Konya Mühendislik Bilimleri Dergisi, c. 12, s. 4, 822-837, 2024
Konya Journal of Engineering Sciences, v. 12, n. 4, 822-837, 2024
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1471289

DEEP LEARNING-BASED ADAPTIVE ENSEMBLE LEARNING MODEL FOR CLASSIFICATION OF MONKEYPOX DISEASE

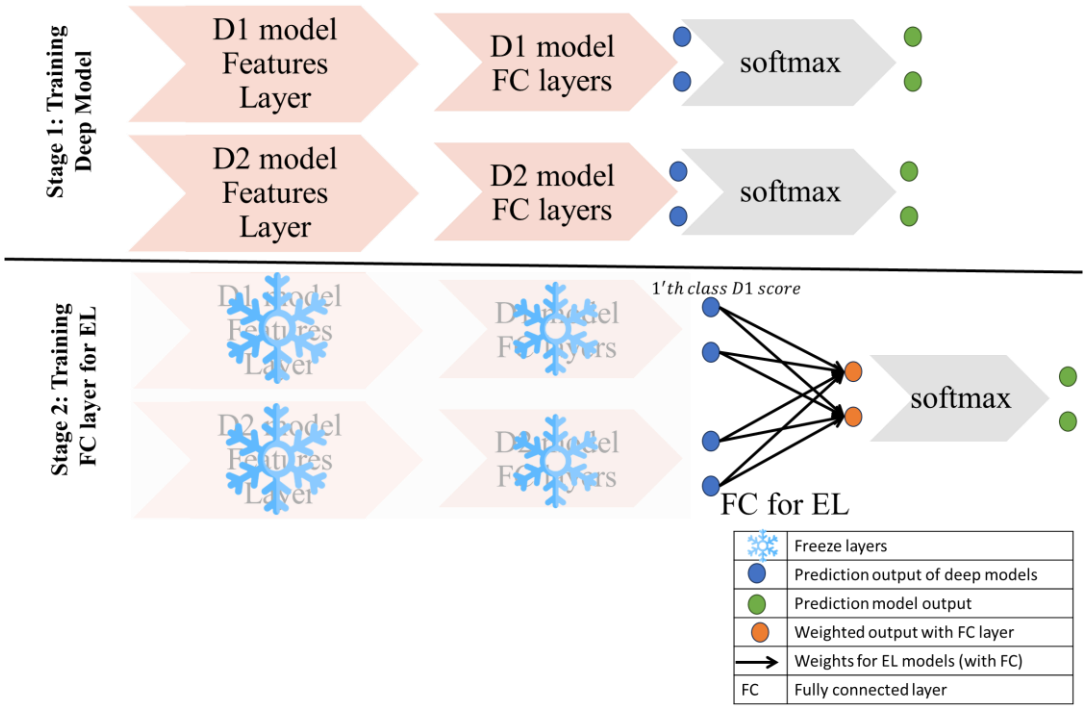
[1,*] Hüseyin ÜZEN, [2] Hüseyin FIRAT

[1]*Bingöl University, Engineering-Architecture Faculty, Computer Engineering Department, Bingöl, TÜRKİYE*
[2]*Dicle University, Engineering Faculty, Computer Engineering Department, Diyarbakır, TÜRKİYE*
[1] huzen@bingol.edu.tr, [2] huseyin.firat@dicle.edu.tr

Highlights

- Adaptive ensemble learning model for monkeypox disease detection
- Adaptive ensemble learning model uses a weight learning approach with a fully connected layer
- Two different deep learning models are combined for the ensemble learning approach
- Monkeypox disease has been successfully detected with two different datasets used in experimental studies

Graphical Abstract



Concatenate strategy of the proposed DL-AEL model

***Corresponding Author:** Hüseyin ÜZEN, huzen@bingol.edu.tr

Konya Mühendislik Bilimleri Dergisi, c. 12, s. 4, 822-837, 2024
Konya Journal of Engineering Sciences, v. 12, n. 4, 822-837, 2024
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1471289

DEEP LEARNING-BASED ADAPTIVE ENSEMBLE LEARNING MODEL FOR CLASSIFICATION OF MONKEYPOX DISEASE

[1,*] Hüseyin ÜZEN, [2] Hüseyin FIRAT

[1] *Bingöl University, Engineering-Architecture Faculty, Computer Engineering Department, Bingöl, TÜRKİYE*
[2] *Dicle University, Engineering Faculty, Computer Engineering Department, Diyarbakır, TÜRKİYE*
[1] huzen@bingol.edu.tr, [2] huseyin.firat@dicle.edu.tr

(Received: 21.04.2024; Accepted in Revised Form: 03.09.2024)

ABSTRACT: Monkeypox a viral disease resembling smallpox often transmitted *via* animal contact or human-to-human transmission. Symptoms include fever, rash, and respiratory issues. Healthcare experts initially may confuse it with chickenpox or measles due to its rarity, but swollen lymph nodes typically distinguish it. Diagnosis involves tissue sampling and polymerase chain reaction (PCR) testing, although PCR tests have limitations like time consumption and false negatives. Deep learning-based detection offers advantages over PCR, including reduced risk of exposure, quicker results, and improved accuracy. In this study, a novel adaptive ensemble learning (AEL)-based model for monkeypox diagnosis is proposed. This proposed ensemble learning model aims to enhance diagnosis accuracy by combining different deep learning models, leveraging an adaptive approach for model combination. Experimental studies using MSLD and MSID datasets show promising results, with ensemble models achieving high accuracy, precision, recall, and F1 scores. The ResNet101+VGG16 (92.46% accuracy, 92.75% precision, 93.22% recall, and 92.98% F1 score) ensemble model performs best for MSLD, while DenseNet121+Xception (97.58% accuracy, 96.57% precision, 95.74% recall, and 96.14% F1 score) excels for MSID. In addition, the proposed AEL model outperforms previous studies using the same datasets, showcasing its potential for improved monkeypox diagnosis.

Keywords: *Adaptive Ensemble Learning, Deep Learning, Monkeypox Disease, Classification*

1. INTRODUCTION

Monkeypox is a viral disease that causes symptoms similar to those of smallpox [1]. It is caused by the monkeypox virus, which belongs to the Orthopoxvirus genus, the same genus as the variola virus, which causes smallpox [2]. Monkeypox was first discovered in laboratory monkeys in 1958, hence its name [3]. The disease primarily occurs in Central and West Africa, particularly in remote areas where people have close contact with animals that may carry the virus. Human-to-human transmission can also occur, typically through respiratory droplets or contact with infected bodily fluids or lesions [4]. Symptoms of monkeypox usually appear within 5 to 21 days after infection and include fever, headache, muscle aches, backache, swollen lymph nodes, chills, exhaustion, and a characteristic rash that often begins on the face and then spreads to other parts of the body. In some cases, the rash may develop into fluid-filled blisters that crust over [5].

Healthcare experts might initially consider comparable skin eruptions like chickenpox or measles as the source of monkeypox due to its rarity. However, swollen lymph nodes often distinguish monkeypox from other pox varieties [6]. These experts will collect a tissue sample from an individual with an ongoing infection on their body to diagnose monkeypox. This sample is then forwarded to a lab for examination utilizing polymerase chain reaction (PCR). The PCR testing approach entails direct exposure to the monkeypox virus, given the likelihood of virus transmission during sample collection, transfer, and analysis of lesions for diagnosis. Moreover, PCR tests are notably time-intensive and susceptible to inaccurate negative outcomes [7].

Computer-aided automated systems in monkeypox detection provide numerous advantages over conventional PCR testing methods [4]. Enhancing diagnostic accuracy and speed stands out as a primary

***Corresponding Author:** Hüseyin ÜZEN, huzen@bingol.edu.tr

incentive behind creating such systems. Monkeypox, characterized by symptoms like fever, rash, and respiratory issues, shares resemblances with other viral illnesses like measles and chickenpox, posing challenges for healthcare experts in diagnosis [8]. In recent times, progress in areas like artificial intelligence and machine learning has elevated them to essential tools for healthcare experts [9]. Deep Learning (DL), a subset of artificial intelligence, facilitates model construction, automatic feature extraction sans human intervention, model training, and result generation [10]. Various imaging modalities find extensive application in medicine, aiding in the diagnosis of diverse ailments like pneumonia, respiratory illnesses like tuberculosis [11], COVID19 [12], and others. DL-based analysis of medical images has garnered considerable attention in recent research endeavors. The drawbacks associated with the PCR test for diagnosing monkeypox underscore the potential advantages of employing DL-based approaches in this context. Firstly, PCR tests necessitate direct exposure to the monkeypox virus during sample collection, transfer, and analysis, raising concerns about the risk of transmission. In contrast, DL-based diagnosis does not involve physical contact with infectious material, minimizing the risk of exposure for healthcare experts. Furthermore, PCR tests are known for their time-consuming nature, often requiring several hours or even days to generate results. In contrast, DL algorithms can provide rapid and automated analysis of medical images, potentially reducing the time required for diagnosis and enabling quicker initiation of appropriate treatment. Another limitation of PCR tests is their susceptibility to false-negative results, which can occur due to various factors such as inadequate sample quality or viral load. DL-based approaches, however, can leverage large datasets to learn complex patterns and features indicative of monkeypox, potentially enhancing diagnostic accuracy and reducing the likelihood of false negatives. Additionally, PCR tests typically require specialized laboratory equipment and trained personnel, limiting their accessibility, especially in resource-limited settings. In contrast, DL-based diagnosis can be implemented using standard medical imaging technology and may offer a more scalable and cost-effective solution for monkeypox diagnosis, particularly in regions with limited healthcare infrastructure. Overall, the advantages of DL-based diagnosis, including reduced risk of exposure, faster turnaround time, enhanced accuracy, and greater accessibility, highlight its potential to overcome the limitations associated with PCR testing for monkeypox diagnosis. By leveraging advanced DL techniques, healthcare experts can improve diagnostic capabilities and ultimately enhance patient care and outbreak management efforts.

Lately, there have been investigations into utilizing DL, specifically convolutional neural networks (CNNs), for diagnosing monkeypox. Nayak et al. [7] used transfer learning-based ResNet10, ResNet18, ResNet50 and SqueezeNet to distinguish monkeypox from measles and chickenpox. As a result of experimental studies on a 4-class MSID dataset, they obtained the best accuracy value with ResNet18 model with 91.11%. Ali et al. [13] developed an ensemble model consisting of InceptionV3, ResNet50, and VGG16 to diagnose monkeypox. They conducted experimental studies on the MSLD dataset containing two classes to assess the performance of the developed model. As a result of these studies, individual usage of the models yielded accuracy values of 74.07%, 82.96%, and 81.48%, respectively. However, they found an accuracy value of 79.26% with the ensemble model comprising the three models. Haque et al. [14] proposed a model combining convolutional block attention module (CBAM) with different models based on transfer learning. The models they used in conjunction with CBAM were MobileNetV2, EfficientNetB3, DenseNet121, Xception, and VGG19. Through experimental studies conducted on the two-class MSLD dataset to analyze the effectiveness of the models, they achieved the most successful results with a model consisting of Xception-CBAM-Dense layers, attaining an accuracy value of 83.89%. Sahin et al. [15] conducted experimental studies on the two-class MSLD dataset to test the MobileNetV2 model proposed for the diagnosis of monkeypox. In these experimental studies, an accuracy of 91.11% was achieved with MobileNetV2. Similarly, Almufareh et al. [16] utilized the EfficientNetB4 model. In experimental studies conducted on both the MSLD and MSID datasets, they found accuracy values of 88.89% and 92%, respectively. Alakus et al. [17] developed Siamese deep learning as a new model. They tested this model on the MSLD dataset and obtained an accuracy of 91.09%. Ural et al. [18] conducted experimental studies using the MSID dataset with the AlexNet and

VGG16 models. Through these experimental studies with the dataset, they achieved accuracy values of 71% and 80%, respectively. Similarly, Uysal et al. [19] developed a model consisting of CNN+LSTM for the diagnosis of monkeypox. This model was tested using the MSID dataset, resulting in an accuracy value of 87%.

In recent years, when DL-based methods were examined for the diagnosis of monkeypox, it was observed that the results were low. In this study, a novel DL-based model is proposed to increase the accuracy of disease diagnosis. This model is a ensemble learning model. The purpose of using the ensemble learning model is to generate more effective results by combining different deep learning models. Moreover, many researchers have been drawn to the idea because numerous trials are required to create a network model from scratch. The fundamental basis of the ensemble learning approach involves taking the average of the outputs of two different models. On the other hand, while some ensemble learning approaches merge two different models by weighting the outputs of neural networks, these weights can provide a more effective approach to combining the models. However, adjusting these weights depends on the approach of the models and the problem at hand.

In this study, an adaptive approach to combining two different CNN-based models for ensemble learning is proposed by adding fully connected layers. In this model, known as Adaptive Ensemble Learning (AEL), instead of traditional voting or averaging outputs, a weight learning approach is employed using a fully connected layer. The backpropagation policy of deep learning is used in the learning phase of the weights. The proposed AEL consists of two stages. In the first stage, two different network models are trained with the training dataset. In the second stage, the softmax layer, which obtains the trained models' full probabilistic scores, is removed, and the outputs of the two models are concatenated. Then, a fully connected layer is applied to the concatenated outputs to obtain the model's output values. Finally, a softmax layer is applied to obtain the probabilistic distribution of this output. In the proposed AEL model, binary combinations of Xception, VGG16, DenseNet121, MobileNet, InceptionV3 and ResNet101 network models are used. Experimental studies were conducted using the two-class MSLD and four-class MSID datasets to analyze the effectiveness of the proposed model in diagnosing monkeypox. With the MSLD dataset, an ensemble model of ResNet101+VGG16 achieved an accuracy of 92.46%, precision of 92.75%, recall of 93.22%, and an F1 score of 92.98%. On the MSID dataset, an ensemble model of DenseNet121+Xception achieved an accuracy of 97.58%, precision of 96.57%, recall of 95.74%, and an F1 score of 96.14%. Upon examining various experimental studies conducted on both datasets, it was observed that the best results were obtained with the ResNet101+VGG16 ensemble model for MSLD and DenseNet121+Xception ensemble model for MSID. Comparisons with other studies using the same datasets from the literature also indicate that the proposed AEL model produces more successful results.

The remainder of the study is organized as follows: In Section 2, the proposed model and the monkeypox datasets used are explained. Section 3 presents the results of the experimental studies and compares them with different studies in literature. The final section, Section 4, is the conclusion section, which provides an overall summary of the study.

2. MATERIAL AND METHODS

In this section, firstly, the proposed deep learning-based adaptive ensemble learning model for the detection and classification of monkeypox disease is discussed in detail. In the following subsection, the monkeypox dataset used in the study is explained.

2.1. Proposed Deep Learning-Based Adaptive Ensemble Learning Model

In this study, a deep learning-based adaptive ensemble learning (DL-AEL) approach is presented for the detection and classification of monkeypox disease. Adaptive ensemble learning, based on adaptive learning, is a deep learning technique that combines the principles of ensemble learning with adaptability, allowing the ensemble to evolve over time and improve its performance. Adaptive learning

takes this concept further by allowing the ensemble to dynamically adjust its composition or the weights of base models based on incoming data or changing patterns. A common approach to adaptability is dynamic weighting, where the ensemble assigns different weights to each base model depending on its recent performance or the current difficulty of the data. Models performing well on the current data may receive higher weights, while those performing poorly may receive lower weights. Determining these weights is the fundamental problem in this approach. In this study, a deep learning approach is used to learn these weights. The overall structure of the proposed DL-AEL model is illustrated in Figure 1.

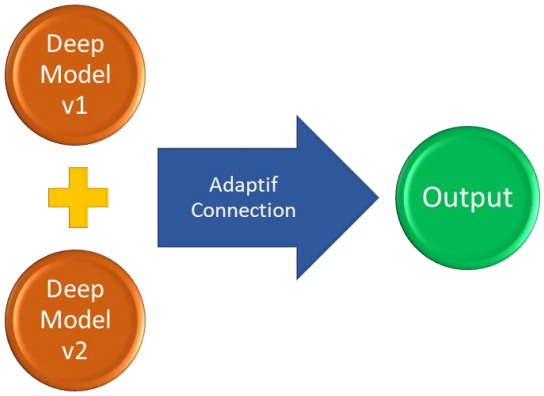


Figure 1. Overall Structure of The Proposed DL-AEL Model

As can be seen in Figure 1, the proposed approach is based on the combination of two deep learning models. Unlike the basic ensemble learning approach, the concatenate strategy relies on dynamic weights. These weights are trained using the learning approach of deep learning and are updated to the most suitable weights for testing. In the combination of network architectures, first, the classification layer, or in other words, the softmax layer, of the networks is removed. Then, the outputs of these network architectures are merged. A fully connected (FC) layer is added to the merged layers. Finally, a softmax layer is added to the obtained output to obtain the prediction output.

2.1.1 Deep learning models used for ensemble learning and concatenate strategy

In this study, Xception, VGG16, DenseNet121, MobileNet, InceptionV3 and ResNet101 network models were used for ensemble learning. These networks were chosen because of their unique strengths and high performance in medical image analysis. The Xception model is an efficient model that provides highly efficient feature extraction using extended depthwise separable convolutions. This model can improve diagnostic accuracy thanks to its ability to effectively analyse the complex patterns and textures of skin lesions [20]. The VGG16 is a traditional network architecture with a simple and regularised layer structure and is the most used high-performance model for image classification tasks. Moreover, the regular structure of the VGG16 is powerful in detecting prominent features of skin lesions [21]. The ResNet model is known for solving the vanishing gradient problem encountered when increasing the depth of deep learning models. This feature allows the model to learn more complex features through deep layers. Its simple and effective structure is suitable for fast and reliable diagnostics [22]. The InceptionV3 is favored for its computational efficiency and high accuracy in image classification tasks. It employs 1x1 convolutions to reduce computational load and multi-scale feature extraction to capture diverse patterns effectively. Its modular design facilitates parallelization and scalability, making it suitable for a wide range of applications and enabling efficient learning [23]. The DenseNet121 model is characterized by the structure in which each layer combines feature maps from all previous layers. This structure enables better propagation of low and high-level features and more effective feature learning. The tightly coupled structure of the model accelerates the learning process and improves its overall performance [24]. The MobileNet is a lightweight model optimized for mobile and embedded devices. It requires low computational power while operating with high efficiency [25]. In addition, its tunable

convolutional structure makes it a powerful method for limited datasets. The combination of different models allows to exploit the strengths of each model and compensate for its weaknesses. This diversity improves the overall performance of the model and increases the diagnostic accuracy. These network models were first trained with monkeypox datasets. Then these models are combined as given in Equation 1.

$$P = softmax\left(FC\left(concat(D_1, D_2)\right)\right) \tag{1}$$

In Equation 1, P is the output prediction vector of the DL-AEL model. FC represents fully connected layers. On the other hand, $D1$ and $D2$ are the pre-trained deep learning models used. These models are binary combinations of Xception, VGG16, DenseNet121, MobileNet, InceptionV3 and ResNet101 network models, respectively.

The concatenating strategy significantly impacts the success. Initially, in the traditional ensemble learning model, the outputs of two network models are averaged, expecting both models to contribute equally. In other words, in this model, referred to as voting, the average of the predictions of the two models is taken as input. However, the contribution of a poorly performing model can mislead the ensemble learning model. Therefore, as shown in Equation 1, the FC layer is applied. The weights in this layer represent the weights in the combination of models for ensemble learning. Through this FC layer, the outputs are weighted. This proposed structure is illustrated in Figure 2.

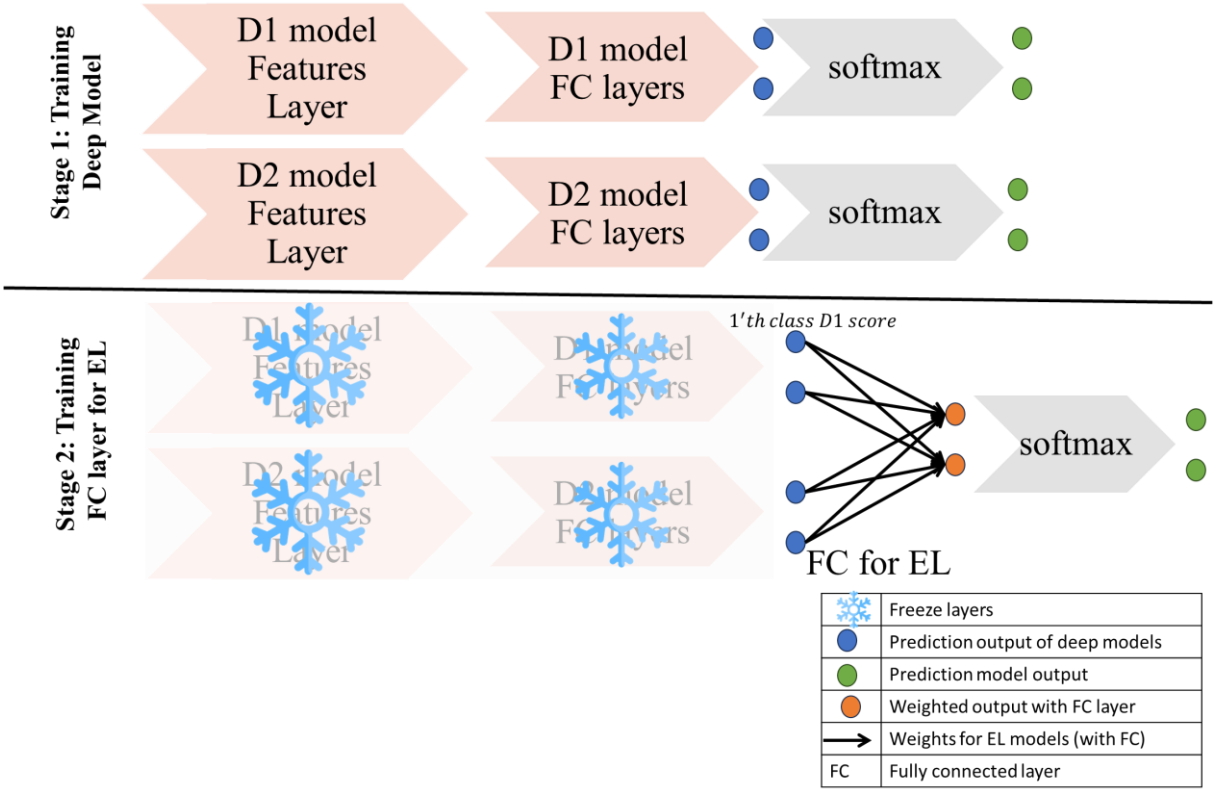


Figure 2. Concatenate strategy of the proposed DL-AEL model

As seen in Figure 2, the proposed DL-AEL approach consists of two stages. In the first stage, the D1 and D2 network models are trained with the training dataset. In the second stage, the softmax layer, which obtains the full probabilistic score of the trained models, is removed, and the outputs of the two models are concatenated. Then, a fully connected layer is applied to the concatenated outputs to obtain the model's output values. Finally, a softmax layer is applied to obtain the probabilistic distribution of this output.

2.1.2 Xception

The Xception (Extreme Inception) is a deep convolutional neural network architecture developed by François Chollet [20]. It builds upon the Inception model by replacing the standard Inception modules with depthwise separable convolutions. This results in a more efficient model with fewer parameters and improved performance. The architecture consists of 36 convolutional layers structured into 14 modules, all using residual connections. Xception's main innovation is the use of depthwise separable convolutions, which decouple spatial and cross-channel convolutions, allowing for more efficient feature extraction. This model has shown strong performance on various image classification tasks and serves as a powerful alternative to traditional Inception models.

2.1.3 VGG16

The VGG16, developed by the Visual Graphics Group at Oxford, is a 16-layer convolutional neural network with 13 convolutional layers and 3 fully connected layers. It uses 3x3 receptive fields in convolutional layers, organized in blocks followed by max-pooling layers. The number of filters doubles after each max-pooling layer, starting from 64 to 512. ReLU activation is used throughout. Originally trained on the ImageNet dataset, VGG16 excels in image classification and object detection due to its simple and uniform architecture. It was a top performer in the 2014 ImageNet Large Scale Visual Recognition Challenge [21].

2.1.4 DenseNet121

The DenseNet121 is a CNN architecture known for its densely connected layers. Developed by Gao Huang et al. [24], it features 121 layers in total, organized into dense blocks where each layer is connected to every other layer in a feed-forward fashion. This connectivity pattern helps in feature reuse and gradient propagation, leading to improved accuracy and efficiency. DenseNet121 uses bottleneck layers and transition layers between dense blocks to control the number of parameters and facilitate downsampling. It has been widely adopted for tasks such as image classification and object detection, offering competitive performance with efficient use of parameters.

2.1.5 MobileNet

The MobileNet is a lightweight CNN architecture designed for mobile and embedded vision applications. Developed by Google researchers, it aims to achieve high accuracy with low computational resources and a small model size. MobileNet primarily utilizes depthwise separable convolutions, which split the standard convolution into separate depthwise and pointwise convolutions. This approach reduces the number of parameters and computations, making the model more efficient for deployment on devices with limited computational power [25].

2.1.6 InceptionV3

The InceptionV3 is a CNN architecture developed by Google. It builds on the original Inception architecture but introduces several improvements to enhance performance and efficiency. The key features of InceptionV3 include the use of inception modules with varying filter sizes (1x1, 3x3, 5x5) to capture multi-scale features efficiently. It also incorporates factorization into smaller convolutions to reduce computational cost while maintaining representation power. InceptionV3 includes batch normalization and ReLU activations after each convolutional layer for faster training and better convergence [23]. This model is known for its strong performance on image classification tasks and is widely used in research and practical applications due to its balance of accuracy and computational efficiency.

2.1.7 ResNet101

ResNet101 is a deep CNN architecture known for its depth and residual learning framework. Developed by Microsoft Research [22], it belongs to the ResNet (Residual Network) family, which addresses the vanishing gradient problem by introducing skip connections or shortcuts that skip one or more layers. These shortcuts enable the gradient to flow more directly during backpropagation, facilitating easier training of very deep networks. ResNet101 specifically has 101 layers, featuring residual blocks where each block consists of multiple convolutional layers with skip connections. ResNet101 specifically uses bottleneck architecture within its residual blocks, which helps reduce computational complexity while maintaining model depth. This architecture has proven effective in various computer vision tasks, offering state-of-the-art performance on tasks like image classification and object detection.

2.1.8 Training the proposed model and loss function

As detailed in the previous section, the training of the proposed model is conducted in two stages. In the first stage, D1 and D2 models are trained. In experimental studies, Xception, VGG16, DenseNet121, MobileNet, InceptionV3, and ResNet101 network models are used for D1 and D2 models. In the second stage, the layers of these models are frozen, and the final layers of the DL-AEL model are trained. Categorical Cross-Entropy loss function is used in the training of the proposed network model. The Categorical Cross-Entropy loss function used is shown in Equation 2.

$$L_{CLF} = -\sum_{k}^{M} Y_k log(P_k) \tag{2}$$

In Equation 2, L_{CLF} is the loss value of the classification model, Y and P are the expected and prediction vectors respectively. M is the number of classes and k is the index of the classes.

2.2. Monkeypox Datasets

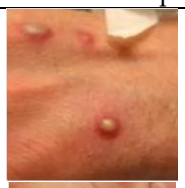
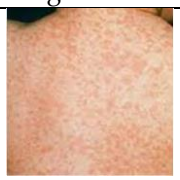
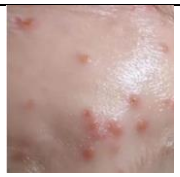
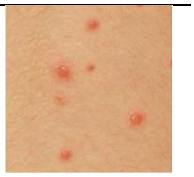
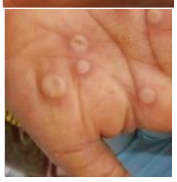
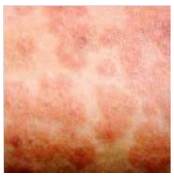
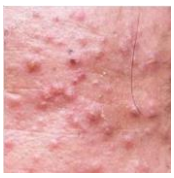
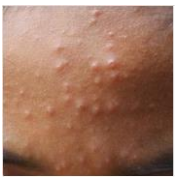
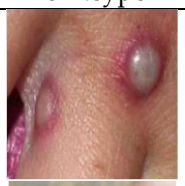
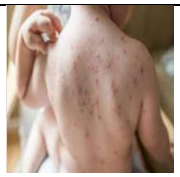
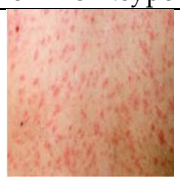
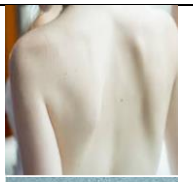
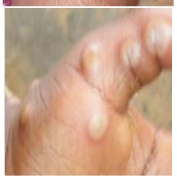
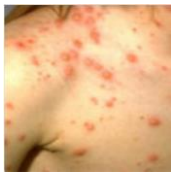
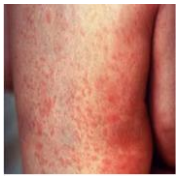
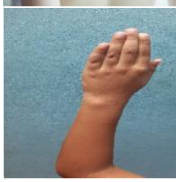
Two different monkeypox datasets were used in this study. The first dataset is the Monkeypox Skin Lesion Dataset (MSLD) [13] which contains 2 classes (monkeypox and non-monkeypox). The other dataset is the Monkeypox Skin Images Dataset (MSID) [26], which contains 4 classes (normal, monkeypox, measles, and chickenpox). The original MSLD dataset contains a total of 228 images. Among these, 102 belong to the monkeypox class, while the remaining 126 represent the others class, meaning cases other than monkeypox (non-monkeypox). However, data augmentation has been applied to the MSLD dataset within the scope of experimental studies. The data augmentation techniques used in this study are as follows: (Horizontal flip: True, Fill mode: Reflective, Shear range: 2%, Height shift Up to: 2%, Zoom range: 2%, Rotation Range Randomly: 0°- 45°, and Width shift Up to: 2%). Following data augmentation, the MSLD dataset used in the study contains 1428 images belonging to the monkeypox class and 1764 images representing the non-monkeypox class. In the experimental studies, 1148 monkeypox images and 1526 non-monkeypox images were used for training in the MSLD dataset. For validation, 140 monkeypox and 126 non-monkeypox images were used, while for testing, 140 monkeypox and 112 non-monkeypox images were used. Another dataset, MSID, contains four classes. The original MSID dataset includes 279 monkeypox, 91 measles, 107 chickenpox, and 293 normal images, totaling 770 images. Data augmentation has been applied to the images in this dataset, increasing the total number of images to 5390. In the MSID dataset, 4312 images are used for training, while 539 images each are used for validation and testing. Sample images of both datasets are given in Table 1.

3. RESULTS AND DISCUSSION

Numerous experimental studies have been conducted to analyze the performance of the proposed

DL-AEL model in detail. These experimental studies are presented in this section. In the rest of the section, firstly, the parameter settings and evaluation criteria used in the experiments are discussed in detail. Then the experimental results and discussion are presented.

Table 1. Sample images in the MSLD and MSID datasets

MSLD				
				
	monkeypox		non-monkeypox	
MSID				
				
	monkeypox	chickenpox	measles	normal

3.1. Parameter Settings

The open-source keras-tensorflow library was employed for designing the deep learning model. The codes are written in Python programming language. Experimental studies were carried out on a computer system equipped with Nvidia 3080 Ti graphics card, 32 GB RAM and Intel i7 processor. The chosen parameters include an image size of 224x244x3 and a batch size of 16. Further parameters encompass a model training duration of 100 epochs, a learning rate (LR) of 0.0001. The optimization process involves the utilization of the Adam optimizer to minimize the loss function, and the LR is dynamically adjusted during training through the application of the ReduceLROnPlateau model in tensorflow. This model ensures a more efficient training regimen by adaptively reducing the LR when the training process reaches a standstill. Categorical Cross Entropy loss function is used for training the proposed DL-AEL model. The Categorical Cross Entropy loss function is a fundamental tool for training and evaluating classification models. The purpose of using this loss function is to measure the difference between the predicted class probabilities and the actual class labels in a multi-class classification problem. In this way, it guides the learning process of the model and helps the model to make better and more accurate predictions.

3.2. Evaluation Metrics

In the experimental studies, the results obtained from the confusion matrix were used to evaluate the performance of the proposed DL-AEL model. The metrics used for evaluation are precision, recall, test accuracy and F1 score. The confusion matrix is a performance measurement tool in deep learning, particularly in classification tasks. It's a table that allows visualization of the performance of an model by displaying the number of true positives (TP), false positives (FP), true negatives (TN), and false

negatives (FN). TP cases occur when the number of instances of the 'Monkeypox' class is correctly identified. TN cases occur when non-Monkeypox (other) cases are correctly identified. FP and FN results are incorrectly predicted results. FP results occur when non-Monkeypox (others) cases are incorrectly identified. FN occur when Monkeypox cases are incorrectly estimated. The models perform well when FP and FN are minimized [27][28]. The evaluation metrics obtained using the confusion matrix are as follows.

Precision: A measure that emphasizes true positive and false positive results. Precision is high when false positive cases are low. It is calculated using Equation 3. *Recall:* A measure that emphasizes true positive and false negative results. Sensitivity is high when false negative cases are low. It is calculated using Equation 4. *F1 score:* A measure that takes into account both precision and sensitivity. It is the harmonic mean of the two measures. It is calculated using Equation 5. *Test Accuracy*: The number of correctly predicted samples among all samples (both monkeypox and non-monkeypox). It is calculated using Equation 6 [29][30].

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 - score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

3.3. Experimental Results and Discussion

In this study, within the scope of the proposed DL-AEL model, binary combinations of Xception, VGG16, DenseNet121, MobileNet, InceptionV3 and ResNet101 network models are used. The MSLD with two classes and MSID with four classes were used in the experimental studies. The classification results obtained in the experimental studies with the MSLD dataset are given in Table 2. When examining Table 2, it can be observed that the individual use of models resulted in the best classification performance with 91.66% accuracy, 91.95% precision, 92.41% recall, and 91.65% F1 score for VGG16, and 88.49% accuracy, 88.92% precision, 89.28% recall, and 88.48% F1 score for ResNet101 models. Considering the combinations of models, the ResNet101 + VGG16 model achieves the highest overall performance with 92.46% accuracy, 92.75% precision, 93.22% recall, and 92.98% F1 score. These results indicate that this model combination is most effective in classifying data points into the relevant two classes in the MSLD dataset. ResNet101+Xception, InceptionV3+VGG16, and ResNet101+InceptionV3 models achieve good performance with accuracy above 90%. This indicates the effectiveness of these models for classification tasks in the MSLD dataset. Table 2 also highlights variations in performance between models. For instance, MobileNet and ResNet101+ MobileNet fall short compared to another models. This could be attributed to various factors. The model's complexity might not be ideal for the size or characteristics of the MSLD dataset. Additionally, the specific architecture of these models might be less suited for the classification task at hand. In addition, combining the VGG16 model, which has a simpler structure than other models, with other models appears to increase performance in terms of accuracy and F1 score.

The classification results obtained in the experimental studies with the MSID dataset are given in Table 3. According to Table 3, it can be observed that the individual use of models resulted in the best classification performance with 96.10% accuracy, 96.11% precision, 92.17% recall, and 93.76% F1 score for Xception, and 96.42% accuracy, 95.23% precision, 95.09% recall, and 95.10% F1 score for DenseNet121 models. Considering the combinations of models, the DenseNet121+Xception model achieves the highest overall performance with 97.58% accuracy, 96.57% precision, 95.74% recall, and 96.14% F1 score. The successful results obtained by DenseNet121 in individual use have also influenced the use of

DenseNet121 in combination with other models. DenseNet121+Xception achieved the most successful result with 97.58% accuracy. In addition, InceptionV3+DenseNet121, MobileNet+ DenseNet121, DenseNet121+VGG16 models achieved better results than other models with accuracy values of 96.66%, 96.47%, 94.99% respectively. Overall, the results indicate that DenseNet121 is a strong performer on its own, and combinations involving it tend to perform well, showcasing its effectiveness in classifying instances in the MSID dataset. ResNet101 and VGG16 models exhibit relatively lower performance in all metrics. In binary combinations, the ResNet101+VGG16 model has the lowest classification results. In summary, Table 3 provides valuable insights into the classification results of various deep learning models and combinations on the MSID dataset, aiding in the selection of the most suitable model or combination for classification tasks on this dataset. As a result, combining models (e.g., DenseNet121+VGG16) improved performance overall compared to individual models, suggesting that ensemble learning techniques provide benefits.

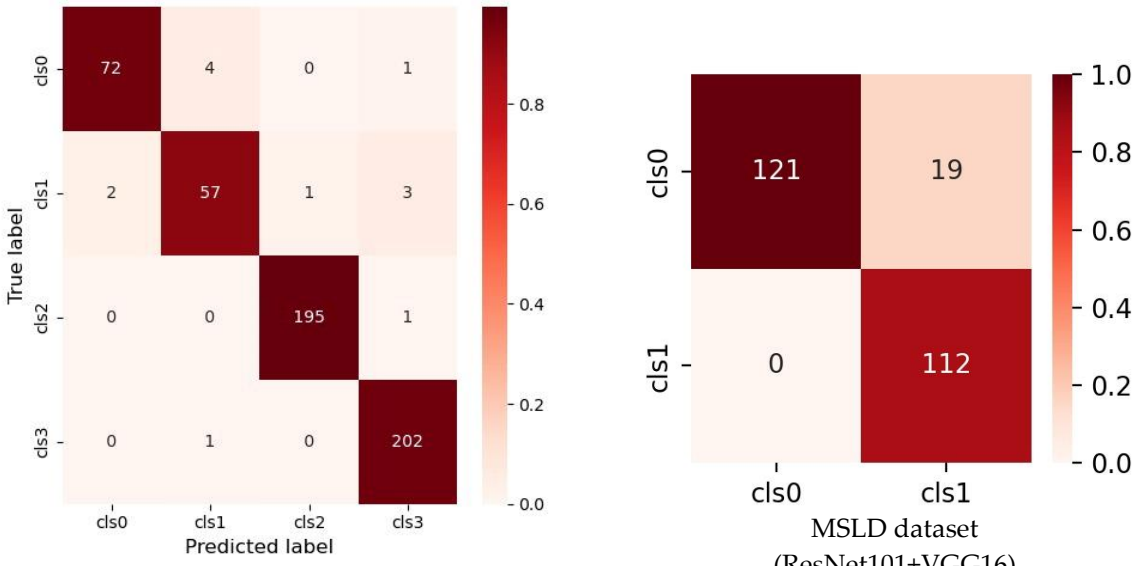
The confusion matrices for the proposed DL-AEL model using both MSID and MSLD datasets are given in Figure 3. The MSID dataset contains four classes. These classes are cls0: chickenpox, cls1: measles, cls2: monkeypox, cls3: normal. The MSLD dataset consists of cls0: monkeypox and cls1: non-monkeypox classes. Firstly, when the confusion matrix of the MSID dataset is analyzed, it is seen that 72 out of 77 chickenpox test images, 57 out of 63 measles, 195 out of 196 monkeypox and finally 202 out of 203 normal patients are correctly classified. In total, 526 out of 539 MSID test images were correctly classified. The test accuracy value obtained in this case is 97.58%. When the confusion matrix obtained using the second dataset, MSLD, is examined, it is seen that 121 out of 140 monkeypox test images and all 112 non-monkeypox images are correctly classified. In total, 233 out of 252 test images were correctly classified. The test accuracy value obtained in this case is 92.46%.

Table 2. Classification results for the MSLD dataset with two classes

Model_Name	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Xception	87.88	88.10	88.55	88.32
VGG16	**91.66**	**91.95**	**92.41**	**91.65**
DenseNet121	87.69	88.68	88.75	87.69
MobileNet	85.31	86.27	86.33	85.31
InceptionV3	88.09	88.19	88.19	88.19
ResNet101	**88.49**	**88.92**	**89.28**	**88.48**
VGG16+Xception	89.28	88.44	89.12	88.78
DenseNet121+Xception	87.69	87.13	86.65	86.89
MobileNet+Xception	88.49	88.54	88.76	88.65
InceptionV3+Xception	89.28	88.04	89.16	88.60
ResNet101+Xception	90.47	90.27	89.81	90.04
DenseNet121+VGG16	88.88	89.10	89.55	89.32
MobileNet+VGG16	88.88	88.46	88.68	88.57
InceptionV3+VGG16	92.06	91.75	91.89	91.82
ResNet101+VGG16	**92.46**	**92.75**	**93.22**	**92.98**
MobileNet+ DenseNet121	87.69	86.42	87.55	86.98
InceptionV3+ DenseNet121	88.88	88.04	88.16	88.10
ResNet101+ DenseNet121	86.90	84.10	85.90	84.99
InceptionV3+ MobileNet	88.88	87.95	87.52	87.73
ResNet101+ MobileNet	82.93	80.66	81.75	81.20
ResNet101+ InceptionV3	90.47	89.28	89.65	89.46

Table 3. Classification results for the MSID dataset with two classes

Model_Name	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Xception	**96.10**	**96.11**	**92.17**	**93.76**
VGG16	91.23	89.91	88.25	88.93
DenseNet121	**96.42**	**95.23**	**95.09**	**95.10**
MobileNet	92.85	92.24	88.99	89.84
InceptionV3	93.50	93.41	90.92	92.05
ResNet101	89.93	87.58	85.37	86.11
VGG16+Xception	94.24	93.28	92.59	92.93
DenseNet121+Xception	**97.58**	**96.57**	**95.74**	**96.14**
MobileNet+Xception	93.50	91.12	91.36	91.24
InceptionV3+Xception	93.87	91.49	92.04	91.76
ResNet101+Xception	93.13	90.89	90.55	90.72
DenseNet121+VGG16	94.99	94.09	94.17	94.13
MobileNet+VGG16	94.80	93.93	94.11	94.02
InceptionV3+VGG16	92.20	90.50	91.05	90.77
ResNet101+VGG16	91.83	90.23	90.91	90.57
MobileNet+ DenseNet121	96.47	95.10	95.27	95.18
InceptionV3+ DenseNet121	96.66	95.41	95.73	95.57
ResNet101+ DenseNet121	93.69	91.15	91.76	91.45
InceptionV3+ MobileNet	94.24	92.74	93.09	92.91
ResNet101+ MobileNet	92.39	91.31	91.93	91.62
ResNet101+ InceptionV3	93.13	92.21	91.89	92.05



MSID dataset
(DenseNet121+Xception)

Figure 3. Confusion matrixs for the proposed DL-AEL model

Table 4 provides a comprehensive comparison of the performance of the proposed DL-AEL model with various studies in the literature on monkeypox disease classification. Our findings indicate that the proposed DL-AEL model outperforms several state-of-the-art models presented in the literature. On the MSLD dataset, the proposed DL-AEL model achieved an accuracy of 92.46%, surpassing the performance of ensemble models such as those proposed by Ali et al. [13] and Haque et al. [14], as well

as individual models like MobileNetv2 (Sahin et al. [15]), EfficientNetB4(Almufareh et al. [16]), and the Siamese Deep Learning Model (Alakus et al. [17]). Furthermore, the DL-AEL model exhibited notable improvements in precision, recall, and F1-score compared to these prior works. When evaluated on the MSID dataset, the proposed DL-AEL model continued to demonstrate its superiority, achieving an impressive accuracy of 97.58%. This performance surpasses that of established models such as EfficientNetB4 (Almufareh et al. [16]), AlexNet (Ural et al. [18]), VGG16 (Ural et al. [18]), CNN+LSTM (Uysal et al. [19]), and ResNet18 (Nayak et al. [7]). Notably, the proposed DL-AEL model exhibited exceptionally high precision, recall, and F1-score, underscoring its effectiveness in accurately classifying medical images from the MSID dataset.

Consequently, the strengths and weaknesses of the proposed Deep Learning-Based Adaptive Ensemble Learning (DL-AEL) model for the detection and classification of monkeypox disease are as follows:

Strengths:

- The DL-AEL model achieved high accuracy, precision, recall, and F1-score on both the MSLD and MSID datasets.
- The model uses dynamic weighting, allowing it to adapt based on the performance of the base models on incoming data. This adaptability improves the overall performance of the ensemble by adjusting the weights of the base models dynamically.
- The model combines multiple deep learning models (e.g., Xception, VGG16, DenseNet121, MobileNet, InceptionV3, and ResNet101) to leverage their strengths. This combination has been shown to enhance classification performance compared to individual models.
- The FC layer and backpropagation policy are used effectively to learn the weights in the ensemble, ensuring that the most suitable weights are applied during testing.
- Comparative studies with other models in the literature demonstrate that the proposed DL-AEL model outperforms state-of-the-art models in terms of accuracy and other performance metrics.

Weaknesses:

- The performance of the ensemble model heavily relies on the quality and representativeness of the training data. Any biases or inaccuracies in the training data could adversely affect the model's performance.
- Training multiple deep learning models and then combining them into an ensemble can be computationally intensive and time-consuming. This complexity might require significant computational resources, which could be a limitation for some applications.
- In the study, publicly available MSLD and MSID datasets, which are frequently used in the literature, are used for training and testing. The performance of the proposed model in real-time applications remains uncertain due to the difficulties in generating datasets.

Table 4. Comparison with different studies in the literature

Literature	Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Ali et al. [13]	Ensemble Model (VGG16, ResNet50, and InceptionV3)	MSLD	79.26	84.00	79.00	81.00
Sahin et al. [15]	MobileNetv2	MSLD	91.11	90.00	90.00	90.00
Haque et al. [14]	Xception-CBAM-Dense layer	MSLD	83.89	90.70	89.10	90.11
Almufareh et al. [16]	EfficientNetB4	MSLD	88.89	90.02	88.00	89.00
Alakus et al. [17]	Siamese Deep Learning Model	MSLD	91.09	89.03	93.36	91.14
Almufareh et al. [16]	EfficientNetB4	MSID	92.00	89.81	95.23	92.44
Ural et al. [18]	AlexNet	MSID	71.00	70.75	67.75	68.00
	VGG16		80.00	78.75	74.5	74.25
Uysal et al. [19]	CNN+LSTM	MSID	87.00	93.00	87.00	90.00
Nayak et al. [7]	ResNet18	MSID	91.11	94.72	90.43	92.55
Proposed model	**DL-AEL**	**MSLD**	**92.46**	**92.75**	**93.22**	**92.98**
		MSID	**97.58**	**96.57**	**95.74**	**96.14**

4. CONCLUSIONS

Monkeypox, a lethal illness, has recently extended its reach to numerous nations. It manifests through distinctive skin eruptions, a feature readily identifiable through imaging techniques. Thus, our study proposes a novel deep learning model designed to differentiate Monkeypox from similar dermatological conditions like chickenpox and measles. The proposed model is called deep learning based adaptive ensemble learning (DL-AEL) model. The proposed DL-AEL model is realized in two stages. In the first stage, two different pre-trained models are trained with training datasets. In the second stage, the outputs of the two models are concatenated by removing the softmax layer that obtains the full probabilistic value of the trained models. Then a fully connected layer is applied to the concatenated outputs. In this way, the output values of the model are obtained. Finally, a softmax layer is applied to obtain the probabilistic distribution of this output. In the proposed DL-AEL model, binary combinations of Xception, VGG16, DenseNet121, MobileNet, InceptionV3 and ResNet101 network models are used. Two different monkeypox datasets were used for the performance analysis of the models obtained with the binary combination. The first dataset is MSLD with two classes (monkeypox or non-monkeypox), while the other dataset is MSID with four classes (monkeypox, chickenpox, measles and normal). As a result of the experimental studies, the best classification performance results with the MSLD dataset were obtained in the ResNet101+VGG16 model with 92.46% accuracy, 92.75% precision, 93.22% recall and 92.98% F1 score, while 97.58% accuracy, 96.57% precision, 95.74% recall and 96.14% F1 score were found in the DenseNet121+Xception model with the MSID dataset. Comprehensive comparisons with different studies from the literature using the same data set suggest that the proposed

DL-AEL model is successful. In future studies, the primary focus will be on expanding the dataset. Additionally, efforts will be made to develop different deep learning models capable of real-time diagnosis from clinical images or patient symptoms.

DECLARATION OF ETHICAL STANDARDS

The paper is conducted in accordance with ethical standards.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

DECLARATION OF COMPETING INTEREST

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

FUNDING / ACKNOWLEDGEMENTS

No dedicated funding was received for this research from commercial, public or non-profit entities.

DATA AVAILABILITY

Monkeypox Skin Lesion Dataset (MSLD): https://www.kaggle.com/datasets/nafin59/monkeypox-skin-lesion-dataset

Monkeypox Skin Images Dataset (MSID): https://data.mendeley.com/datasets/r9bfpnvyxr/6

5. REFERENCES

[1] A. Gessain, E. Nakoune, and Y. Yazdanpanah, "Monkeypox," N. Engl. J. Med., vol. 387, no. 19, pp. 1783–1793, 2022.

[2] M. E. Wilson, J. M. Hughes, A. M. McCollum, and I. K. Damon, "Human monkeypox," Clin. Infect. Dis., vol. 58, no. 2, pp. 260–267, 2014.

[3] B. L. Ligon, "Monkeypox: A review of the history and emergence in the Western hemisphere," Semin. Pediatr. Infect. Dis., vol. 15, no. 4, pp. 280–287, 2004.

[4] K. Chadaga et al., "Application of Artificial Intelligence Techniques for Monkeypox: A Systematic Review," Diagnostics, vol. 13, no. 5, pp. 1–16, 2023.

[5] World Health Organization, "Multi-country monkeypox outbreak in non-endemic countries," 2022. [Online]. Available: https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON385. [Accessed: 01-Apr-2024].

[6] K. L. Koenig, C. K. Beÿ, and A. M. Marty, "Monkeypox 2022 Identify-Isolate-Inform: A 3I Tool for frontline clinicians for a zoonosis with escalating human community transmission," One Heal., vol. 15, no. June, p. 100410, 2022.

[7] T. Nayak et al., "Detection of Monkeypox from skin lesion images using deep learning networks and explainable artificial intelligence," Appl. Math. Sci. Eng., vol. 31, no. 1, 2023.

[8] R. Pramanik, B. Banerjee, G. Efimenko, D. Kaplun, and R. Sarkar, "Monkeypox detection from skin lesion images using an amalgamation of CNN models aided with Beta function-based normalization scheme," PLoS One, vol. 18, no. 4 April, pp. 1–21, 2023.

[9] A. Bohr and K. Memarzadeh, The rise of artificial intelligence in healthcare applications. INC,

2020.
[10] M. A. Myszczynska et al., "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," Nat. Rev. Neurol., vol. 16, no. 8, pp. 440–456, 2020.
[11] H. Firat and H. Üzen, "Detection of Pneumonia Using A Hybrid Approach Consisting of MobileNetV2 and Squeeze- and-Excitation Network," Turkish J. Nat. Sci., vol. 13, no. 1, pp. 54–61, 2024.
[12] M. F. Ozdemır and D. Hanbay, "A Novel Covid-19 Detection System Based on PSO and Hybrid Feature Using Support Vector Machines," Comput. Sci., vol. IDAP-2022, pp. 120–129, 2022.
[13] S. N. Ali et al., "Monkeypox Skin Lesion Detection Using Deep Learning Models: A Feasibility Study," arXiv Prepr. arXiv2207.03342, 2022.
[14] E. H. Md, R. A. Md, S. N. Razia, and I. Salekul, "Classification of Human Monkeypox Disease Using Deep Learning Models and Attention Mechanisms," arXiv:2211.15459, 2022.
[15] V. H. Sahin, I. Oztel, and G. Yolcu Oztel, "Human Monkeypox Classification from Skin Lesion Images with Deep Pre-trained Network using Mobile Application," J. Med. Syst., vol. 46, no. 11, 2022.
[16] M. F. Almufareh, S. Tehsin, M. Humayun, and S. Kausar, "A Transfer Learning Approach for Clinical Detection Support of Monkeypox Skin Lesions," Diagnostics, vol. 13, no. 8, pp. 1–16, 2023.
[17] T. B. Alakuş, "Prediction of Monkeypox on the Skin Lesion with the Siamese Deep Learning Model," Balk. J. Electr. Comput. Eng., vol. 11, no. 3, pp. 225–231, 2023.
[18] A. B. Ural, "A Computer-Aided Feasibility Implementation to Detect Monkeypox from Digital Skin Images with Using Deep Artificial Intelligence Methods," Trait. du Signal, vol. 40, no. 1, pp. 383–388, 2023.
[19] F. Uysal, "Detection of Monkeypox Disease from Human Skin Images with a Hybrid Deep Learning Model," Diagnostics, vol. 13, no. 10, 2023.
[20] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 1800–1807, 2017.
[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14, 2015.
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-Decem, pp. 770–778.
[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 2818–2826, 2016.
[24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 2261–2269, 2017.
[25] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.
[26] D. Bala, "Monkeypox Skin Images Dataset (MSID)," mendeley data, 2022. [Online]. Available: https://data.mendeley.com/datasets/r9bfpnvyxr/6. [Accessed: 15-Mar-2024].
[27] A. S. Jaradat et al., "Automated Monkeypox Skin Lesion Detection Using Deep Learning and Transfer Learning Techniques," Int. J. Environ. Res. Public Health, vol. 20, no. 5, 2023.
[28] H. Fırat and H. Üzen, "Transfer Öğrenme Kullanılarak Derï Lezyon Görüntülerïnden Maymun Çïçeğï Hastalığının Tespïtï," Adıyaman Üniversitesi Mühendislik Bilim. Derg., vol. 11, no. 22, pp. 148–164, 2024.
[29] S. Surati, H. Trivedi, B. Shrimali, C. Bhatt, and C. M. Travieso-González, "An Enhanced Diagnosis of Monkeypox Disease Using Deep Learning and a Novel Attention Model Senet on Diversified Dataset," Multimodal Technol. Interact., vol. 7, no. 8, 2023.

[30] A. Sorayaie Azar, A. Naemi, S. Babaei Rikan, J. Bagherzadeh Mohasefi, H. Pirnejad, and U. K. Wiil, "Monkeypox detection using deep neural networks," BMC Infect. Dis., vol. 23, no. 1, pp. 1–13, 2023.