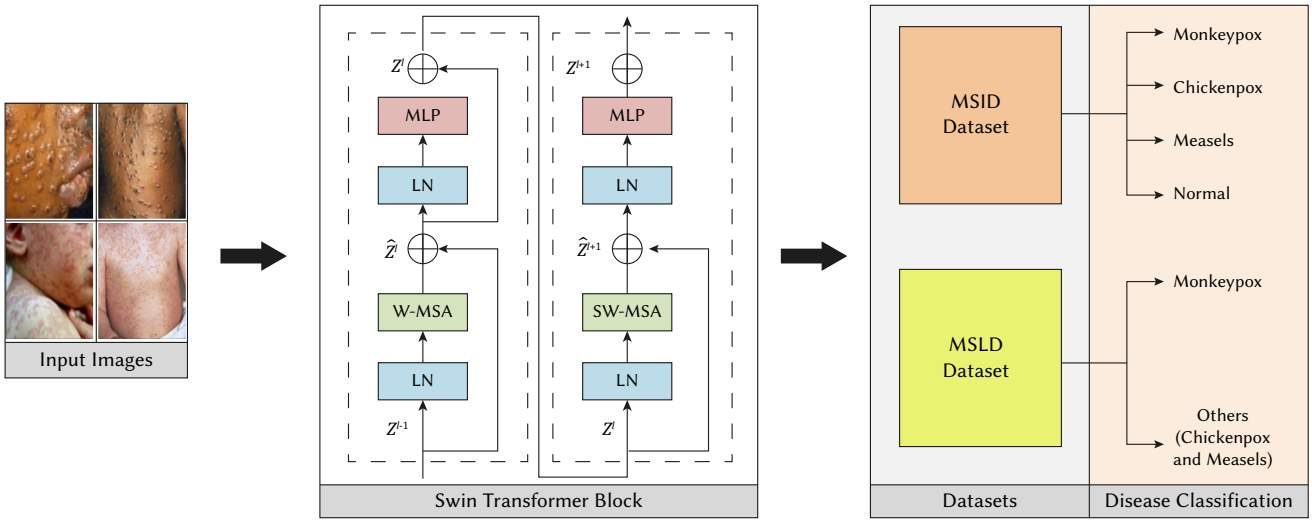


Fig. 4. Proposed Architecture of Swin Transformer.

E. Swin Transformer

For image classification using the Swin Transformer architecture (Fig. 4), the encoding and decoding process is summarized as follows:

1. **Encoding Process:**
   i. Take an input image of *size* ($H \times W \times C$), where H is the height, W is the width, and C is the number of channels (e.g., RGB channels).
   ii. Divide the input image into non-overlapping patches of fixed size. Each patch is treated as a token.
   iii. Each patch is linearly embedded into a lower-dimensional space using a learnable linear projection. This step converts the pixel values of each patch into a high-level feature representation.
   iv. Augment the patch embeddings with positional embeddings to provide spatial information to the model. Positional embeddings encode the spatial relationships between patches in the image.
   v. Process the patch embeddings in a hierarchical manner through multiple stages, with each stage consisting of several transformer blocks.
   vi. Within each stage, apply multiple transformer blocks to process the patch embeddings. Each transformer block consists of self-attention layers followed by feedforward layers. Self-attention helps capture global dependencies within the patches.
   vii. After processing within each transformer block, merge the patch embeddings to form larger patches. This step allows the model to capture both local and global information effectively.
   viii. Before passing the merged patches to the next stage, apply down sampling operations to reduce spatial resolution. Down sampling helps increase the receptive field while managing computational complexity.
2. Decoding Process:
   i. Perform global average pooling over the output of the final stage. This aggregates information from all patches into a single feature vector.
   ii. Pass the aggregated feature vector through a classification head, which typically consists of one fully connected layer followed by a sigmoid activation function for MSLD dataset. For the MSID dataset, two fully connected layers are followed by a softmax activation function.
   iii. The output of the softmax layer represents the predicted probabilities of the input image belonging to each class in the classification task.

The Swin Transformer primarily focuses on encoding the input image into a compact representation and then uses this representation to make predictions about the image's class. The decoding process is relatively straightforward, as it involves only global pooling and a classification head.

F. Pre-Trained Deep Learning Models

The effectiveness of the proposed model in identifying monkeypox disease is evaluated by comparing its results with the well-known deep learning models, including ResNet50, VGG16, DenseNet121, DenseNet169, Xception, VGG19, InceptionResNetV2 and MobileNetV2.

1. ResNet

He et al. [40] proposed the ResNet network which served as the foundation for the ILSVRC 2015 and COCO 2015 classification challenges. They achieved a classification error rate of 3.57% on ImageNet, making the model the most accurate of all competitors. Deep residual learning was inspired by the failure of several nonlinear layers to learn identity mappings and degradation problems (ResNet). The design of ResNet is known as a network-in-network (NIN), and it is based on many stacked residual units. The network is made up of these residual units, which are like a set of building blocks. The ResNet architecture is constructed out of various building blocks formed by a collection of residual units. Convolutional and pooling layers make up the residual units. ResNet's architecture is comparable to that of the VGG network consisting of 33 filters, although it is approximately eight times deeper than the VGG network. This is because global average pooling is used instead of fully connected layers. The residual module of ResNet is updated to get better precision, which was accomplished by modifying it to make use of identity mappings. In this work, the ResNet model is created with 50 layers, and pre-trained weights from ImageNet are loaded into it. Finally, the existing fully connected layer is replaced with the new dense layer with a sigmoid activation function, and it is trained using the monkeypox dataset. In our work, sigmoid activation function is used to classify whether the disease is monkeypox or others (chickenpox, measles).

2. VGG Network

The CNN model, the VGGNet was developed in [41] specifically for the ILSVRC-2014 competition. The model achieved a top-5 error rate of 7.5% on the validation set, securing the second position in the competition. In VGGNet, the depth of the architecture is increased to 16 and 19 layers by employing tiny 3x3 convolution filters. Despite the increased depth, the number of parameters is reduced, resulting

in a more efficient network. These two new architectures are named VGG-16 and VGG-19. The VGG model is made up of convolution layers that have multiple consecutive 3x3 convolutions. After that comes a 2x2 max pooling layer, followed by two fully connected layers, with the final layer functioning as the softmax output. In most cases, the model is constructed with only three convolutional layers of size three stacked one on top of the other in increasing depth. The max pooling method shrinks the volume size (down sampling). To accomplish the fine-tuning of the VGG-16 and VGG-19, the original fully connected layer is truncated and a new dense layer is added. Furthermore, both the models utilize the ImageNet weights and a new dense layer is retrained with a new target dataset whereas the weight values are also updated. The sigmoid activation function is applied in the final layer for binary classification.

3. MobileNetV2

The MobileNetV2 network is constructed using the MobileNetV1 model as a foundation, and nonlinearities in the model's narrow layers that contain its constituent parts were addressed in this model [42]. Two new features have been added to the MobileNetV2 over its predecessor. First, some bottlenecks can form between layers linearly; second, shortcuts can form between bottlenecks. The architecture of the MobileNetV2 network includes depth-wise separable filters and their combination stages. This model uses the 1x1 deep convolution filter for each layer and inputs are divided into two layers for examination using depthwise separable convolutional filters. The characteristics collected by filtering them out are joined in the combination stage to create a new layer. ReLU and Batchnorm activation functions are used in the framework of the MobilNetV2 model. MobileNet V2 is fine-tuned by loading the ImageNet pre-trained weights and including the customized dense layer to the top layer. The model's newly added layer is retrained using the monkeypox dataset, and its weight values are changed to reflect the new dataset. The target dataset's classes are matched to the number of classes in the new layer, and the sigmoid function is turned on to classify diseases.

4. DenseNet

The DenseNet121 is one of the models from DenseNet family which contains 121 layers. It is the most popular image classification model. The models were initially created in the Torch framework, and are then converted to the Caffe format. The ImageNet database serves as the pre-training set for all DenseNet models. Like this, the well-known DenseNet169 architecture is primarily employed for classification problems. The architecture consists of a transition layer, dense layers, maxpool layers, and convolutional layers. The architecture includes the ReLU as an activation function, and softmax as an activation function at the final layer. The convolutional layer extracts the features from the image, while the maxpool layer reduces the dimensionality of its inputs. Subsequently, the flatten layer converts the output into a single 1D array, resembling an artificial neural network input, which is then processed by the fully connected layer for classification [43].

5. Xception

The Xception [44] model, an extension of the inception network, employs separable convolution layers in place of traditional convolution layers. Its 36 convolution layers are utilised to extract the features from the input. The Xception model incorporates a depth-wise separable convolutional neural network alongside standard convolutions. The term "Separable Convolutional" pertains to the depth-wise separable convolution, which serves as a replacement for the traditional convolutional layer. Its primary objective is to reduce computational expenses while performing similar functions as the conventional convolutional layer. In the initial depth-wise separable convolutional layer, the point-wise convolution is applied after the depth-wise convolution. The Xception model utilizes a customized depth-wise separable convolution layer, where a 1x1 convolution is performed on each channel before applying the depth-wise convolution operator.

6. InceptionResNetV2

Relative connections and an enhanced version of the Inception architecture are combined in the InceptionResNetV2 design, which has 164 deep residual layers. A series of filters are concatenated to form each branch (1x1, 3x3, 5x5, etc.). The split-transform-merge architecture of the inception module is regarded to have good representational capabilities in its deep layers [45]. The architecture is constructed by fusing several construction components. origin, stem, Average Pooling, Dropout, and Softmax blocks, as well as 20 x Inception-ResNet-B, Reduction-B, 10 x Inception-ResNet-C, and 10 x Inception-ResNet-A, are among the featured blocks [46]. The output of each Inception-ResNet block can be scaled in several ranges thanks to the scale factors employed in the blocks. The network will accept images up to 299 x 299 x 3 in size.

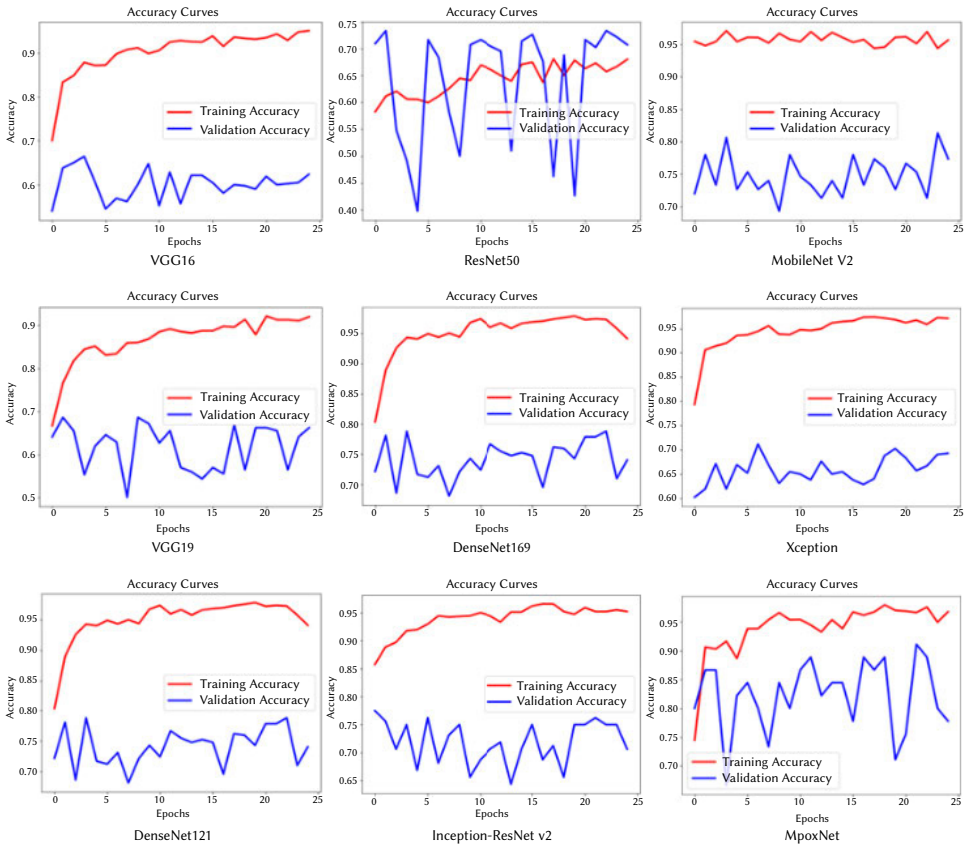
G. Performance Evaluation

The model's predictive performance is assessed by utilizing various metrics, including accuracy, precision, recall, and F1-score. From [47] the formulae for the performance measures are adopted. Accuracy measures a method's ability by comparing accurately predicted cases to the overall number of cases. The precision refers to the proportion of accurately anticipated positive cases to the overall number of expected positive cases. A low percentage of false positives is associated with high precision. The recall metric measures the proportion of correctly predicted positive instances, by the total number of observations belonging to the positive class in the actual dataset. In the event of an unbalanced class distribution, the F1 Score is calculated. This is especially important when there are many false negatives in the data.

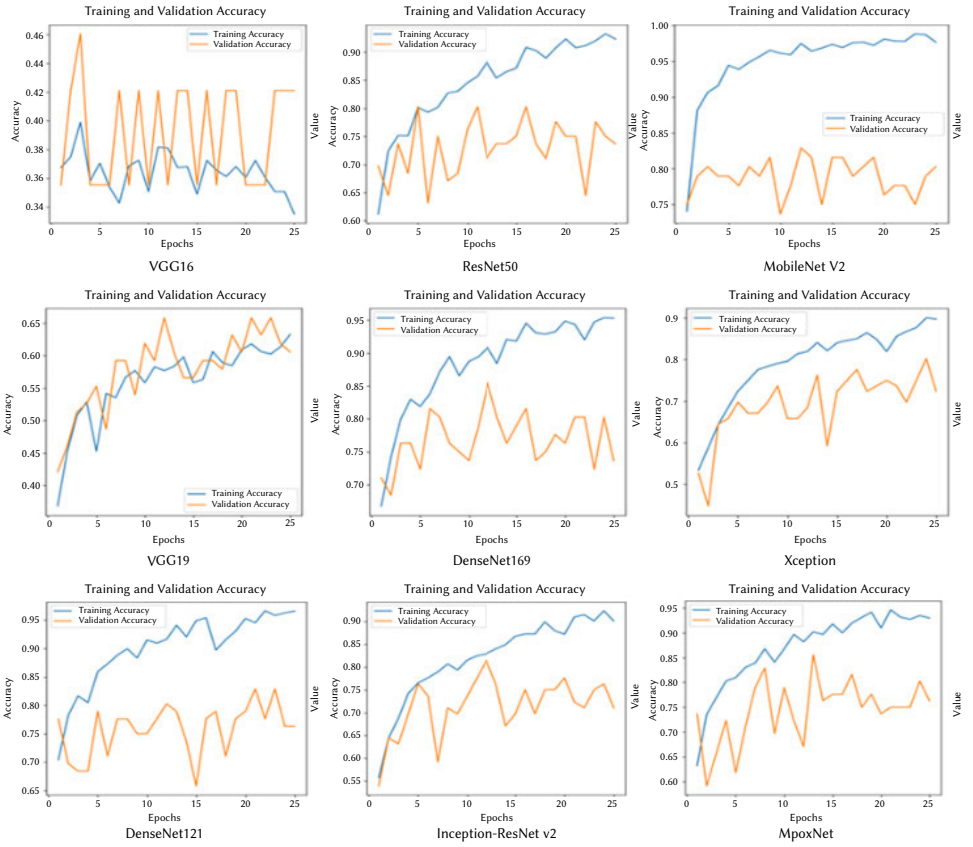
IV. Experimental Results

In our experiment, the performance of the proposed model is quantitatively compared to that of various pre-trained DL models including ResNet50, VGG19, VGG16, and MobileNetV2. Since the size of the dataset is small the deep learning models with a smaller number of trainable parameters are selected to investigate the effect of the sample size used for training. Due to the limited size of the image collection, the deep models are initialized using the weights learned from ImageNet and then fine-tuned with data. During training, different augmented images for each class are employed to compensate for the class imbalance in the original dataset. Initially, all the models are trained with 25 epochs. The training and validation variations obtained by the different models are depicted in Fig. 5. a. and Fig. 5. b. for MSLD and MSID dataset respectively.

Similarly, the loss values obtained by the different models are presented in Fig. 6. a. and Fig. 6. b. for MSLD and MSID dataset respectively. Comparing the performance of the five different models, modified MobileNet V2 consistently achieves the highest accuracy with fewer epochs. Consequently, the model contains a lesser number of parameters compared with the other models. Following the proposed model MobileNet V2 shows better performance. Moreover, ResNet 50, and VGG16 achieve satisfactory performance and VGG19 doesn't perform well with the small dataset. The effectiveness of the models is evaluated using the classification accuracy, and its comparison is depicted in Fig. 7. a. and Fig. 7. b. for MSLD and MSID dataset respectively.

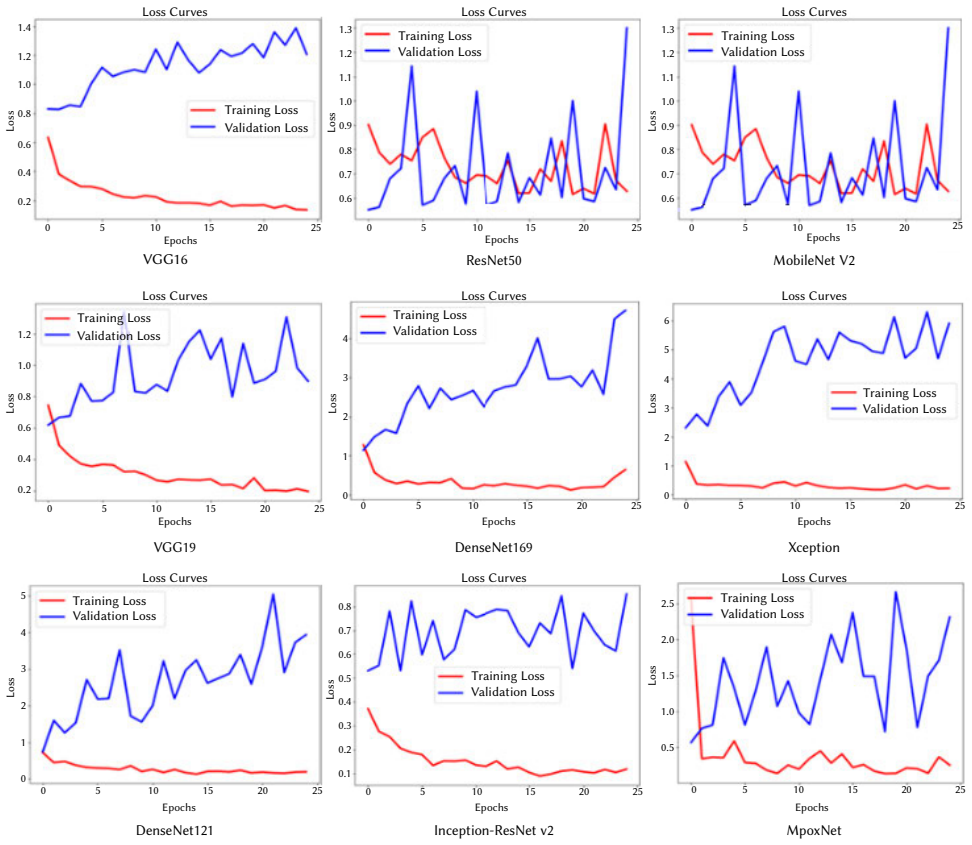


(a)

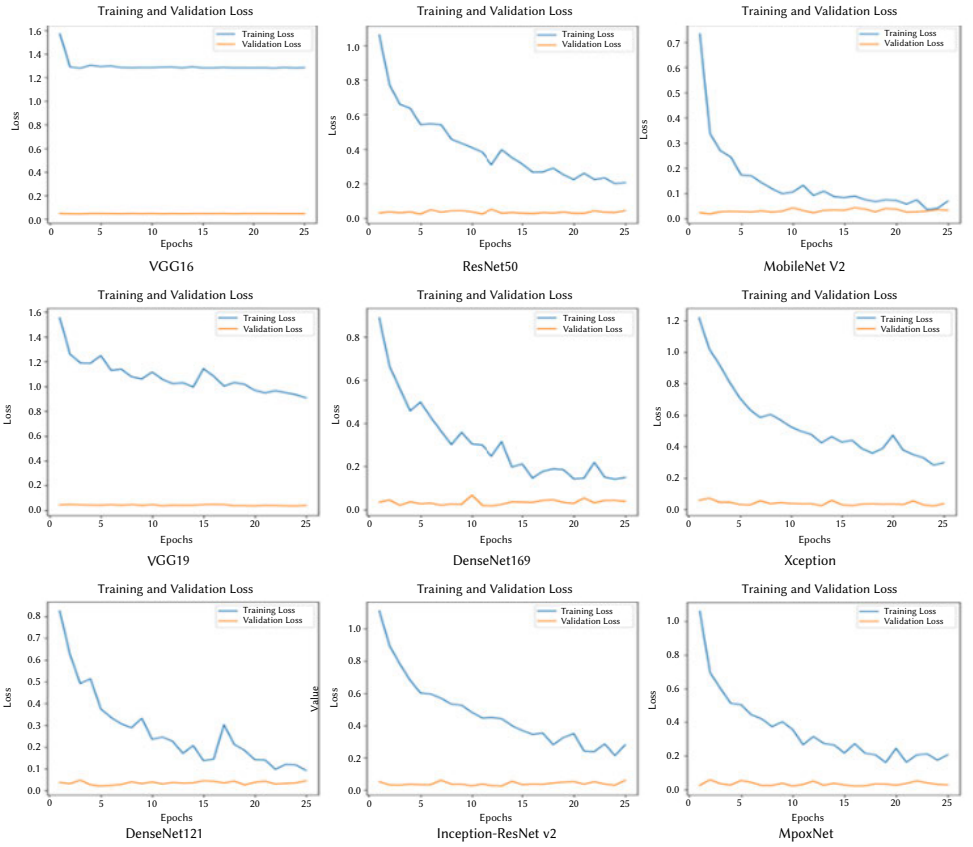


(b)

Fig. 5. (a) Accuracy variations of various DL models and MpoxNet during training and validation process for MSLD dataset. (b) Accuracy variations of various DL models and MpoxNet during training and validation process for MSID dataset.

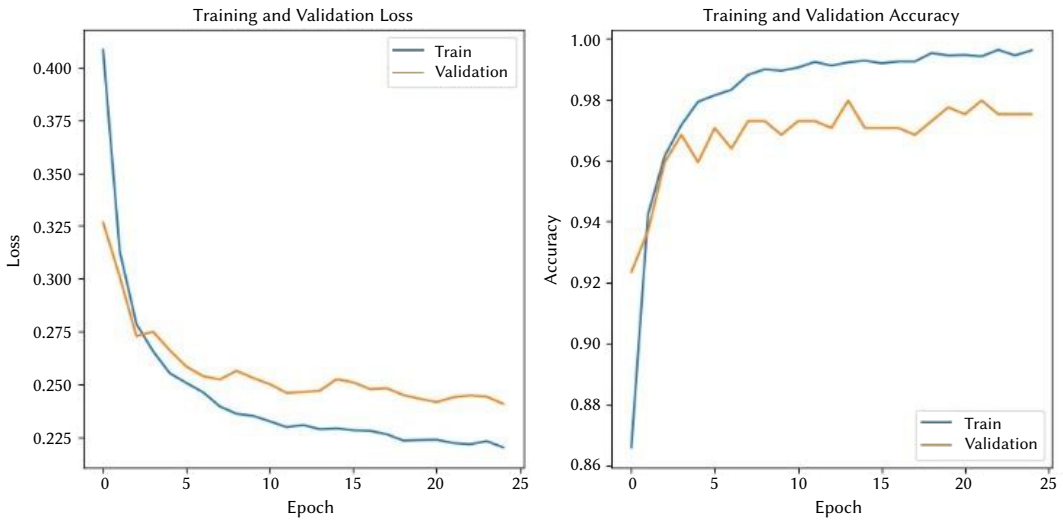


(a)

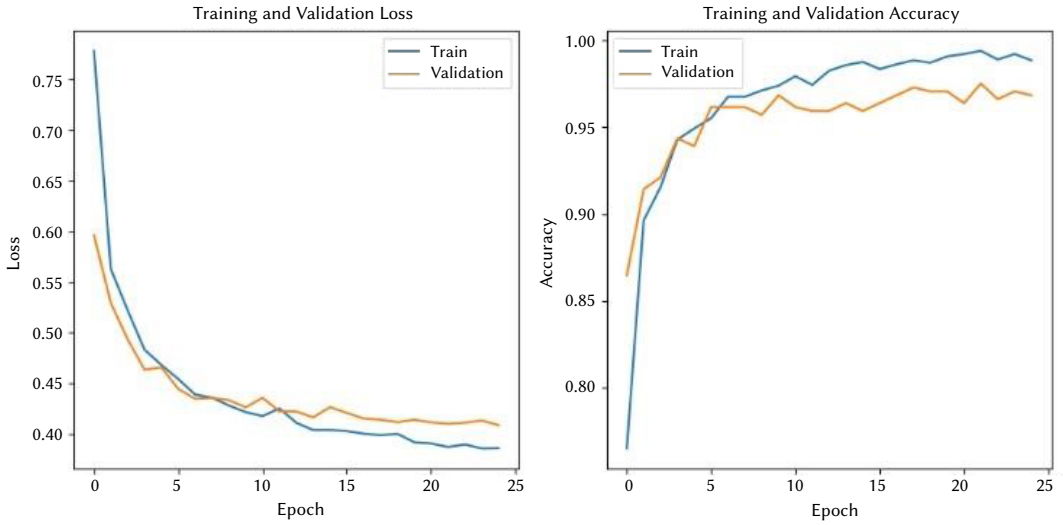


(b)

Fig. 6. (a) Loss variations of various DL models and MpoxNet during training and validation process for MSLD dataset. (b) Loss variations of various DL model and MpoxNet during training and validation process for MSLD dataset.



(a)



(b)

Fig. 7. (a) Loss and Accuracy variations of Swin Transformer model on MSLD dataset. (b) Loss and Accuracy variations of Swin Transformer model on MSID dataset.

Further analysis of the model performance is conducted by calculating its precision, recall, and F1-score and its comparison is shown in the Tables I and II.

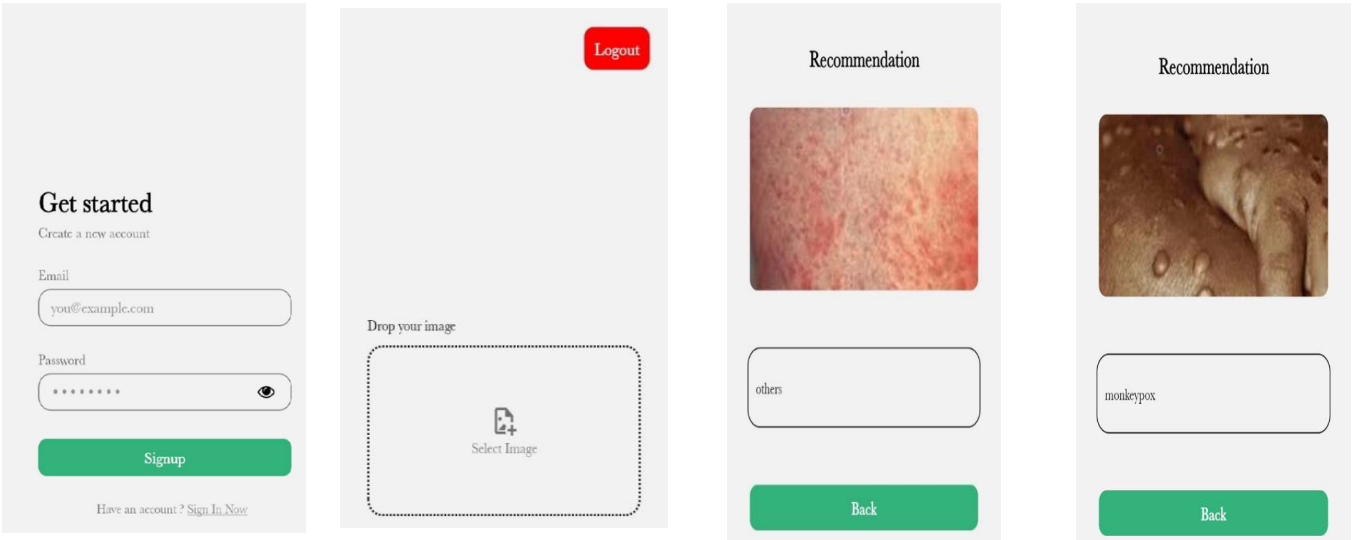
TABLE I. Performance Metrics Comparison of the Proposed Model With Other DL Models Using MSLD Dataset

Models	MSLD Dataset			
	Accuracy	Precision	Recall	F1 score
MobileNetV2	84.44	0.87	0.84	0.84
ResNet 50	82.22	0.84	0.82	0.82
VGG16	73.33	0.76	0.71	0.72
VGG19	68.89	0.68	0.66	0.66
DenseNet 121	77.78	0.78	0.72	0.74
DenseNet 169	88.38	0.89	0.88	0.88
Inception-ResNet V2	86.67	0.88	0.87	0.87
Xception	89.26	0.89	0.87	0.86
MpoxNet	**94.82**	**0.94**	**0.93**	**0.93**
Swin Transformer	**98.08**	**0.98**	**0.98**	**0.98**

TABLE II. Performance Metrics Comparison of the Proposed Model With Other DL Models Using MSID Dataset

Models	MSID Dataset			
	Accuracy	Precision	Recall	F1 score
MobileNetV2	80.36	0.78	0.75	0.75
Resnet 50	73.24	0.72	0.73	0.7
VGG16	60.87	0.62	0.64	0.64
VGG19	64.56	0.66	0.67	0.64
DenseNet 121	76.78	0.75	0.77	0.73
DenseNet 169	73.02	0.71	0.73	0.69
Inception-ResNet V2	71.10	0.71	0.62	0.63
Xception	72.49	0.66	0.75	0.66
MpoxNet	**73.54**	**0.75**	**0.78**	**0.73**
Swin Transformer	**98.84**	**0.97**	**0.96**	**0.99**

Table I and Table II provide a comprehensive comparison of the performance metrics of various deep learning (DL) models, including MobileNetV2, ResNet50, VGG16, VGG19, DenseNet121, DenseNet169,



(a) (b) (c) (d)

Fig. 8. Screenshots of Mobile Application. (a) Login Page (b) Image Uploading (c) Prediction of Monkeypox Disease Class (d) Prediction of Other Diseases Class (chickenpox and measles).

Inception-ResNet V2, Xception, MpoxNet, and Swin Transformer, using two different datasets: MSLD (Table II) and MSID (Table III). The table II presents accuracy, precision, recall, and F1 score for each DL model using the MSLD dataset. Swin Transformer exhibits the highest accuracy (98.08%), followed closely by MpoxNet (94.82%) and Xception (89.26%). These models outperform others in accurately classifying monkeypox images. It also achieves the highest precision (0.98), indicating its ability to minimize false positives. MpoxNet follows closely with a precision of 0.94. Swin Transformer and DenseNet169 both demonstrate the highest recall (0.98), indicating their ability to correctly identify true positives. It again achieves the highest F1 score (0.98), reflecting a balance between precision and recall.

Similarly, table II presents the performance metrics of DL models using the MSID dataset. Swin Transformer achieves the highest accuracy (98.84%), indicating its effectiveness in accurately classifying monkeypox images within the MSID dataset. It also exhibits the highest precision (0.97), followed by MpoxNet (0.75). Swin Transformer demonstrates the highest recall (0.96), indicating its ability to correctly identify true positives. It achieves the highest F1 score (0.99), indicating a balance between precision and recall, making it the most reliable model for monkeypox detection in the MSID dataset.

Across both datasets, Swin Transformer consistently outperforms other DL models in all performance metrics, demonstrating its robustness and reliability in monkeypox detection. MpoxNet also performs well, especially in terms of accuracy and precision, making it a viable alternative to Swin Transformer. The impressive performance of Swin Transformer suggests its potential for real-world applications in healthcare, where accurate and efficient disease detection is critical. While Swin Transformer shows promising results, further research and validation on larger and more diverse datasets are necessary to confirm its generalizability and effectiveness in real-world scenarios.

These tables highlight the superiority of Swin Transformer and MpoxNet in accurately detecting monkeypox disease, showcasing their potential to revolutionize disease diagnosis in healthcare.

V. Mobile Application

Creating a mobile application that integrates deep convolutional neural networks (CNNs) for diagnosing monkeypox disease has the potential to bring revolutionary change to the healthcare industry. This app aims to offer a convenient and precise tool for early detection, leading to improved medical outcomes. The app is designed with a user-friendly interface to ensure seamless interaction with various features and functionalities. Users can easily capture or upload images for analysis, enhancing the convenience and accessibility of the diagnosis process. The proposed modified MobileNetV2 model is integrated with the mobile application due to its memory efficiency compared to other deep learning models during deployment. Users can capture or upload images of skin lesions or rashes related to monkeypox within the mobile application. These images are then analyzed using the modified MobileNetV2 model to identify patterns and indicators of monkeypox disease. Overall, the deep learning-based mobile application for monkeypox detection has the capability to assist in early identification, prompt treatment, and efficient disease management, thereby making a valuable contribution to public health initiatives. The mobile application is built using the React Native framework, enabling cross-platform compatibility for both iOS and Android. This approach saves development time and resources since separate codebases for each platform are not required. Fig. 8 displays screenshots of the mobile application developed for monkeypox disease detection.

VI. Discussion

The proposed model outperformed previous studies in terms of performance metrics. The experiments conducted confirmed that the trained model achieves successful detection of a Monkeypox image within an impressive 44-second and 31-second timeframe for MSLD and MSID datasets, respectively. This remarkable result indicates that the system could be utilized as a real-time application, enabling timely and efficient detection of Monkeypox cases. The outstanding performance and real-time capabilities of our proposed system have promising implications, offering significant advancements in the field of Monkeypox detection. Furthermore, its success contributes to the overall progress and development of AI-based diagnostic tools in the healthcare sector. Despite achieving promising results, our work has certain limitations that should be acknowledged. The primary constraint of this research is associated with the dataset used. The current dataset lacks clinical approval and solely comprises skin images. To develop more robust models suitable for real-world applications, it is crucial to collect additional features, such as laboratory test data, to enhance the model's reliability and practicality.

VII. Conclusion

This study's findings hold both theoretical and practical implications for the field of disease identification, particularly in the context of emerging infectious diseases like monkeypox. The development and evaluation of MpoxNet, alongside the introduction of the Swin Transformer model, contribute significantly to the advancement of automated monkeypox detection systems. The theoretical contribution lies in the exploration of deep learning techniques for improving the accuracy and efficiency of disease identification, thereby enhancing early detection capabilities and potentially curbing the spread of infectious diseases.

From a practical standpoint, the integration of MpoxNet with a mobile application offers several advantages for real-world implementation. The rapid and precise detection capabilities of MpoxNet empower healthcare professionals to swiftly diagnose monkeypox, enabling timely intervention and containment measures. Moreover, the accessibility of the mobile application facilitates broader deployment in regions with limited healthcare infrastructure, ultimately aiding in the global fight against the spread of monkeypox.

Despite the promising findings, this study is not without limitations. Firstly, the evaluation is limited to two specific datasets, potentially constraining the generalizability of the findings to broader population samples or different geographic regions. Additionally, the reliance on deep learning models may pose challenges related to model interpretability and transparency, which are crucial for gaining trust and acceptance among healthcare practitioners and policymakers.

For future research, it is imperative to expand the evaluation of MpoxNet and the Swin Transformer model on more diverse datasets representing a broader range of demographic and geographic characteristics. Additionally, exploring methods to enhance the interpretability of deep learning models for disease identification could improve trust and adoption in clinical settings. Furthermore, investigating the integration of real-time data streams and advanced analytics techniques could enhance the predictive capabilities of automated disease detection systems, paving the way for proactive and preventive healthcare interventions.

Appendix

A. 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.

B. Data Availability

The data will be made available on request.

C. Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors have no information to acknowledge.

References

[1] The World Health Network Declares Monkeypox a Pandemic – Press —June 22, 2022. Accessed: [Online]. Available: https://www.worldhealthnetwork.global/monkeypoxpressrelease.

[2] WHO, "Monkeypox Fact Sheet," 2024. Accessed: [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/monkeypox.

[3] S. N. Ali et al., "Monkeypox skin lesion detection using deep learning models: A preliminary feasibility study," arXiv preprint arXiv:2207.03342, 2022.

[4] E. Alakunle, U. Moens, G. Nchinda, and M. I. Okeke, "Monkeypox virus in Nigeria: infection biology, epidemiology, and evolution," Viruses, vol. 12, no. 11, p. 1257, 2020.

[5] M. Moore and F. Zahra, "Monkeypox," in StatPearls [Internet], StatPearls Publishing, 2021.

[6] L. D. Nolen, L. Osadebe, J. Katomba, J. Likofata, D. Mukadi, B. Monroe, and M. G. Reynolds, "Extended human-to-human transmission during a monkeypox outbreak in the Democratic Republic of the Congo," Emerging Infectious Diseases, vol. 22, no. 6, pp. 1014, 2016.

[7] P. Y. Nguyen, W. S. Ajisegiri, V. Costantino, A. A. Chughtai, and C. R. MacIntyre, "Reemergence of human monkeypox and declining population immunity in the context of urbanization, Nigeria, 2017–2020," Emerging Infectious Diseases, vol. 27, no. 4, pp. 1007, 2021.

[8] Q. Gong, C. Wang, X. Chuai, and S. Chiu, "Monkeypox virus: a re-emergent threat to humans," Virologica Sinica, 2022.

[9] T. Islam, M. A. Hussain, F. U. H. Chowdhury, and B. R. Islam, "Can artificial intelligence detect monkeypox from digital skin images?" bioRxiv, 2022.

[10] J. Sun, L. Peng, T. Li, D. Adila, Z. Zaiman, G. B. Melton-Meaux, and C. J. Tignanelli, "Performance of a chest radiograph AI diagnostic tool for COVID-19: A prospective observational study," Radiology: Artificial Intelligence, vol. 4, no. 4, pp. e210217, 2022.

[11] A. Akbarimajd, N. Hoertel, M. A. Hussain, A. A. Neshat, M. Marhamati, M. Bakhtoor, and M. Momeny, "Learning-to-augment incorporated noise-robust deep CNN for detection of COVID-19 in noisy X-ray images," Journal of Computational Science, vol. 63, p. 101763, 2022.

[12] M. Momeny, A. A. Neshat, M. A. Hussain, S. Kia, M. Marhamati, A. Jahanbakhshi, and G. Hamarneh, "Learning-to-augment strategy using noisy and denoised data: Improving generalizability of deep CNN for the detection of COVID-19 in X-ray images," Computers in Biology and Medicine, vol. 136, p. 104704, 2021.

[13] J. Liu, N. Dey, N. Das, R. González Crespo, F. Shi, and C. Liu, "Brain fMRI segmentation under emotion stimuli incorporating attention-based deep convolutional neural networks," Applied Soft Computing, vol. 122, p. 108837, 2022.

[14] S. Kadry, E. Herrera-Viedma, R. González Crespo, S. Krishnamoorthy, and V. Rajinikanth, "Automatic detection of lung nodule in CT scan slices using CNN segmentation schemes: A study," Procedia Computer Science, vol. 218, pp. 2786-2794, 2023.

[15] N. Gessert, T. Sentker, F. Madesta, R. Schmitz, H. Kniep, I. Baltruschat, and A. Schlaefer, "Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting," IEEE Transactions on Biomedical Engineering, vol. 67, no. 2, pp. 495-503, 2019.

[16] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 2016, pp. 1397-1400.

[17] E. H. Mohamed and W. H. El-Behaidy, "Enhanced skin lesions classification using deep convolutional networks," in 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 2019, pp. 180-188.

[18] M. A. Al-Masni, D. H. Kim, and T. S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," Computer Methods and Programs in Biomedicine, vol. 190, p. 105351, 2020.

[19] S. Jinnai, N. Yamazaki, Y. Hirano, Y. Sugawara, Y. Ohe, and R. Hamamoto, "The development of a skin cancer classification system for pigmented skin lesions using deep learning," Biomolecules, vol. 10, no. 8, p. 1123, 2020.

[20] W. Sae-Lim, W. Wettayaprasit, and P. Aiyarak, "Convolutional neural networks using MobileNet for skin lesion classification," in 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), Chonburi, Thailand, 2019, pp. 242-247.

[21] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, "Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet," in International Conference on Advanced Machine Learning Technologies and Applications, Springer, Singapore, 2020, pp. 165-176.

[22] Harangi, B., Baran, A., & Hajdu, A. (2018, July). Classification of skin

lesions using an ensemble of deep neural networks. In 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 2575-2578). IEEE.

[23] J. Steppan and S. Hanke, "Analysis of skin lesion images with deep learning," arXiv preprint arXiv:2101.03814, 2021.

[24] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-sources dermatoscopic images of common pigmented skin lesions," Scientific Data, vol. 5, no. 1, pp. 1-9, 2018.

[25] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 2018, pp. 168-172.

[26] M. Combalia, N. C. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, and J. Malvehy, "Bcn20000: Dermoscopic lesions in the wild," arXiv preprint arXiv:1908.02288, 2019.

[27] T. Hu, M. Khishe, M. Mohammadi, G.-R. Parvizi, S. H. Taher Karim, and T. A. Rashid, "Real-time COVID-19 diagnosis from X-Ray images using deep CNN and extreme learning machines stabilized by chimp optimization algorithm," Biomedical Signal Processing and Control, vol. 68, Article 102764, 2021.

[28] H. Sharma, J. S. Jain, P. Bansal, and S. Gupta, "Feature extraction and classification of chest X-Ray images using CNN to detect pneumonia," in 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 227–231.

[29] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, "Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms," International Journal of Medical Informatics, vol. 144, p. 104284, 2020.

[30] M. V. Madhavan, A. Khamparia, D. Gupta, S. Pande, P. Tiwari, and M. S. Hossain, "Res-CovNet: An internet of medical health things driven COVID-19 framework using transfer learning," Neural Computing and Applications, pp. 1–14, 2021.

[31] R. Sandeep, K. Vishal, M. Shamanth, and K. Chethan, "Diagnosis of visible diseases using CNNs," in Proceedings of International Conference on Communication and Artificial Intelligence, Springer, 2022, pp. 459–468.

[32] K. Glock, C. Napier, T. Gary, V. Gupta, J. Gigante, W. Schaffner, and Q. Wang, "Measles rash identification using transfer learning and deep convolutional neural networks," in 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 3905–3910.

[33] M. M. Ahsan, M. R. Uddin, M. Farjana, A. N. Sakib, K. A. Momin, and S. A. Luna, "Image data collection and implementation of deep learning-based model in detecting monkeypox disease using modified VGG16," arXiv preprint arXiv:2206.01862, 2022.

[34] C. Sitaula and T. B. Shahi, "Monkeypox virus detection using pre-trained deep learning-based approaches," Journal of Medical Systems, vol. 46, no. 11, pp. 1-9, 2022.

[35] V. Kumar, "Analysis of CNN features with multiple machine learning classifiers in diagnosis of monkeypox from digital skin images," medRxiv, 2022. [CrossRef].

[36] S. Maqsood and R. Damaševičius, "Monkeypox detection and classification using deep learning-based features selection and fusion approach," in 2023 IEEE International Systems Conference (SysCon), Vancouver, BC, Canada, 2023, pp. 1-8.

[37] M. Hussain, M. A. Khan, R. Damaševičius, A. Alasiry, M. Marzougui, M. Alhaisoni, and A. Masood, "SkinNet-INIO: multiclass skin lesion localization and classification using fusion-assisted deep neural networks and improved nature-inspired optimization algorithm," Diagnostics, vol. 13, no. 18, p. 2869, 2023.

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

[39] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, ... and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[40] D. Bala and M. S. Hossain, "Monkeypox Skin Images Dataset (MSID)," Mendeley Data, V6, 2023. doi: 10.17632/r9bfpnvyxr.6.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 770-778.

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[43] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510-4520.

[44] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261-2269.

[45] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251-1258.

[46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818-2826.

[47] P. Pandiyan, R. Thangaraj, M. Subramanian, R. Rahul, M. Nishanth, and I. Palanisamy, "Real-time monitoring of social distancing with person marking and tracking system using YOLO V3 model," International Journal of Sensor Networks, vol. 38, no. 3, pp. 154-165, 2022.

Dr. S. Sadesh

Dr. S. Sadesh is presently working as a Professor & Head in the Department of Artificial Intelligence and Data Science at Velalar College of Engineering and Technology. His areas of interest are Web Mining, Data Mining and Data Analytics. He has published 23 papers in International Journals.

Dr. Rajasekaran Thangaraj

Rajasekaran Thangaraj received his Ph.D in the area of Deep Learning from Anna University in the year 2022. He has published 25 articles in reputed International Journals and presented 18 papers at various International Conferences. He published 4 book chapters and 5 Indian patents.

Dr. Pandiyan P

Pandiyan P received his PhD Degree in Instrumentation and Control Engineering from the National Institute of Technology. His research interests include design and simulation of MEMS based logic devices, Energy harvesting and Artificial Intelligence.

Dr. R. Devi Priya

Dr. R. Devi Priya is working as professor in the department of Computer Science and Engineering in KPR Institute of Engineering and Technology and associated with Research Centre for IoT and Artificial Intelligence. She has published more than 60 papers in reputed journals and conferences.

Dr. Palanichamy Naveen

Naveen P. completed his Ph.D. in the field of Image Processing and Machine Learning from Kalasalingam Academy of Research and Education. He has about 13+ years of teaching experience at various levels and is currently holding the responsibility of Coordinator - Sponsored Research in the Centre for Research and Development, as well as the position of Assistant Professor in the Department of Electrical and Electronics Engineering at KPR Institute of Engineering and Technology, Coimbatore, India. His research interests include Image Processing and Machine Learning. He has published around 25 papers in International Journals and International Conferences. He also published 4 patents.