

## Deep learning based detection of monkeypox virus using skin lesion images

Tushar Nayak<sup>a</sup>, Krishnaraj Chadaga<sup>b</sup>, Niranjana Sampathila<sup>a,\*</sup>, Hilda Mayrose<sup>a</sup>,  
 Nitila Gokulkrishnan<sup>a</sup>, Muralidhar Bairy G<sup>a</sup>, Srikanth Prabhu<sup>b</sup>, Swathi K. S<sup>c</sup>,  
 Shashikiran Umakanth<sup>d</sup>

<sup>a</sup> Department of Biomedical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India

<sup>b</sup> Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India

<sup>c</sup> Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India

<sup>d</sup> Department of Medicine, Dr. T.M.A. Pai Hospital, Manipal Academy of Higher Education, Udupi, Karnataka, 576101, India

### ARTICLE INFO

#### Keywords:

Deep learning  
 Disease diagnosis  
 Image processing  
 Monkeypox virus  
 Machine learning  
 Transfer learning

### ABSTRACT

As we set into the second half of 2022, the world is still recovering from the two-year COVID-19 pandemic. However, over the past three months, the outbreak of the Monkeypox Virus (MPV) has led to fifty-two thousand confirmed cases and over one hundred deaths. This caused the World Health Organisation to declare the outbreak a Public Health Emergency of International Concern (PHEIC). If this outbreak worsens, we could be looking at the Monkeypox virus causing the next global pandemic. As Monkeypox affects the human skin, the symptoms can be captured with regular imaging. Large samples of these images can be used as a training dataset for machine learning-based detection tools. Using a regular camera to capture the skin image of the infected person and running it against computer vision models is beneficial. In this research, we use deep learning to diagnose monkeypox from skin lesion images. Using a publicly available dataset, we tested the dataset on five pre-trained deep neural networks: GoogLeNet, Places365-GoogLeNet, SqueezeNet, AlexNet and ResNet-18. Hyperparameter was done to choose the best parameters. Performance metrics such as accuracy, precision, recall, f1-score and AUC were considered. Among the above models, ResNet18 was able to obtain the highest accuracy of 99.49%. The modified models obtained validation accuracies above 95%. The results prove that deep learning models such as the proposed model based on ResNet-18 can be deployed and can be crucial in battling the monkeypox virus. Since the used networks are optimized for efficiency, they can be used on performance limited devices such as smartphones with cameras. The addition of explainable artificial intelligence techniques LIME and GradCAM enables visual interpretation of the prediction made, helping health professionals using the model.

### 1. Introduction

The first half of the year 2022 saw the gradual decline of the severity of the COVID-19 pandemic, after its third wave that began in January 2022. However, just a few weeks later a new threat emerged which quickly grew into a global outbreak and could soon become a pandemic. The Human Monkeypox disease is not a new and novel disease [1]. The first infection identified as early as 1970, with cases increasing over the following decade. This is also not the first Human Monkeypox outbreak, as there was the 2003 Midwest Monkeypox Outbreak and the 2017–2019 Nigeria Monkeypox Outbreak [2]. There have also been few cases of the disease infection cropping up in isolated cases in the United Kingdom, Singapore, and other parts of the United States of America [3]. However, the current 2022 Monkeypox outbreak has spread to over a hundred

countries and territories over the past nine months [4]. Due to the mode of transmission of the virus, it is comparatively less contagious [5]. However, the need for a low cost and rapid detection system is of paramount importance since it is still spreading.

The virus is a member of the Poxviridae family [6]. Mammalian species such as squirrels, rats and other primates have been identified as natural hosts of the virus. The disease that the virus causes is an infectious disease that lasts between two to four weeks, with the usual onset about five to twenty-one days post exposure. Currently known symptoms include fever, muscle and headache, shivering, swollen lymph nodes and blistering rashes [7]. The onset of these rashes is within three days and appears on the face, palms, and foot sole of the patient. It can also spread to mouth, eyes and genitals. This phase is followed by the skin eruption phase where the lesions worsen in four stages. It begins with lesions

\* Corresponding author.

E-mail address: [niranjana.s@manipal.edu](mailto:niranjana.s@manipal.edu) (N. Sampathila).

<https://doi.org/10.1016/j.medntd.2023.100243>

Received 7 March 2023; Received in revised form 25 May 2023; Accepted 26 May 2023

2590-0935/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

having flat bases (macules) and progresses to become raised and firm lesions (papules). These papules then get filled with puss (pustules) which then forms a solid crust [8].

The current method for the detection of Human Monkeypox is by using a Polymerase Chain Reaction (PCR) test. However, the results obtained by this test is not accurate as they can be inconclusive due to the virus remaining in the blood for a short time [9]. It also requires additional information such as current stage of the rashes, patient age and the dates of onset of fever and rashes. PCR tests also require a considerable amount, hampering its availability in rural or remote areas. A system that is independent of these metrics and can use real time data for near perfect diagnosis using readily available devices could be the solution to creating

an effective and efficient diagnosis modality for Monkeypox. The usage of Artificial Intelligence and its sub-units is not new to the world of healthcare [10]. Deep neural networks for computer vision applications in the domain of healthcare can leverage the wide availability of healthcare data to train Convolutional Neural Networks that can use pre-existing devices to solve emerging problems [11]. This is the approach taken here to develop a model for the detection of Monkeypox using RGB images of skin lesions taken from regular cameras on smartphones. The sample images of this dreaded disease are described in Fig. 1.

A few researches have been conducted which use deep learning to diagnose Monkeypox disease. Sitaula et al. [12] used deep learning to

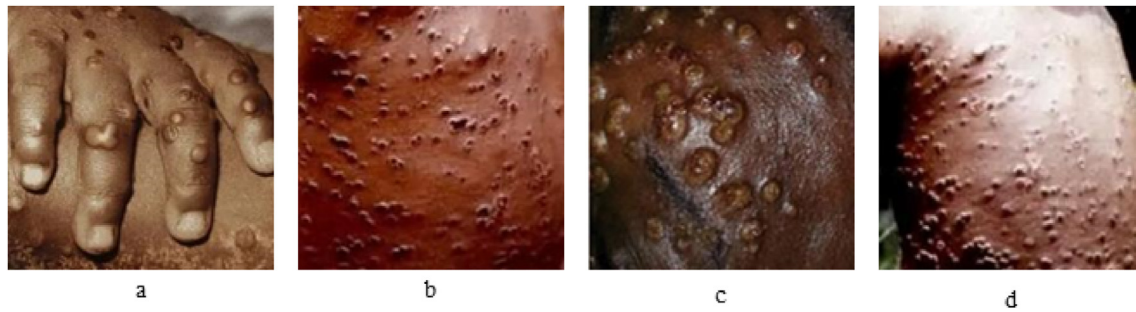


Fig. 1. Examples of skin lesions of Monkeypox patients.



Fig. 2. Illustrative example of the dataset. Images a-g:Class 'MonkeyPox. Images h-n: Class "Others".

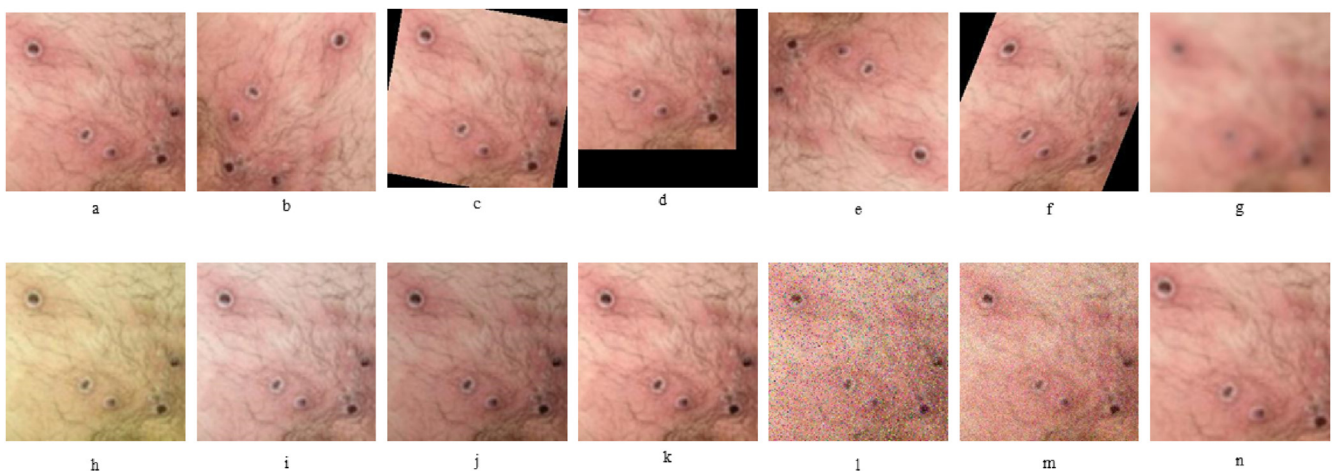


Fig. 3. Illustrative example of the fourteen-fold augmentation on an image from the 'Monkeypox' class. Image a: original image. Images b-n: Augmented images.

diagnose monkeypox virus. Thirteen different models were used to train and test the data. They were further ensembled for optimization. An average accuracy of 87.13% was obtained by the models. In another research, deep pre-trained network was used to classify monkeypox using skin lesion images [13]. An android application was developed to facilitate this research. A maximum accuracy of 91.11% was obtained. Alakus et al. [14] used Wart DNA sequences and deep learning models to distinguish monkeypox from warts. Three stages were used in the classification process and maximum accuracy of 96.08% was obtained. "AI-BIRUS Earth Radius Optimization Algorithm" was used to classify images in Ref. [15]. Ten different evaluation metrics were used in the study. An accuracy of 98.8% was obtained by the algorithm. The monkeypox can cause the next pandemic and it is critical that the resources should be utilized efficiently. AI can help in many ways including disease

diagnosis. In this study, we utilize a variety of transfer learning models to classify monkeypox images.

The objectives of this research are as follows.

- The utilization of deep learning models such as GoogLeNet, Places365-GoogleNet, SqueezeNet and AlexNet and ResNet-18 to detect monkeypox virus with high accuracy.
- Comparison of the above models regarding performance metrics such as accuracy, precision, recall and f1-score. We also compare our research with other existing researches.
- Further discussion about how the models can be utilized in real time or accurate diagnosis.

The approach taken here is to leverage the performance of modern

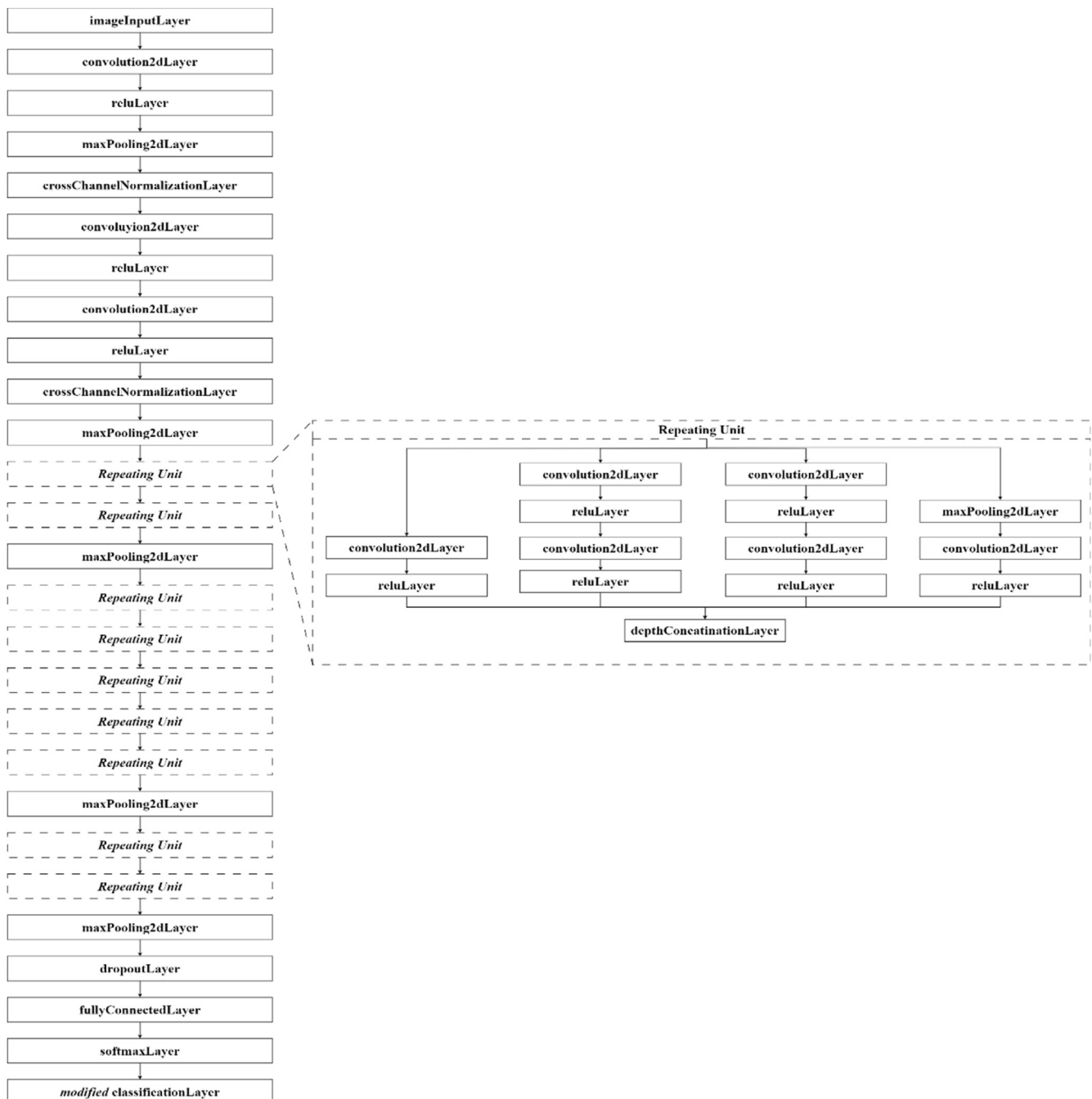


Fig. 4. (a) Architecture of GoogLeNet ImageNet & Places365-Standard. (b) Repeating unit.

smartphones and the capabilities of their cameras to develop a fully self-contained diagnostic modality. The usage of modified pre-trained networks aided by providing a solid foundation for further hyper-parameter tuning during the training of these networks are also explained in depth. The networks selected have also been built around efficiency, ensuring reliable and efficient performance on relatively low-performance devices such as smartphones. Validation accuracies above 95% have been obtained by most models.

## 2. Methodology

The methodology section is divided into five sections: data collection, data augmentation, transfer learning approaches and the experimental setup used.

### 2.1. Data collection

The dataset which is used to train the transfer learning models is titled “Monkeypox Skin Lesion Dataset” [16]. It is a binary classification

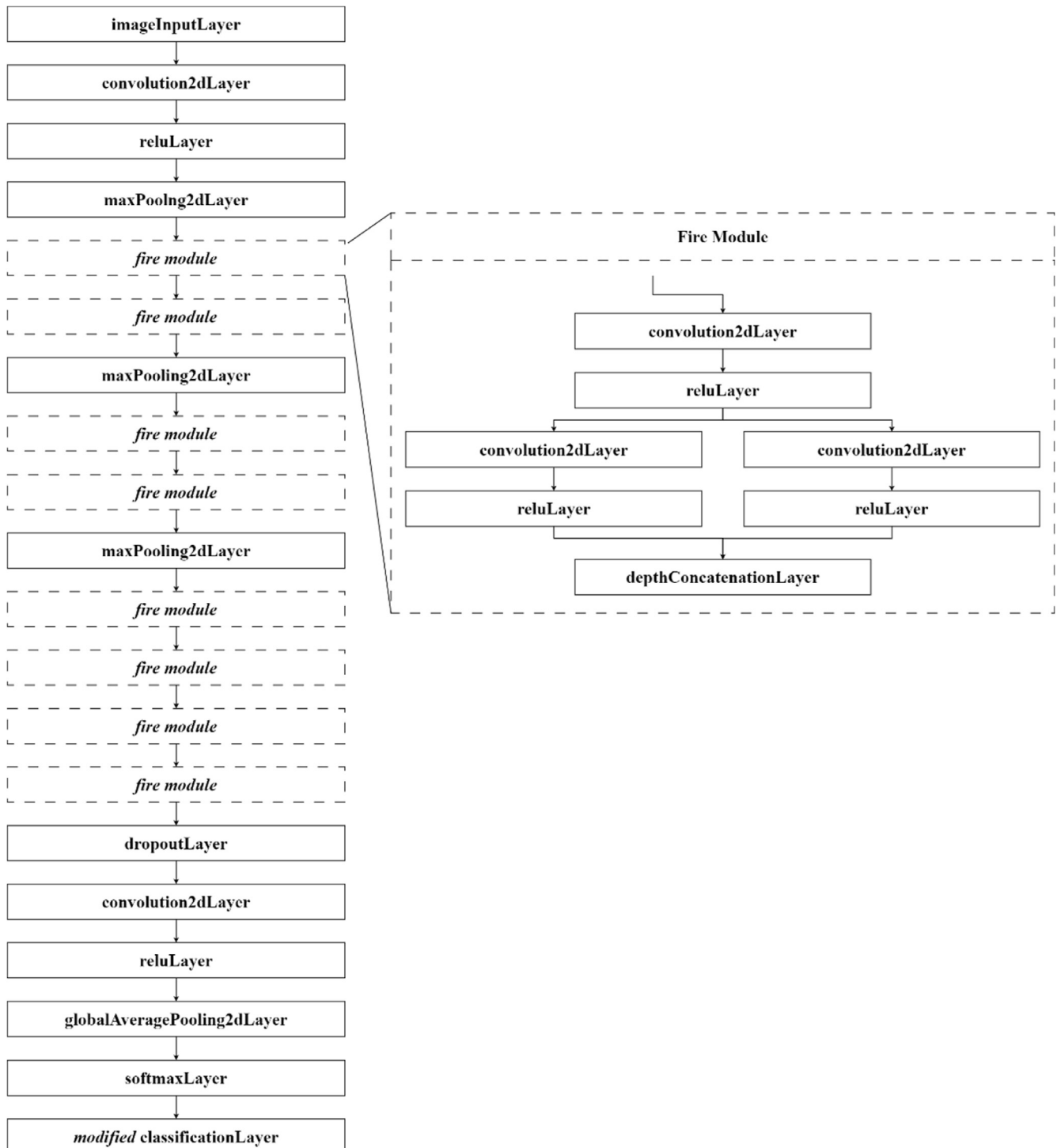


Fig. 5. Architecture of SqueezeNet (a) Basic structure of SqueezeNet (b) Repeating unit.



dataset for Monkeypox v/s non-Monkeypox images and is available on Kaggle (an online community of data scientists and machine learning practitioners). The dataset is a collection of 228 images which comprises of 102 Monkeypox & 126 other (chickenpox & measles) cases. The images have a resolution of  $224 \times 224 \times 3$ . Fig. 2 describes some of the images of both classes.

## 2.2. Data augmentation

The dataset obtained had its original dataset augmented by the creators themselves. Augmentation of datasets in training deep neural networks is used to increase the size of the dataset to improve the performance of the network [17]. They used augmentation methods such as rotation, translation, reflection, shear, hue, saturation, contrast and brightness jitter, noise and scaling. In total, fourteen-fold augmentation was performed on the original dataset, bringing the number of images in each class to 1428 in 'Monkeypox' class & 1764 in 'Others' class. Some of the augmented images are described in Fig. 3.

## 2.3. Deep learning approaches

For the evaluation of the model on the given dataset, a pilot run was run with the augmented dataset on multiple pre-trained deep neural networks in MathWorks MATLAB's Deep Network Designer. The selected networks were GoogLeNet, Places365-GoogLeNet, SqueezeNet and AlexNet and ResNet-18. This selection was mainly made due to the performance efficiency of the networks, with the first three networks being built specifically with efficiency in mind. Due to the pre-established performance reliability of AlexNet and ResNet-18 in past researches involving skin lesion classification, they were included too. Due to the availability of these networks on MATLAB 2022b as add-on packages, all the trials on all the networks were run on the Experiment Manager app within MATLAB 2022b with all relevant pre-trained networks installed and loaded.

### 2.3.1. GoogLeNet & GoogLeNet trained on Places-365 dataset

GoogLeNet is a deep convolutional neural network developed by researchers at Google Inc as a variant of Inception Network [18]. GoogLeNet architecture has twenty-seven layers (when pooling layers are included) which have nine inception modules. There are about one hundred layers (building blocks) used in the architecture. The architectural details of the auxiliary classifiers are as follows: Average pooling later of filter size  $5 \times 5$  and stride 3,  $1 \times 1$  convolution with 128 filters for dimension reduction and ReLU activation, fully connected layer with 1025 outputs and ReLU activation, Dropout Regularization with dropout ratio = 0.7, Softmax classifier with 1000 classes output like the main softmax classifier. The GoogLeNet architecture is designed to be computationally efficient and work on devices with low system memory.

The main difference between GoogLeNet and Places365-GoogLeNet is the dataset used for pre-training. GoogLeNet used the data from the ImageNet data from the ILSVRC 2014 Classification Challenge. Places365-GoogLeNet is the same network, pretrained on the *Places365-Standard dataset* [19]. In both, GoogleNet & Places365-GoogleNet, there are two modifications made in the fully connected layer and classification layer. These layers have been trained with datasets that default it to 1000 classes on both layers. Since the given dataset used here, output size = 2 on the fully connected layer which when trained, assigns two classes to the classification layer. The architecture if GoogLeNet is given in Fig. 4.

### 2.3.2. SqueezeNet

SqueezeNet is a deep neural network for computer vision developed by researchers at DeepScale, University of California, Berkeley, and Stanford University with the intention of creating a smaller neural network with fewer parameters to be more computationally efficient [20]. The architecture of SqueezeNet starts with a convolution layer (conv1), ending with final convolution layer (conv10). It has Eight Fire

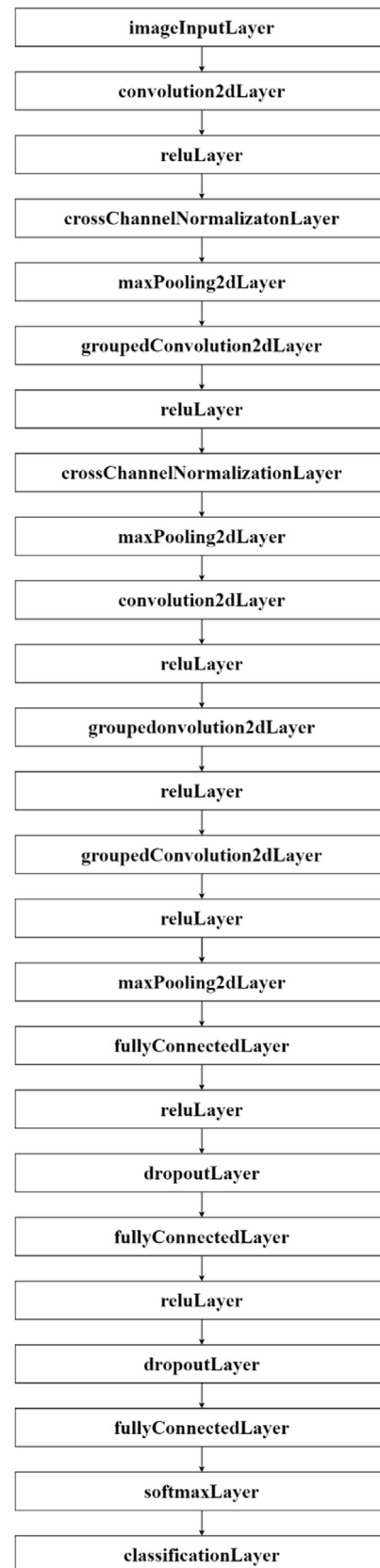


Fig. 6. Architecture of AlexNet.

modules. It also performs max-pooling with stride of 2 after conv1, fire 4, fire 8 and conv10. Dropout with a ratio of 50% is applied after fire 9

module. To adjust to dataset, final convolution layer conv10 has modified NumFilters = 2 since the dataset has two classes. No additional layers are added. The architecture of SqueezeNet is given in Fig. 5.

2.3.3. AlexNet

AlexNet is a convolutional neural network developed by Alex

Krizhevsky, Ilya Sutskever and Geoffrey Hinton and won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012. The architecture of AlexNet has 25 layers, out of which are five convolutional layers. The first, second and fifth convolutional layers also have Max-Pooling Layers. These Max-Pooling layers are overlapped, having strides of 2 with filter size  $3 \times 3$ . Dropout layer in the network decreases overfitting

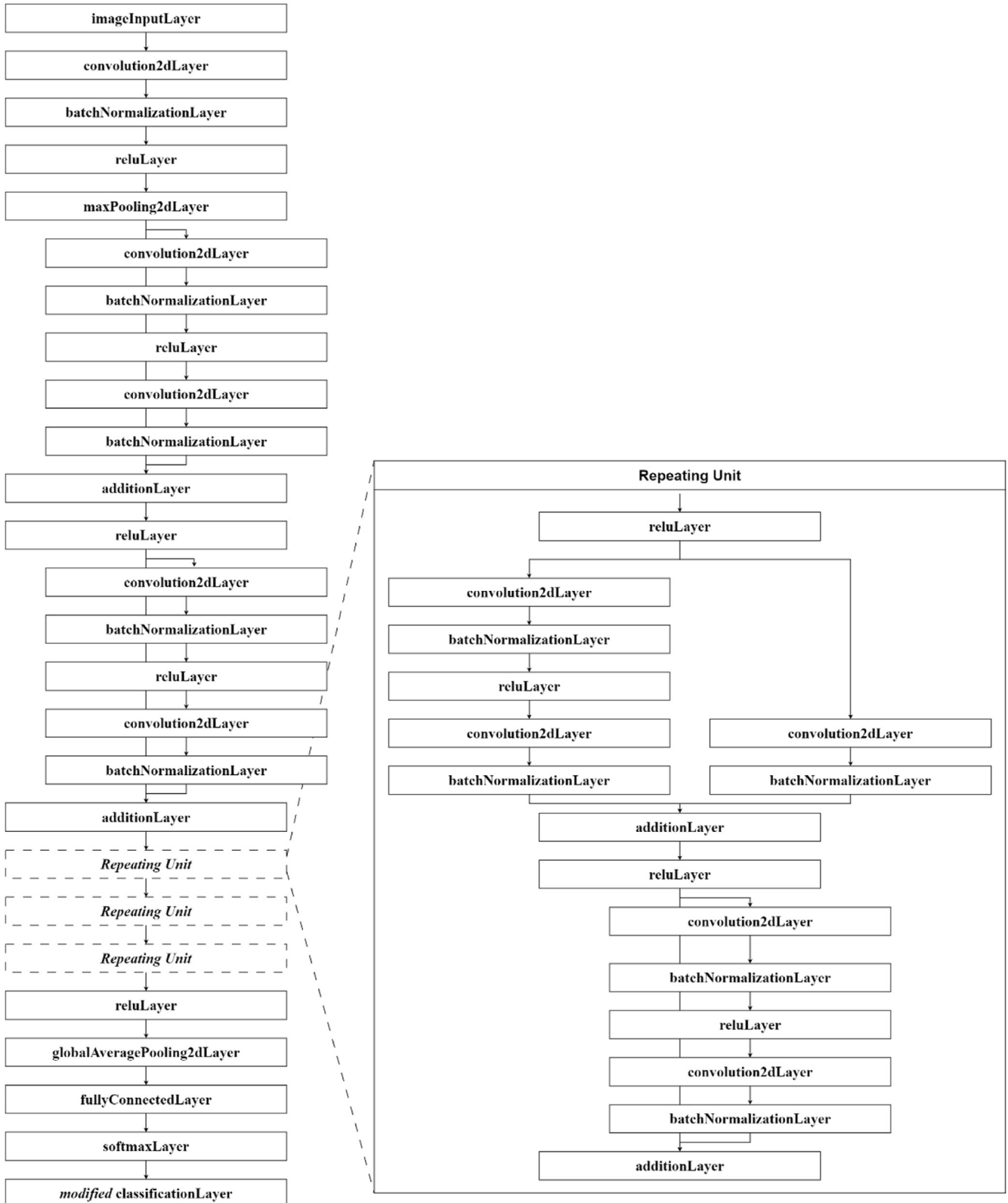


Fig. 7. (a) Architecture of ResNet-18 (b) Repeating unit.

**Table 1**  
Hyperparameters used across all trials.

Parameter	Values
Mini Batch Size	[16 32 64]
Learning Rate	[0.01 0.001]

**Table 2**  
Distribution of Training & Validation data.

Dataset	Training (80%)		Validation (20%)	
	Class 'Monkeypox'	Class 'Others'	Class 'Monkeypox'	Class 'Others'
Augmented Dataset	1411	1142	353	286
Original Dataset	82	101	20	25

of the network and it is recommended to use augmented dataset for training [21]. The architecture of AlexNet is given in Fig. 6.

**2.3.4. ResNet-18**

ResNet is a reliable and performant deep neural network that won ILSVRC 2014 Classification Challenge and has given good results on the ImageNet dataset. ResNet has many types, depending on the number of layers in the network. ResNet-18 used here has eighteen layers [22]. Using shortcut connections in the architecture, ResNets can solve the

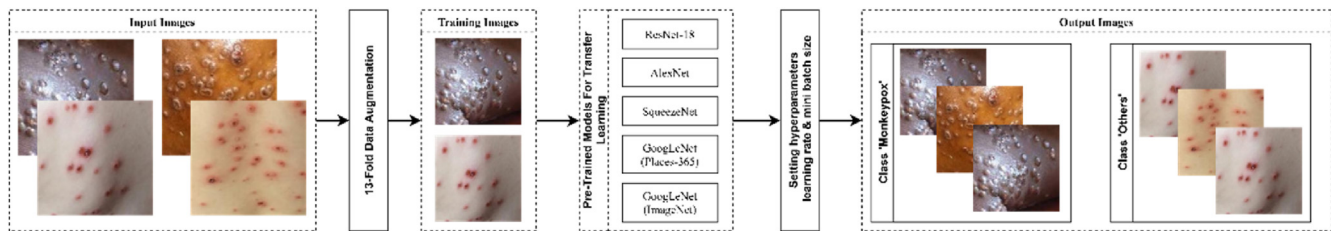
vanishing problem. Residual Networks have also been widely used in healthcare applications. The architecture of ResNet is described in Fig. 7.

**2.3.5. Vision transformer**

Vision Transformer (ViT) are transformer-based architecture that achieve robust performance in image classification tasks. Self-attention mechanism is used in the ViT that allows it to selectively focus on different parts of the image. The Vision Transformer extracts global features by dividing the image into a sequence of patches and feeding them through a transformer-encoder. This is in start contrast to Convolutional Neural Networks such as GoogLeNet, ResNet-18, AlexNet & SqueezeNet that all use fixed size filters to extract local features, not global features. Described in our study is ViT-B16. It is a ViT model that has a transformer encoder trained on the ImageNet-21 K & ImageNet-1K with 16 layers and a patch size of 16 × 16 pixels [23].

**2.4. Experimental setup**

The experiment was conducted on a laptop with an Intel Core i7 10750H and 16 GB memory. The execution environment for the training and validation of the networks was set to the single GPU: Nvidia GeForce GTX 1650 with 4 GB VRAM. The Experiment Manager on Mathworks MATLAB 2022b was used to perform the experiment. The experiment was run four times, once for each network with varying hyperparameters: Mini Batch Size and Learning Rate. The details of the hyperparameters used for tuning are detailed in Table 1. Given the variation of the Learning Rate & Mini Batch Size, the Epoch Size for all trials was



**Fig. 8.** Flow diagram used in this research to classify monkeypox.

**Table 3**  
Training and Validation Results for all the models.

Trial	Learning Rate	Mini Batch Size	Training Accuracy (%)	Training Loss	Validation Accuracy (%)	Validation Loss
<b>GoogLeNet</b>						
1	0.001	32	100	7.8E-06	97.8697	0.0744
2	0.001	16	100	0.00014	97.4937	0.1144
3	0.001	64	100	3.5E-05	97.11779	0.1272
4	0.0001	16	100	0.0002	96.49123	0.2138
<b>GoogLeNet (Places-365)</b>						
1	0.001	64	100	3E-05	97.61905	0.0948
2	0.001	32	100	8.3E-07	97.36842	0.0940
3	0.001	16	100	4.7E-06	97.36842	0.1280
4	0.0001	32	100	3.1E-05	96.11529	0.2099
<b>SqueezeNet</b>						
1	0.001	16	100	7.0777E-06	98.08153	0.1813
2	0.001	32	100	0.0002	96.36591	0.1505
3	0.001	64	100	1.126E-05	95.3634	0.2747
4	0.001	128	100	0.000012038	94.61153	0.245
<b>AlexNet</b>						
1	0.0001	16	100	5.6624E-07	97.61905	0.1058
2	0.0001	32	100	0.000020707	96.2406	0.1995
3	0.001	64	100	0.000011415	96.11529	0.198
4	0.0001	128	100	0.001	96.61404	0.1862
<b>ResNet18</b>						
1	0.001	32	100	1.33E-06	99.4987	0.0375
2	0.001	16	100	1.71E-07	99.4987	0.0235
3	0.001	64	100	1.51E-05	97.7444	0.0784
4	0.0001	32	100	0.0001	97.619	0.0662

**Table 4**  
Values from confusion matrices for all the deep learning models.

Trial	True Positive	True Negative	False Positive	False Negative
<b>GoogLeNet</b>				
1	345	436	5	12
2	347	431	10	10
3	344	431	10	13
4	340	430	11	17
<b>GoogLeNet (Places-365)</b>				
1	345	434	7	12
2	345	432	9	12
3	342	435	6	15
4	343	424	17	14
<b>SqueezeNet</b>				
1	384	434	7	9
2	336	433	8	21
3	340	421	20	17
4	333	422	19	24
<b>AlexNet</b>				
1	344	435	6	13
2	332	436	5	25
3	337	430	11	20
4	328	435	6	29
<b>ResNet-18</b>				
1	353	441	0	4
2	355	439	2	2
3	346	434	7	11
4	345	434	7	12

standardized to 200 Epochs, with validation frequency set to 5 Epochs.

Table 2 contains the dataset that was used during the training and validation of the three deep networks. The networks were trained and validated only on the augmented data set, with 80% of the combined images from both labels used for training and 20% used for validation. During validation, the data set was randomized for more accurate validation. Fig. 8 describes the entire process used in this research to classify monkeypox skin lesion images using deep learning.

### 3. Results & discussion

The results are explained in depth in this section. The first section

**Table 5**  
Calculated evaluation parameters for all deep learning models.

Trial	Sensitivity	Specificity	Precision	Accuracy	F1-Score
<b>GoogLeNet</b>					
1	0.966386555	0.988662132	0.985714286	0.978696742	0.975954738
2	0.971988796	0.977324263	0.971988796	0.974937343	0.971988796
3	0.963585434	0.977324263	0.971751412	0.971177945	0.967651195
4	0.952380952	0.975056689	0.968660969	0.964912281	0.960451977
<b>GoogLeNet (Places-365)</b>					
1	0.966386555	0.984126984	0.980113636	0.976190476	0.973201693
2	0.966386555	0.979591837	0.974576271	0.973684211	0.970464135
3	0.957983193	0.986394558	0.982758621	0.973684211	0.970212766
4	0.960784314	0.961451247	0.952777778	0.961152882	0.956764296
<b>SqueezeNet</b>					
1	0.977099237	0.984126984	0.982097187	0.980815348	0.979591837
2	0.941176	0.981859	0.976744	0.963659	0.958631
3	0.952380952	0.954648526	0.944444444	0.953634085	0.948396095
4	0.932773109	0.9569161	0.946022727	0.946115288	0.939351199
<b>AlexNet</b>					
1	0.963585434	0.986394558	0.982857143	0.976190476	0.973125884
2	0.929971989	0.988662132	0.985163205	0.962406015	0.956772334
3	0.943977591	0.975056689	0.968390805	0.961152882	0.956028369
4	0.918767507	0.986394558	0.982035928	0.956140351	0.94934877
<b>ResNet18</b>					
1	0.988795518	1	1	0.994987469	0.994366197
2	0.994397759	0.995464853	0.994397759	0.994987469	0.994397759
3	0.969187675	0.984126984	0.980169972	0.977443609	0.974647887
4	0.966386555	0.984126984	0.980113636	0.976190476	0.973201693
<b>ViT B18</b>					
1	0.7926	0.6851	0.4977	0.7155	0.6114

describes the performance metrics used in this research. The second section is used to explain the results obtained by the various transfer learning models. In the final sub section, we describe how the models can be used in real time to combat the monkeypox virus.

#### 3.1. Performance metrics

The performance of the three networks has been evaluated using results obtained from the associated confusion matrix for each trial. For evaluation the parameters used are precision, sensitivity, specificity, accuracy and F1-Score [24].

- Confusion matrix:** It is a  $2 \times 2$  matrix during a binary classification. It consists of true positives, true negatives, false positives and false negative values. True positive cases occur when the number of class ‘Monkeypox’ samples identified correctly. True negative cases occur when the non-Monkeypox (‘others’) cases are identified accurately. False positive and false negative results are wrongly predicted results. False positive results occur when non-Monkeypox (‘others’) are identified wrongly. False negative occurs when Monkeypox cases are predicted wrong. The models perform well when false positive and false negative cases are minimized.
- Precision:** Precision is a metric which emphasizes on the true positive and false positive results. The precision is high when the false positive cases are low. It is calculated using equation (1).

$$Precision = \frac{True\ positives}{True\ positives + False\ positives} \tag{1}$$

- Specificity:** Specificity is a metric which focuses on true negative and false positive results. The specificity is also high when the false positive cases are low. It is calculated using equation (2).

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives} \tag{2}$$



- **Sensitivity (Recall):** Sensitivity is a metric which emphasizes on true positive and false negative results. The sensitivity is high when the false negative cases are low. It is calculated using equation (3).

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ negatives} \tag{3}$$

- **F1-score:** F1-score is a metric which considers both precision and recall. It can be calculated using equation (4).

$$F1\ score = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \tag{4}$$

- **Accuracy:** Number of samples predicted accurately among all samples (Both Monkeypox and non Monkeypox). It is calculated using equation (5).

$$Accuracy = \frac{True\ positives + True\ negatives}{True\ positives + True\ negatives + False\ positives + False\ negatives} \tag{5}$$

### 3.2. Model evaluation

In this study, five deep learning models: GoogLeNet, GoogLeNet (Places-365), SqueezeNet, AlexNet and ResNet18 were used to classify monkeypox from other similar diseases such as chicken pox and measles. The fast computer-aided diagnosis using skin lesion images can be highly

beneficial if the virus spreads rapidly. Table 3 shows the validation accuracies and losses obtained by the classifiers. The learning rate consisted of two values (0.001 and 0.0001), which was changed iteratively. Four values were used to define mini batch size (16, 32, 64,128), which was changed iteratively. Further, batch size of 128 was only provided to the SqueezeNet and AlexNet models. All the models obtained a training accuracy of 100%. ResNet18 obtained the highest validation accuracy of 99.49% during two trials.

GoogLeNet obtained the highest accuracy of 97.86% during the first run and the validation loss was 0.07. For the above run, the mini batch size was set to 32 and the learning rate was set to 0.001%. GoogLeNet (Places-365) obtained the highest accuracy of 97.61% during the first trial and the validation loss obtained was 0.09%. The mini batch size of set to 64 and a learning rate of 0.001 was used. The SqueezeNet model obtained a maximum accuracy of 98.08% and a validation loss 0.18 during the first run. The batch size was set to 16 and learning rate of 0.001 was used. A maximum accuracy of 97.61% was obtained by the AlexNet model during its first run. The validation loss obtained was 0.10. The mini batch size was set 16 and a learning rate of 0.0001 was used. ResNet18 obtained a maximum accuracy of 99.50% during the first two runs. The losses obtained were 0.0375 and 0.0235 respectively. The accuracy obtained was equal when the batch size was set to 16 and 32. A learning rate of 0.001 was utilized.

The values obtained by the confusion matrices are described in Table 4. It can be observed that the number of false positive cases obtained by the GoogleNet model was only five. During the third run of GoogleNet (places-365), only six false positive cases were observed.

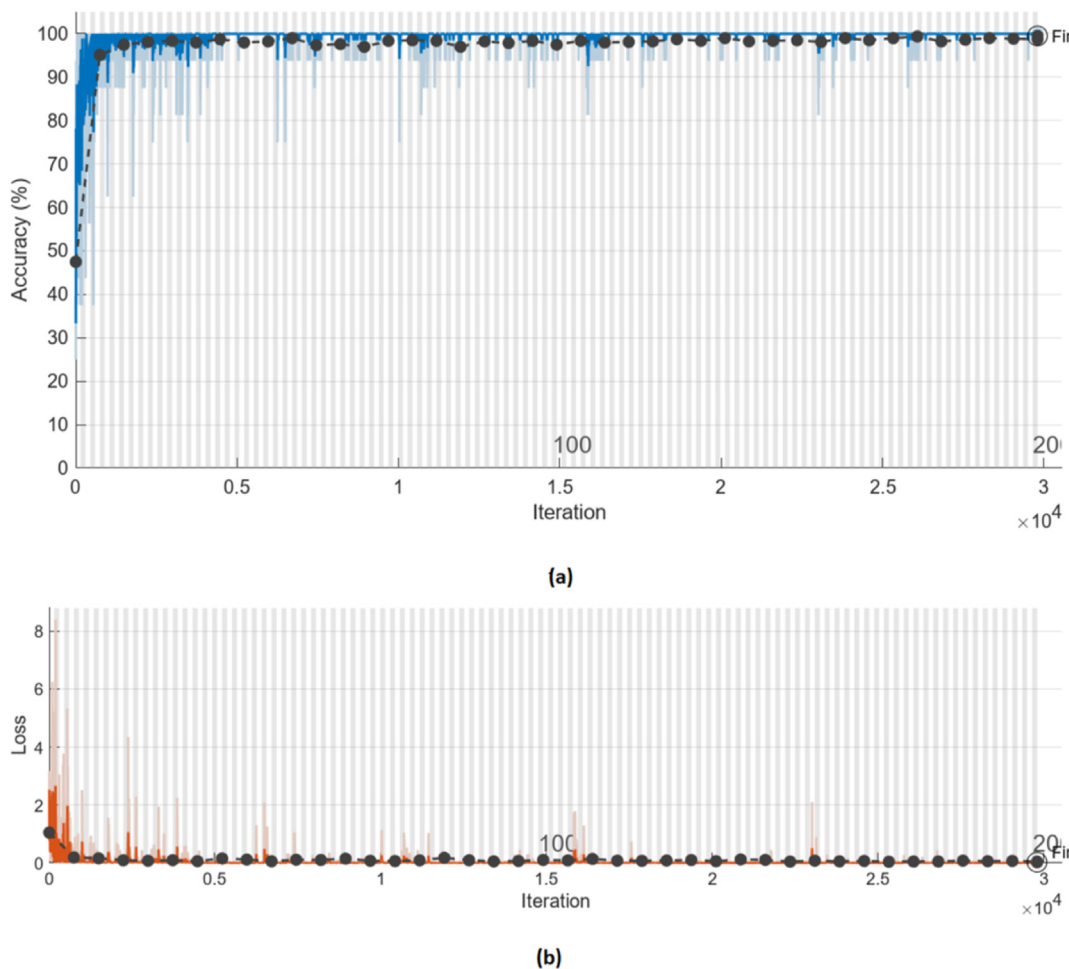


Fig. 9. Accuracy and loss curve obtained for the best performing ResNet18 model (a) Accuracy curve – Training (Blue) & Validation (Black) (b) Loss curve. Training (Red) & Validation (Black)

However, there were 15 false negative cases. During the first run of SqueezeNet, the number of both false positive and false negative cases obtained was less than 10. When compared to other models, the number of false positive cases obtained by the AlexNet was comparatively higher. During the first run of ResNet18, there were no false positive results obtained. The number of false negative results obtained was four. From the table, it can be seen that the false negative cases were comparatively more compared to false positive cases.

Table 5 describes the metrics such as sensitivity, specificity, precision, accuracy and F1-score obtained by the test dataset. Among all the models, the ResNet18 was able to obtain a maximum sensitivity of 99.43%. The maximum sensitivity obtained by the GoogleNet, GoogleNet (places-365), SqueezeNet and AlexNet were 92.19%, 96.63%, 97.70% and 96.35%, respectively. The ResNet18 obtained a specificity of 100% during the first run. The maximum specificity obtained by the GoogleNet, GoogleNet (places-365), SqueezeNet and AlexNet were 98.86%, 98.63%, 98.41% and 98.86%. The ResNet18 was able to obtain a precision of 100% during the first run. The maximum precision obtained by the GoogleNet, GoogleNet (places-365), SqueezeNet and AlexNet were 98.57%, 98.27%, 98.20% and 98.51%, respectively. The highest F1-score was obtained by the ResNet18 model. The maximum F1-score obtained by the GoogleNet, GoogleNet (places-365), SqueezeNet, AlexNet and ResNet18 were 97.59%, 97.32%, 97.95%, 97.31% and 99.49%, respectively. It can be concluded that the ResNet18 was the best performing model. In summary, all the five models were able to perform extremely well in

classifying monkeypox skin lesion images. The models can serve as the foundation of locally running applications on smartphones, enabling real time and instantaneous detection of infection. The accuracy and loss curve obtained for the ResNet18 model (best run) is depicted Fig. 9. The accuracies of all the models (all runs) are compared in Fig. 10.

#### 4. Discussion

ResNet-18 could perform better than GoogLeNet, GoogLeNet (Places 365), and SqueezeNet because it has a more straightforward and effective architecture that enables it to learn more complex features with fewer inputs. ResNet-18 employs residual blocks that combine and skip connections between layers, assisting in the prevention of vanishing gradient problem and enhancing gradient flow. Additionally, ResNet-18 uses less memory, and it has fewer convolutional layers than GoogLeNet and GoogLeNet (Places 365). Also, it does not make use of inception modules like GoogLeNet and GoogLeNet (Places 365), which have numerous branches with various filter sizes and pooling operations, which can amplify noise and redundancy in feature maps. Finally, ResNet-18 has more channels than SqueezeNet, which makes use of blocks that use squeeze and excitation to lessen the number of channels and then re-weight them according to their importance. That could lead to under-fitting and loss of information.

As an alternate to convolutional neural networks, a vision transformer model ViT B16 has been trained and validated on the same dataset. On using similar hyperparameters during the training and validation, it has been observed that the transformer-based model performs poorly in comparison to the convolutional neural networks. These can be chalked up to several reasons including but not limited to the limited size of the dataset as vision transformers require large training datasets since it has more parameters. The complexity of the dataset might also play an important role in explaining its superior performance on pre-trained convolutional neural networks as ResNet models perform well on datasets with high intraclass variation and low interclass variation. In stark contrast, vision transformers work better on datasets with high interclass variation and low intraclass variation.

The decisions made by deep learning models can be explained using explainable techniques called explainable artificial intelligence (XAI). Two techniques used in this work are Locally Interpretable Model-

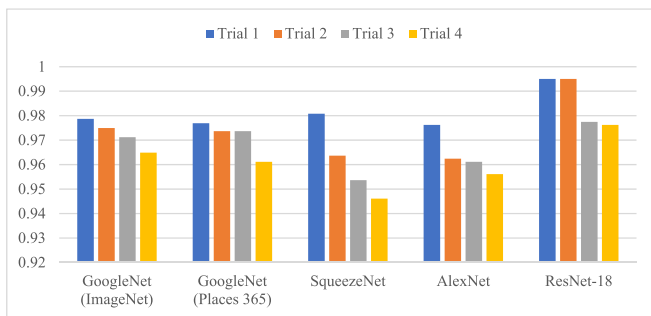


Fig. 10. Graphical comparison of the validation accuracies for all four trials.

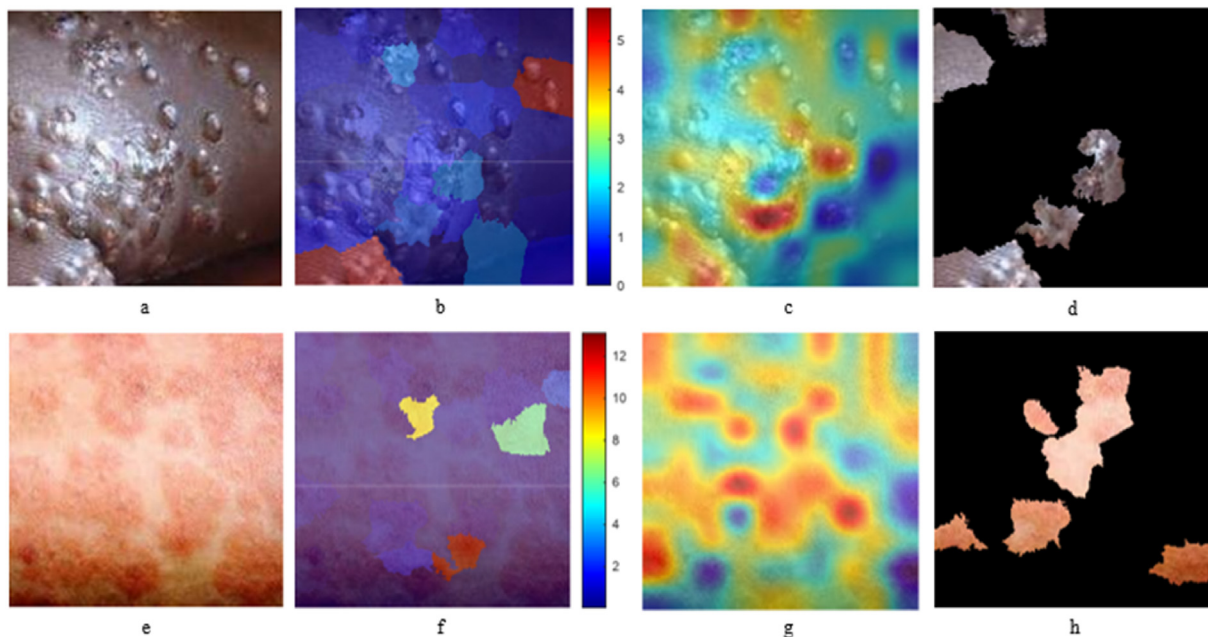


Fig. 11. 11(a) and 11(e) represent class monkeypox and class normal respectively. 11(b) and 11(f) illustrate their respective LIME images. 11(c) and 11(g) are the GradCAM visualization for the two classes. 11(d) and 11(h) show the top 7 features from the LIME visualization.

agnostic Explanations (LIME) and Grad-CAM. LIME is a visualization tool that highlights the important features as determined by the convolutional neural network to dictate the training in making predictions. LIME explains the predictions of the classifier by locally approximating it using an interpretable model [25]. A LIME object can be generated using a specific query point and a defined number of significant predictors. This will result in the creation of an artificial data set, which will be used to train a simple and easy-to-understand model that explains the predictions for the synthetic data centred around the query point. LIME is described in Fig. 11 (b) for class ‘Monkeypox’ and 11(f) for class ‘Normal’ is represented by super pixels that use a regression tree. The warmer areas represent the features of higher importance while the cooler areas have lower importance in the decisions leading to the prediction.

This is supported by the explanations made by Grad-CAM, as represented in Fig. 11 (c) for class ‘Monkeypox’ and 11(g) for class ‘Normal’. Grad-CAM is used to understand how the deep learning network makes its classification decisions [26]. Grad-CAM uses the gradient information

entering the last convolutional layer of the CNN to understand the importance of each neuron in making the desired decision. Like LIME explanations, the Grad-CAM explanations have regions with the warmer regions representing those that are more responsible for making predictions while the cooler regions are not as important to the decision-making process.

As represented in Fig. 11 (d) and 11(e) for classes ‘Monkeypox’ and ‘Normal’ respectively, the most important features in the decision making can also be used to create an effective segment of only the relevant features. These have been generated by creating a mask comprising of the top seven features used that is then multiplied with the original image to create a super pixel segmentation of the most important features.

The predictions made by the other convolutional neural networks can also be visualised using the same explainable AI techniques as represented in Fig. 12. The same image from the dataset has been used to test the five convolutional neural networks and explain the difference in their accuracies. As observable, the top performing ResNet-18 based trial is

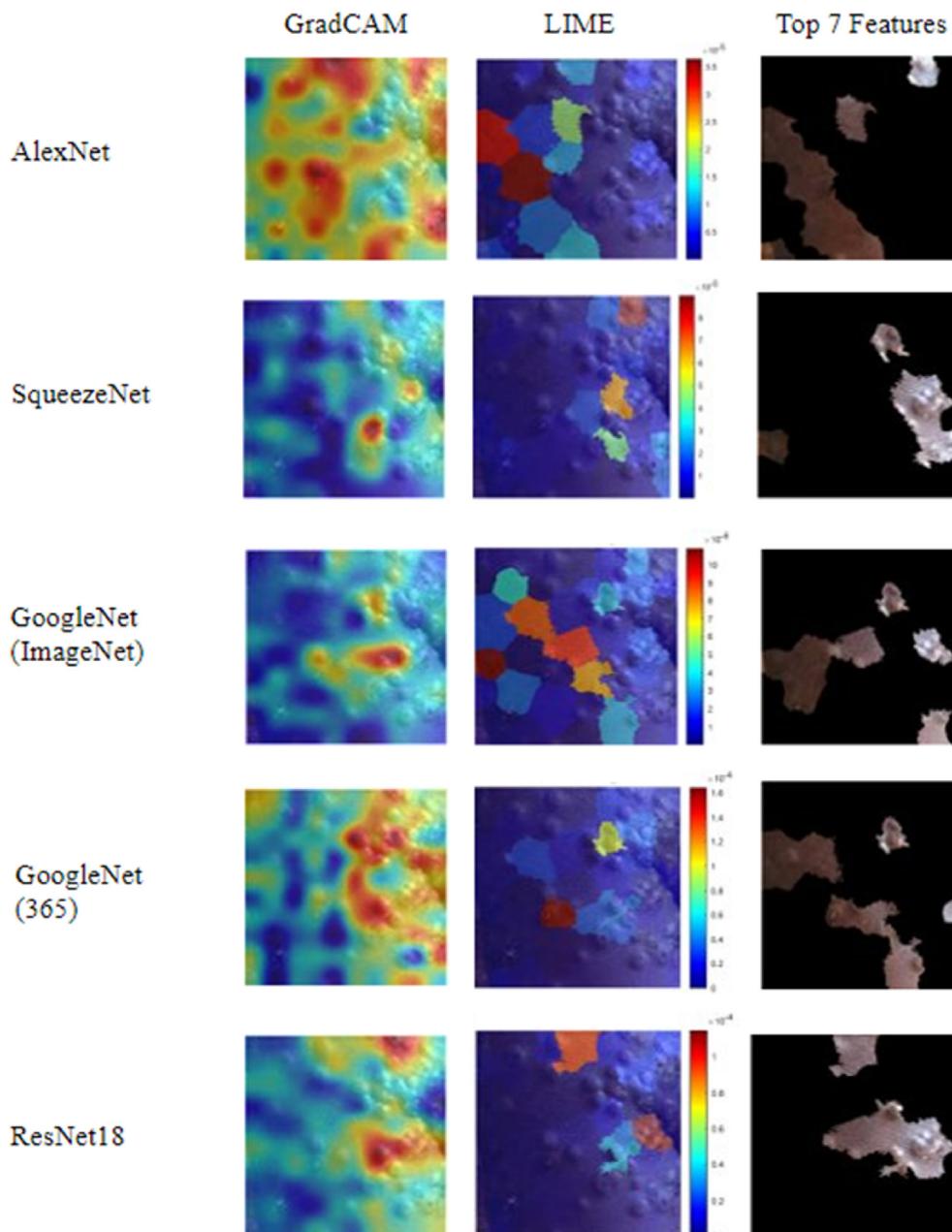


Fig. 12. Comparison of all convolutional neural network predictions' visualizations using GradCAM & LIME and further segmenting the top seven features using LIME.

**Table 6**

Comparison of researches which use deep learning to diagnose monkeypox virus.

Reference	Published year	Dataset	Deep learning models used	Results obtained by the best model				
				Accuracy	Sensitivity	Specificity	Precision	F1-score
Sitaula et al. [12]	2022	Monkeypox skin lesion images [27, 28]	13 baseline models and one ensemble model	87.13%	85.47%	–	85.44%	85.40%
Sahin et al. [13]	2022	Monkeypox skin lesion images from Kaggle [16]	Six baseline models	91.11%	90%	–	90%	90%
Abdelhamid et al. [15]	2022	Monkeypox skin lesion images from Kaggle [16]	Four baseline models	98.8%	62%	–	99.8%	76%
Alakus et al. [14]	2022	DNA sequences of monkeypox and human papilloma virus [23]	Custom BiLSTM model	96.08%	–	–	–	–
Akin et al. [28]	2022	Monkeypox skin lesion images from Kaggle [16]	12 baseline models	98.25%	96.55%	–	100%	98.25%
Our proposed study	2022	Monkeypox skin lesion images from Kaggle [16]	Five baseline models	99.49%	99.43%	100%	100%	99.49%

able to make its accurate predictions by accurately identifying the features from the skin lesions in the image, while the other models use other regions too.

In recent times, the Monkeypox virus has spread to over 75 countries, threatening to be the next pandemic [27]. Deep learning has slowly permeated the medical field in recent years, carrying innovation and solutions that are transforming the face of healthcare industry [10]. Deep learning enables the healthcare personnel to examine data at breakneck speeds while maintaining accuracy. It is an intelligent combination of mathematics and programming which filters patient data at an astonishing rate. They are accurate, fast and efficient. They also have the ability to learn new interesting patterns [11]. This can help gain valuable insights in disease diagnosis. Deep learning has been used to diagnose various diseases including monkeypox.

A few researches have been published which use deep learning techniques to diagnose this dreaded virus. Sitaula et al. [12] used pre-trained models to diagnose monkeypox. A maximum accuracy of 87.13% was obtained by the models. Xception and DenseNet-169 models were ensemble to optimize the accuracy. Explainability was obtained using gradient weighted class activation mapping. Sahin et al. [13] developed a mobile application to classify monkeypox images. An android application was developed using Java. The mobile captured the images and transferred it to a CNN model. A maximum accuracy of 91.11% was obtained. Abdelhamid et al. [15] used AI-Biruni Earth Radius Optimization for classification. A maximum accuracy of 98.8% was obtained by the models. In yet another research, Alakus et al. [14] used deep learning to diagnose monkeypox from Warts using DNA sequencing. A maximum accuracy of 96.08% was obtained by the classifiers. The authors concluded that DNA sequences can be used to diagnose the monkeypox virus from other similar diseases such as smallpox and measles. Explainable artificial intelligence (XAI) was used along with CNN to classify monkeypox skin lesion images in another study [28]. 12 deep learning models were used to classify 572 skin lesion images into two classes. Among all the models, MobileNetV2 obtained the highest accuracy of 98.25%. Table 6 gives a detailed comparison of various researches which use deep learning to diagnose monkeypox.

## 5. Conclusion

This study used four deep learning models: GoogLeNet, Places365-GoogLeNet, SqueezeNet, ResNet and AlexNet to diagnose Monkeypox from skin lesion images. Through this experiment, we conclude that GoogLeNet & Places365-GoogLeNet form a solid foundation for a deep learning-based software for fast diagnosis of Monkeypox. These models are also very lightweight and resource-efficient, making them ideal for application in various healthcare facilities. Further, these models can also be implemented in rural areas where the prevalence of PCR tests is limited. Combining a computer with a camera or a modern smartphone

with this model implemented can serve as the first step of instant diagnosis. ResNet-18 can be used due to its robustness in scenarios where the device hardware is not present. On severely performance-limited devices with slower hardware, SqueezeNet and AlexNet can also serve as a good foundation for a CNN model to detect the presence of Monkeypox infection in the patient with images of skin lesions with a minimal drop in accuracy. Given the ubiquity of smartphones and their ever-improving cameras and processing units, the feasibility of using these deep learning-based models is excellent. In smartphones released over the past few years, chip designers such as Apple, Google, Qualcomm and MediaTek have designed dedicated hardware in their system-on-chips to accelerate machine learning and deep learning tasks. Leveraging the high-resolution smartphone sensors coupled with the dedicated high-performance hardware for running these models, this method to identify the presence of monkeypox infection will help bring an effective, reliable, low-cost and reliable monkeypox detection to the masses and in places where the infrastructure to support PCR testing is unavailable.

## Credit author statement

Tushar Nayak: Software, Writing - original draft; Krishnaraj Chadaga: Data curation, Formal analysis; Niranjana Sampathila: Conceptualization, Project administration; Hilda Mayrose: Investigation, Methodology; Nitila Gokulkrishnan: Software, resources; Muralidhar Bairy G: Writing - review & editing, Visualization; Srikanth Prabhu: Funding acquisition, Supervision; Swathi K S: Writing - review & editing, Funding acquisition; and Shashikiran Umakanth: Validation, Supervision.

## Funding sources

This research received no external funding.

## 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.

## Acknowledgements

We would like to thank Manipal Academy of Higher Education for the computing laboratory facilities.

## References

- [1] Rizk JG, Lippi G, Henry BM, Forthall DN, Rizk Y. Prevention and treatment of monkeypox. *Drugs* 2022 Jun;82(9):957–63. <https://doi.org/10.1007/s40265-022-01742-y>.



- [2] Yinka-Ogunleye A, Aruna O, Dalhat M, Ogoina D, McCollum A, Disu Y, Mamadu I, Akinpelu A, Ahmad A, Burga J, Ndoreraho A. Outbreak of human monkeypox in Nigeria in 2017–18: a clinical and epidemiological report. *Lancet Infect Dis* 2019 Aug 1;19(8):872–9. [https://doi.org/10.1016/S1473-3099\(19\)30294-4](https://doi.org/10.1016/S1473-3099(19)30294-4).
- [3] Zachary KC, Shenoy ES. Monkeypox transmission following exposure in healthcare facilities in nonendemic settings: low risk but limited literature. *Infect Control Hosp Epidemiol* 2022 Jul;43(7):920–4. <https://doi.org/10.1017/ice.2022.152>.
- [4] Chadha J, Khullar L, Gulati P, Chhibber S, Harjai K. Insights into the monkeypox virus: making of another pandemic within the pandemic? *Environ Microbiol* 2022 Oct;24(10):4547–60. <https://doi.org/10.1111/1462-2920.16174>.
- [5] Uwishema O, Adekunbi O, Peñamante CA, Bekele BK, Khoury C, Mhanna M, Nicholas A, Adanur I, Dost B, Onyeaka H. The burden of monkeypox virus amidst the Covid-19 pandemic in Africa: a double battle for Africa. *Annals of medicine and surgery* 2022 Aug 1;80:104197.
- [6] Di Giulio DB, Eckburg PB. Human monkeypox: an emerging zoonosis. *Lancet Infect Dis* 2004 Jan 1;4(1):15–25. [https://doi.org/10.1016/S1473-3099\(03\)00856-9](https://doi.org/10.1016/S1473-3099(03)00856-9).
- [7] De Baetselier I, Van Dijk C, Kenyon C, Coppens J, Michiels J, de Block T, Smet H, Coppens S, Vanroye F, Bugert JJ, Giral P. Retrospective detection of asymptomatic monkeypox virus infections among male sexual health clinic attendees in Belgium. *Nat Med* 2022 Aug 12:1–5. <https://doi.org/10.1038/s41591-022-02004-w>.
- [8] Altindis M, Puca E, Shapo L. Diagnosis of monkeypox virus—An overview. *Trav Med Infect Dis* 2022 Sep 13:102459. <https://doi.org/10.1016/j.tmaid.2022.102459>.
- [9] Mondolfi AP, Guerra S, Muñoz M, Luna N, Hernandez MM, Patino LH, Reidy J, Banu R, Shrestha P, Liggayy B, Umeaku A. Evaluation and validation of an RT-PCR assay for specific detection of Monkeypox virus (MPXV). *J Med Virol* 2022 Oct 21. <https://doi.org/10.1002/jmv.28247>.
- [10] Chadaga K, Prabhu S, Sampathila N, Nireshwalya S, Katta SS, Tan RS, Acharya UR. Application of artificial intelligence techniques for monkeypox: a systematic review. *Diagnostics* 2023 Feb 21;13(5):824. <https://doi.org/10.3390/diagnostics13050824>.
- [11] Norgeot B, Glicksberg BS, Butte AJ. A call for deep-learning healthcare. *Nat Med* 2019 Jan;25(1):14–5. <https://doi.org/10.1038/s41591-018-0320-3>.
- [12] Sitaula C, Shahi TB. Monkeypox virus detection using pre-trained deep learning-based approaches. *J Med Syst* 2022 Oct 6;46(11):78. <https://doi.org/10.1007/s10916-022-01868-2>.
- [13] Sahin VH, Oztel I, Yolcu Oztel G. Human monkeypox classification from skin lesion images with deep pre-trained network using mobile application. *J Med Syst* 2022 Oct 10;46(11):79. <https://doi.org/10.1007/s10916-022-01863-7>.
- [14] Alakus TB, Baykara M. Comparison of monkeypox and Wart DNA sequences with deep learning model. *Appl Sci* 2022 Oct 11;12(20):10216. <https://doi.org/10.3390/app122010216>.
- [15] Abdelhamid AA, El-Kenawy ES, Khodadadi N, Mirjalili S, Khafaga DS, Alharbi AH, Ibrahim A, Eid MM, Saber M. Classification of monkeypox images based on transfer learning and the Al-Biruni Earth Radius Optimization algorithm. *Mathematics* 2022 Oct 2;10(19):3614. <https://doi.org/10.3390/math10193614>.
- [16] Ali SN, Ahmed M, Paul J, Jahan T, Sani SM, Noor N, Hasan T. Monkeypox skin lesion detection using deep learning models: a feasibility study. arXiv preprint arXiv:2207.03342, <https://doi.org/10.48550/arXiv.2207.03342>; 2022.
- [17] Lewy D, Mańdziuk J. An overview of mixing augmentation methods and augmentation strategies. *Artif Intell Rev* 2023 Mar;56(3):2111–69. <https://doi.org/10.1007/s10462-022-10227-z>.
- [18] Ballester P, Araujo R. On the performance of GoogLeNet and AlexNet applied to sketches. In: Proceedings of the AAAI conference on artificial intelligence. 30; 2016. <https://doi.org/10.1609/aaai.v30i1.10171>. Feb 21-No. 1.
- [19] Matei A, Glavan A, Talavera E. Deep learning for scene recognition from visual data: a survey. November 11–13. In: InHybrid artificial intelligent systems: 15th international conference, HAIS 2020. Gijón, Spain: Cham: Springer International Publishing; 2020. p. 763–73. [https://doi.org/10.1007/978-3-030-61705-9\\_64](https://doi.org/10.1007/978-3-030-61705-9_64). Proceedings 2020 Nov 4.
- [20] Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X, Xu B. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *Eur Radiol* 2021 Aug;31:6096–104. <https://doi.org/10.1007/s00330-021-07715-1>.
- [21] Lu S, Lu Z, Zhang YD. Pathological brain detection based on AlexNet and transfer learning. *J comp. sci.* 2019 Jan 1;30:41–7. <https://doi.org/10.1016/j.jocs.2018.11.008>.
- [22] Wu Z, Shen C, Van Den Hengel A. Wider or deeper: revisiting the resnet model for visual recognition. *Pattern Recogn* 2019 Jun 1;90:119–33. <https://doi.org/10.1016/j.patrec.2019.01.006>.
- [23] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J. An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, <https://doi.org/10.48550/arXiv.2010.11929>; 2020.
- [24] Jiao Y, Du P. Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quantitative Biology* 2016 Dec;4(4):320–30. <https://doi.org/10.1007/s40484-016-0081-2>.
- [25] Ribeiro MT, Singh S, Guestrin C. Why should i trust you?" Explaining the predictions of any classifier. In: InProceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016 Aug 13. p. 1135–44. <https://doi.org/10.1145/2939672.2939778>.
- [26] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 618–26. <https://doi.org/10.48550/arXiv.1610.02391>.
- [27] Gessain A, Nakoune E, Yazdanpanah Y. Monkeypox. *New England J. Med.* 2022 Nov 10;387(19):1783–93. <https://doi.org/10.1056/NEJMra2208860>.
- [28] Akin KD, Gurkan C, Budak A, Karataş H. Classification of monkeypox skin lesion using the explainable artificial intelligence assisted convolutional neural networks. *Avrupa Bilim ve Teknoloji Dergisi* 2022;(40):106–10. <https://doi.org/10.31590/ejasat.1171816>.