Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19862-w

SkinMarkNet: an automated approach for prediction of monkeyPox using image data augmentation with deep ensemble learning models

Aqsa Akram[1] · Arwa A. Jamjoom[2] · Nisreen Innab[3] · Nouf Abdullah Almujally[4] · Muhammad Umer[5] · Shtwai Alsubai[6] · Gianluca Fimiani[7]

Received: 3 May 2024 / Revised: 21 June 2024 / Accepted: 13 July 2024

Abstract
Monkeypox, a rare but potentially fatal viral disease, poses a significant public health challenge due to its potential for outbreaks and complications. Detecting monkeypox lesions early and accurately is vital to effectively manage and control the disease. This study introduces a novel method for classifying monkeypox lesions, employing data augmentation methods and a framework based on an ensemble of three transfer learning models called "SkinMarkNet". The dataset used in this research consists of skin lesion images collected from the Kaggle data repository, encompassing diverse demographics and lesion characteristics. This research uses image data augmentation techniques to tackle the scarcity of annotated data. This augmentation enriches the training dataset, thereby improving the model's ability to perform effectively. Moreover, the novelty of this research work lies in the usage of three popular transfer learning models(Inception, Xception, and ResNet) for feature extraction and ensemble learning. The SkinMarkNet achieves promising results showing an accuracy of 90.615% for monkeypox lesion classification, outperforming traditional machine learning and deep learning methods utilized in recent research works. In addition, thorough comparative analysis is done with machine learning models and contemporary approaches to validate the efficacy of the proposed method. Overall, the findings underscore the potential of leveraging advanced deep learning architectures and data augmentation strategies for improving monkeypox lesion classification, thereby facilitating early diagnosis and intervention in clinical settings and public health surveillance efforts.

Keywords Bioinformatic · Technology in healthcare · Biomedical image data · Mpox detection · SkinMarkNet · Skin lesion detection

1 Introduction

Monkeypox, a member of the Orthopoxvirus genus and the Poxviridae family is a zoonotic disease. Apparently, it is similar to other skin-related diseases like chickenpox, smallpox, and measles [1]. Monkeys, Rodents, and other wild animals are the main sources of transfer of this disease in humans. Humans become infected with monkeypox when they come into

Extended author information available on the last page of the article

contact with infected animals through hunting, handling, or consumption of infected animal meat. However, human-to-human transmission of monkeypox can also occur through close contact with infected individuals. This happens when a person comes into contact with bodily fluids or lesion material of an infected person [2]. This disease is more popular among men rather than women. According to the World Health Organization (WHO), 53% male cases are reported and 47% female cases are reported till now. Same in the case of age, monkeypox targets all age groups but it mainly affects 21-30% of age group [3]. The virus now known as monkeypox was first recognized in a monkey's body in a laboratory in Copenhagen, Denmark in 1958 [4]. The first case of monkeypox among humans was recorded during a campaign to remove smallpox in the Democratic Republic of the Congo in 1970 [5]. Monkeypox naturally occurs in individuals living in or near humid rainforests in central and western Africa. This disease is transferred through many resources, including direct contact with infected animals, animal bites, respiratory droplets, and contact with bodily fluids such as those found in the eyes, nose, or mouth [6].

The outbreak of MonkeyPox has underscored the urgent need for effective diagnostic tools to manage and contain its spread. Accurate and timely identification of MonkeyPox lesions is critical for preventing outbreaks and ensuring appropriate medical intervention. Traditional diagnostic methods, though effective, are often labor-intensive and time-consuming. Recent advancements in artificial intelligence, particularly deep learning, have demonstrated significant potential in automating and enhancing diagnostic processes in dermatology [7, 8].

The primary treatment for monkeypox is vaccination; however, although FDA-approved vaccinations are readily available, human use of them has not occurred in the United States yet. Monkeypox is treated in other nations with vaccines meant to treat smallpox, such as Tecovirimat, VIGIV, Cidofovir, and Brincidofovir [9]. Polymerase chain reaction (PCR) is used to identify the monkeypox virus in humans [10]. For the classification of biomedical pictures, numerous machine learning (ML) and deep learning (DL) algorithms are now in use. Because of their increased capacity to learn and analyze complex data, the most recent developments in deep learning, especially the multiple Convolutional Neural Network (CNN) algorithms, have brought new developments to several fields of medical science [11].

Deep learning, mainly the Convolutional Neural Network (CNN) model has presented unlimited potential in the field of medicine. Stimulated by the structure and function of the brain, artificial neurons in CNN take input, process it, and generate output, similar to the human brain. In CNN, the input is the image pixels, and its hidden layers perform various operations such as convolution, pooling, rectified linear units, and fully connected layers to extract features from the image. Since CNN is a feed-forward network, information flows only in one direction. Due to limited datasets for monkeypox classification, a machine learning approach called transfer learning is used to overcome this challenge. This technique uses a large labeled dataset to train a deep neural network model, which can then be used as a feature extractor for smaller datasets with limited labels [12].

Pre-trained models such as Inception-v3, Xception, and ResNet50 are commonly considered due to their exceptional performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). These models employ deep learning techniques, such as additional classifiers, mapping identifiers, and smoothing using Rectified Linear Units (ReLu) [13]. To improve the accuracy of image classification for monkeypox, the individual pre-trained models including Inception-v3, Xception, and ResNet50 were trained on the dataset. Each of these models produced results for image classification. However, further enhancement in the classification accuracy is done through ensemble method using all the trained models. This ensemble method involves combining the results of each model and adding extra hidden lay-

ers to obtain a final prediction. The proposed ensemble method produced satisfactory results, indicating that the combined output of multiple models can lead to better performance in image classification tasks [14].

In this context, we introduce "SkinMarkNet" This innovative approach leverages the power of image data augmentation to enhance the training dataset, thereby improving the robustness and accuracy of the predictive models. By employing a deep ensemble learning strategy, SkinMarkNet integrates multiple deep learning architectures to capitalize on their strengths and mitigate their weaknesses, resulting in superior predictive performance [15, 16]. This method not only accelerates the diagnostic process but also provides a scalable and reliable solution for MonkeyPox detection, addressing a critical gap in current healthcare practices. Below are the main objectives of this proposed study:

1. Proposed a novel ensemble model "SkinMarkNet" incorporating the novelty of data augmentation and extra hidden layers in achieving high accuracy for the classification of chickenpox, monkeypox, measles, and normal skin marks.
2. The performance of the proposed model is compared with a light gradient boosting machine (lgbm), random forest classifier (rfc), logistic regression (lr), and extra tree classifier (etc).
3. To present the first study to achieve such precision in differentiating among these specific skin-related conditions.
4. Addressed a critical research gap by focusing on accurately detecting monkeypox, unlike other studies that typically concentrate on healthy and unhealthy skin.

Furthermore, the paper is divided as follows: Section 2 explains all past work related to this research. Section 3 explains how we conducted the experiment, and Section 4 discusses the results. In Section 6, we conclude this proposed research work, and in Section 5 we describe the limitations of this work and suggest possible directions for future research.

2 Related work

Monkeypox, a significant public health concern in West and Central Africa, extended its impact to the United States in 2003. The outbreak originated from contact with infected dogs housed with imported dormice and Gambian pouched rats. Over 70 cases were recorded in the US. Nigerian tourists contracted monkeypox in Israel, the UK, Singapore, and the US on various occasions. In May 2022, monkeypox cases emerged in previously unaffected countries. Ongoing studies aim to understand the disease's epidemiology, vectors, and transmission dynamics for effective control [13]. Previous studies focused on the classification of monkeypox disease using deep learning techniques. To enhance the dataset, they collect skin lesion images from open-source websites and employ data augmentation techniques along with 3-k-fold cross-validation. The researchers evaluate the performance of pre-trained models, including VGG-16, ResNet50, and InceptionV3, for the classification of monkeypox and other ailments. Among these models, ResNet50 demonstrates the highest accuracy rate, making it the model of choice for further analysis. ResNet50 achieves an average accuracy of 82.96%, while VGG-16 achieves 81.48%. By ensembling these three models, the authors attain an accuracy of 79.26% [14].

The study incorporated the Xception transfer learning model and employed additional techniques such as Grad-Cam and LIME. Furthermore, a community approach was developed using both the Xception model and the DenseNet model. By leveraging this proposed ensemble approach, the researchers evaluated the performance scores on a publicly available

dataset. After careful analysis of the experimental outcomes, the study achieved commendable results with an average precision of 85.44%, a recall of 85.47%, an F1 score of 85.40%, and an accuracy of 87.13%. These metrics demonstrate the effectiveness of the ensemble method and highlight its ability to accurately classify monkeypox and distinguish it from other diseases [17].

In another research study, a thorough evaluation and initial probability assessment of five pre-trained deep learning methods were conducted, along with the integration of a convolutional block attention mechanism (CBAM) and dense layers. The primary objective was to categorize monkeypox infection accurately using skin lesions as the basis. The Monkeypox Skin Lesion Dataset (MSLD) is employed for training and evaluating the models. By integrating the CBAM module, which incorporates both channel and spatial attention, the projected system effectively emphasizes significant feature maps and concentrates on inter-channel needs within the affected sections of the image. Their projected model demonstrates promising classification performance, achieving a validation accuracy of 83.89%, utilizing the Xception-CBAM-Dense layer architecture [18].

This study involved the creation of the Monekypox2022 image dataset, followed by the proposal of an altered VGG16 model through two distinct experiments. The recommended altered VGG16 model achieved an accuracy of 97% in the first study, and 88% in the second study, particularly during classifying monkeypox patients. The researchers also discussed the application of Local Interpretable Model-Agnostic Explanations (LIME) for feature extraction and prediction with their model [19].

This Study conducted an investigation using the MSLD dataset, where they examined the performance of various pre-trained deep learning models. Their findings indicated that the MobileNetV2 and EfficientNetB0 models delivered noteworthy results. Building upon these outcomes, the researchers proceeded to develop a mobile application for the categorization of monkeypox disease. To ensure efficient deployment, they converted the whole model into a TensorFlow lite model, which was further enhanced with metadata [20].

2.1 Most relevant state of the art related works

Monkeypox Detection Using Transfer Learning on Convolutional Neural Networks
Researchers have explored the use of transfer learning on pre-trained convolutional neural networks (CNNs) to classify monkeypox lesions from skin images [21]. The study utilized various pre-trained models like VGG16, ResNet50, and InceptionV3, fine-tuning them on a monkeypox image dataset. The results demonstrated that transfer learning can effectively identify monkeypox with high accuracy, highlighting the potential of CNNs in dermatological disease classification.

Automated Diagnosis of Monkeypox from Clinical Images Using Deep Learning
This research focused on developing a deep learning-based automated diagnostic tool for monkeypox [22]. The authors designed a deep CNN architecture specifically tailored for monkeypox detection. By training the model on a labeled dataset of monkeypox and other skin conditions, the study achieved high sensitivity and specificity, proving the efficacy of deep learning in distinguishing monkeypox from similar diseases.

Ensemble Learning Techniques for Monkeypox Disease Classification
This paper investigates the use of ensemble learning techniques, such as Random Forest, Gradient Boosting Machines, and Voting Classifiers, to improve the accuracy of monkeypox diagnosis [23]. By combining the predictions of multiple classifiers, the ensemble approach

showed improved performance over individual models, suggesting that ensemble methods can enhance the robustness and reliability of monkeypox classification systems.

Using Deep Learning to Enhance Early Detection of Monkeypox Outbreaks
In this study, researchers employed recurrent neural networks (RNNs) and long short-term memory (LSTM) networks to analyze time-series data related to monkeypox outbreaks [24]. By integrating epidemiological data and clinical reports, the deep learning models could predict potential outbreak hotspots, enabling timely interventions and better resource allocation.

Machine Learning Approaches for the Classification of Monkeypox and Other Poxviruses
This article presents a comparative study of various machine learning algorithms, including Support Vector Machines (SVM), k-nearest Neighbors (k-NN), and Decision Trees, for the classification of monkeypox and other poxviruses [25]. The study provided insights into the strengths and limitations of each algorithm, concluding that SVMs with radial basis function kernels achieved the best performance in distinguishing monkeypox from other poxvirus infections.

3 Materials and methods

This section presents the source of the dataset, techniques employed to augment the dataset, explains the utilized transfer learning models, a detailed explanation of the proposed methodology accompanied by a figure, individual evaluations of the models by using matrices, and the subsequent evaluation of the proposed "SkinMarkNet" model.

3.1 Dataset

The dataset for this study on monkeypox skin lesion detection was sourced from Kaggle. Specifically, we accessed the monkeypox dataset repository, which includes a collection of biomedical images related to monkeypox disease [26]. We acknowledge and cite the repository owner for making this dataset publicly available on Kaggle. The dataset includes four types of classes chickenpox, measles, monkeypox, and normal. Each class contains a different number of images. The dataset contains a total of 770 images. First of all the dataset was preprocessed to ensure consistency and remove any irrelevant or corrupted images, adjusting the size of the images, balancing images in each class, making sure the pixel values are in a standard range 224×224 and in RGB format, and even adding some variations to the images to make the dataset more diverse. All these steps were aimed at getting the image data ready for training my transfer learning models (Tables 1 and 2).

Table 1 Details of the dataset

Dataset	Details
Dataset Source	https://www.kaggle.com/datasets/dipuiucse/monkeypoxskinimagedataset
Dataset name	Monkeypox skin image dataset (MSID)
Availability	This dataset is available publically
Collection source	This dataset is collected from internet-based sources
Total classes	This dataset contains 4 classes chicken pox, measles, monkeypox, and normal
Total images	770

Table 2 Splitting dataset

Data Split	Data Split Values
Training data	539
Validation data	154
Testing data	77
Total	770

The preprocessed dataset is split into three subsets in the following steps: training, validation, and testing. Approximately 70% of the dataset is set aside for training, 20% for validation, and 10% for testing, as per the amended split ratio. The model is trained on the training set; the validation set helps with hyper-parameter tuning and performance evaluation; and the testing set is used to test the performance of the finished model.

As we already stated we have a limited dataset of monkeypox disease and to overcome the shortage of data, we expand the raw data using a data generator. In this step, data augmentation was implemented using the ImageDataGenerator class from the Keras library. The dataset exhibited class imbalance, with measles and chickenpox classes having fewer images compared to monkeypox and normal classes. To address this, data augmentation techniques were applied only to the measles and chickenpox classes. By augmenting these two classes, additional variations were created, increasing their sample size and balancing the dataset. The augmented images along with the original images were used to create a more representative and balanced dataset for training the model. Now we have 1172 images after data augmentation. Table 3 represents augmented data. This combination gave us two sets of augmented data, making the utilized dataset more diverse and better equipped for the training of machine and transfer learning models effectively. Figure 1 represents all 4 classes of the dataset.

3.2 Machine learning models for monkeypox classification

In this section, we provide a brief description of the machine learning classification methods utilized in the study, along with details on the calibration process of the classifier. The following machine learning techniques were employed in the current study, chosen for their widespread use in classification tasks. All models were implemented using scikit-learn.

3.2.1 Light gradient boosting machine

LightGBM is a swift, efficient, and distributed gradient-boosting structure DT-embedded algorithm. LightGBM is extensively employed in boosting algorithms across multiple ML tasks like classification, ranking, and regression [27, 28]. Boosting methods yield a powerful

Table 3 Class-wise details of original and augmented dataset

Classes	Original Data	Augmented Data
Chickenpox	107	300
Monkeypox	279	279
Measles	91	300
Normal	293	293
Total	770	1172

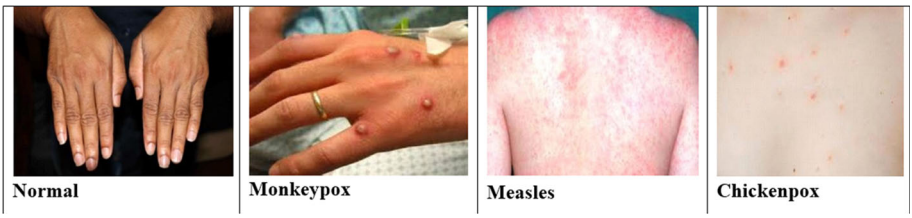

Fig. 1 Dataset examples of multi-class images

learning model by combining multiple weak ML algorithms. Through successive iterations, the importance of misclassified data points is enhanced while that of correctly classified diminished by such boosting algorithms. A greater intention is ensured by this iterative process to the misclassified classifier in the subsequent training sessions. Eventually, all individual ML models are linearly united, with adjustments made to the combined model weights based on the classifiers' error rates.

3.2.2 Random forest classifier

Being an ensemble learning procedure, RF is very proficient at both regression and classification tasks [29, 30]. It harnesses the collective strength of numerous decision trees alongside a technique called Bootstrap and Aggregation, or bagging. This approach involves randomly selecting rows and features from the dataset to generate sample datasets for each tree a process known as Bootstrap. Subsequently, the Aggregation step consolidates the predictions of all individual trees to yield the ultimate results. While RF constructs multiple decision trees and averages their predictions, methods like gradient boosting (GB) and XGBoost build models sequentially to rectify the errors of preceding models. RF demonstrates proficiency with unseen data, exhibits reduced susceptibility to overfitting, and maintains computational efficiency.

3.2.3 Logistic regression

LR is a performance-enhancing combined multiple RF and AdaBoost-like classifiers type of ensemble learning using an iterative ensemble approach. Consequently, a strong classifier is constructed by it [31, 32]. LR identifies the correlation between the categorical dependent and various independent variables. Additionally, it computes the posterior probability p by fitting the data into the logistic function of an event. The classifier's weights and sample training are fixed in each successive iteration aimed at boosting the basic underlying concept to correctly ascertain the target class of the provided data. LR's classification y^* is illustrated as follows

$$y^* = ln(\frac{p}{1-p}) \tag{1}$$

3.2.4 Extra trees classifier

Extremely randomized trees or extra trees, are part of the ensemble learning procedures category, similar to RF, where multiple individual DT results are aggregated [33, 34]. ETs are superior in performance in comparison to RF algorithms. The baseline difference between ET Regressor and RF is the utilization of bootstrap aggression, which is used by RF while

ET doesn't. Instead, it uses the entire training dataset to construct its DTs. ET Regressor, instead of determining the best-split point after all features are taken into consideration, selects features' subset randomly, and eventually a random split point is selected. Overfitting in the model is mitigated by reducing the variance aided by this added randomness. The benefits of ET regressor are proven when datasets are high-dimensional and computational efficiency is a priority.

3.3 Transfer learning models

The simple indication behind transfer learning is to train a deep neural network on a large dataset, and then use this pre-trained model as a feature extractor for a smaller dataset with limited labeled data [35]. By doing this, we speed up the recognition process and achieve faster progress when modeling the new and different tasks. Moreover, since the pre-trained model has already learned many features that are beneficial for various tasks, we reduce the number of parameters needed to train the model and only train the last few layers of the model on the smaller dataset. This research aims to progress the accuracy and efficiency of image classification and we achieve this using an ensemble method. Specifically, we train three of the latest CNN models individually on the dataset and evaluate their accuracy on testing data. After examining the accuracy of each model, we combine them to create a more accurate and efficient ensemble model.

In this study, we used three pre-trained models, namely Inception-v3, Xception, and ResNet50, as feature extractors in the ensemble model. The idea was to leverage the strengths of each model to improve the overall performance of the ensemble model. The hyperparameter details of all transfer learning models are shown in Table 4.

Table 4 Hyperparameters of Inception, ResNet, Xception, and Ensemble via Voting Classifier

Model	Hyperparameter	Value
Inception	Learning Rate	0.001
	Batch Size	32
	Optimizer	Adam
	Epochs	50
	Dropout Rate	0.5
ResNet	Learning Rate	0.0001
	Batch Size	64
	Optimizer	SGD
	Epochs	100
	Weight Decay	0.0005
Xception	Learning Rate	0.0002
	Batch Size	32
	Optimizer	RMSprop
	Epochs	75
	Dropout Rate	0.3
Voting Classifier (Ensemble)	Classifiers	Inception, ResNet, Xception
	Voting Type	Soft
	Weights	[0.33, 0.33, 0.34]

3.3.1 Individual models

Inception-v3: Inception-v3 builds upon the foundations of Inception-v2 and introduces various enhancements to achieve superior efficiency in image recognition tasks. It employs a strategy of factorizing 55 convolutions into two smaller 33 convolutions, leading to faster calculation [36].

3.3.2 ResNet50

Residual learning, a key concept in ResNet, aims to address these challenges by learning residual mappings. ResNet models typically utilize ReLU activations. ResNet50, a variant of Residual Network, comprises 50 layers [37].

3.3.3 Xception model

The Xception model is a CNN architecture that gained recognition as an "extreme inception" approach with a total of 36 convolutional layers. The motivation behind selecting the Xception procedure stems from its exceptional effectiveness and impressive performance documented in prior research [38, 39] in addressing the COVID-19 detection problem, leveraging its proven capabilities in this domain helped us in this study.

3.4 Proposed ensemble model (SkinMarkNet)

In this proposed methodology, we have integrated the feature representations of three distinct deep convolutional neural networks (dCNNs) Inception-v3, Xception, and ResNet50. Each of these models has been trained and evaluated independently on the above-mentioned dataset using standard hyperparameters, including a batch size of 32 and a learning rate of 1×10^{-4}, over 100 epochs. However, we have taken a step further by creating an ensemble model that combines the strengths of all three models. By merging the feature representations extracted from Inception-v3, Xception, and ResNet50. The ensemble model aims to capture a wide range of image characteristics and patterns. This ensemble model harnesses the unique capabilities of each model in recognizing different aspects of the images, such as intricate details, complex structures, and global context. To construct the ensemble model, we aggregate the predictions from each model and make a final decision based on a combination of their outputs by using a concatenation ensemble approach. To further optimize results and attain high accuracy, and uniqueness of this research, additional hidden layers were introduced. These hidden layers, unlike those found in previous research, intricately analyze images. This modification enhances the ensemble model, named SkinMarkNet enabling it to easily and accurately differentiate among the four classes of diseases, achieving notable success in obtaining high accuracy The high accuracy attained by the proposed model is a result of the collective intelligence and complementary strengths of the constituent models. By adopting this ensemble approach, we aim to boost the whole performance and robustness of the classification system. By integrating Inception-v3, Xception, and ResNet50 models, we capitalize on their distinct advantages and strive to attain superior accuracy in this image classification task compared to using any single model in isolation. Figure 2 represents the proposed methodology.

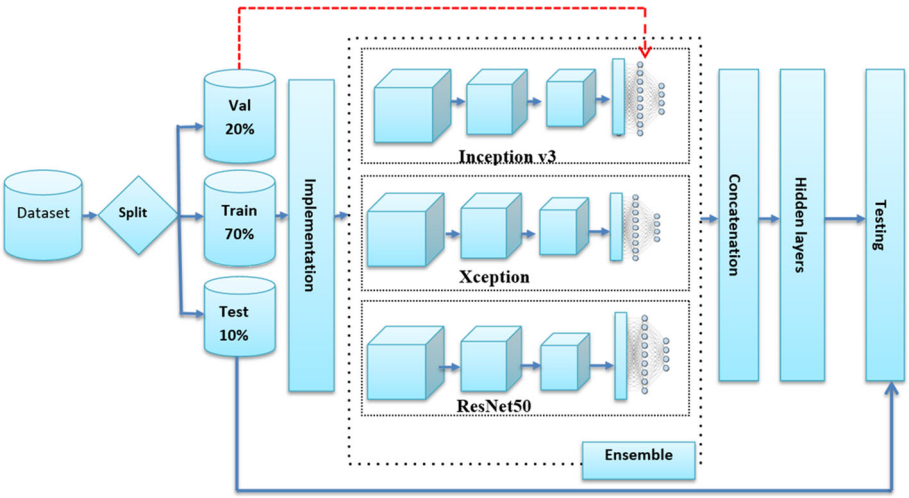

Fig. 2 Proposed Methodology

3.5 Evaluation parameters

The evaluation of the deep learning models encompasses metrics such as recall, precision, F1 score, and accuracy. Given the binary classification nature of this study, the assessment also involves the use of a confusion matrix. This matrix serves as a tabular tool for delineating the model's performance in classifying test data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Precision assesses the ratio of correctly predicted positive instances to the total instances predicted as positive. The highest achievable precision score for a model is 1, while the lowest is 0.

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

Recall, also known as true positive rate (TPR) or sensitivity, signifies the classifier's capability to correctly identify all positive samples. It is computed as the ratio of TP (True Positives) divided by the sum of TP and FN (False Negatives). The highest possible recall score for a model is 1, with a minimum score of 0.

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

The F1 score, sometimes referred to as the F Measure, represents the trade-off between precision and recall. It illustrates the balance between these two metrics, with a model achieving a maximum F1 score of 1 and a minimum of 0.

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

Table 5 System configuration for models experimentation

Parameters	Configuration
Processor	Intel(R) Core(TM) i5-5300U
Clock speed	2.30 GHz
RAM	8.00 GB
GPU	Intel® HD Graphics 5500
Tools	Gpu and Google Colab
Language	Python3

4 Results and experiments

This proposed methodology integrates feature representations from three CNN models Inception- v3, Xception, and ResNet50. Each model is trained and evaluated independently on the dataset. We create an ensemble model by merging their feature representations to capture a wide range of image characteristics and patterns. The ensemble model combines the strengths of each model to achieve higher overall accuracy. We aggregate their predictions through a concatenation ensemble approach to make the final classification decision. The addition of extra hidden layers makes this ensemble model more unique and effective. This ensemble approach enhances performance and robustness compared to using a single model. The experiment is conducted on the system mentioned in Table 5.

The results obtained from these models demonstrated their effectiveness in tackling the image recognition task on the monkeypox dataset. The models showcased high accuracy, precision, recall, and F1 scores, indicating their ability to correctly classify and recognize images. The experiments confirmed the suitability of these models on this dataset and highlighted their potential for accurate image classification tasks. We used 100 epochs to train the model and to improve accuracy. The results of this study showcase the performance of this proposed ensemble model paralleled to each model individually trained and evaluated on the same dataset

Table 6 shows that the individual models, ResNet50, Inception-v3, and Xception, achieved accuracies of 36-61%, 82-85%, and 82-84% respectively. While LGBM, RFC, LR, ETC performance is much better when compared with individual transfer learning models. The best accuracy is obtained by ETC which is 75%. The second best-performing machine learning model is RFC with 72% of accuracy. However, the ensemble model 'SkinMarkNet" demonstrated significantly improved performance with an accuracy of 89- 90%, showcasing

Table 6 Comparison of all transfer learning, machine learning, and proposed model results (in %)

Model	Accuracy	Recall	Precision	F1-Score
LGBM	64	57	61	59
RFC	72	65	67	66
LR	68	63	63	63
ETC	75	68	73	71
ResNet50	61	42	38	36
Inception-v3	84	84	85	84
Xception	83	82	83	82
SkinMarkNet	90	89	90	89

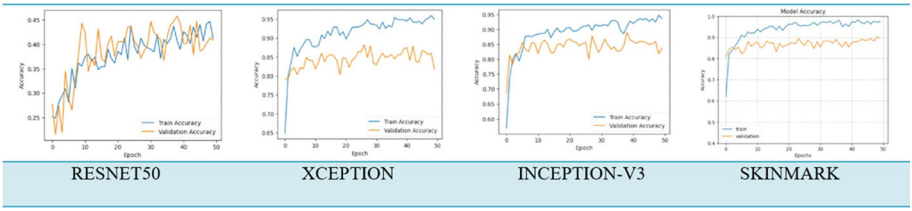

Fig. 3 Proposed model training and validation accuracy

its effectiveness in enhancing classification accuracy. ResNet50 doesn't perform as well as Xception, InceptionV3, and the ensemble model and this could be because ResNet50 has a simpler design. Its simplicity might make it struggle to understand and capture detailed features in the pictures. In comparison, the other models and the combined ensemble model seem to do a better job in this regard.

Figure 3 illustrates the accuracy of models. The accuracy of the proposed model was initially low at epoch 1, but it steadily increased as the epochs progressed. This trend suggests that the model was effectively trained and demonstrated improved performance on the testing dataset as the accuracy increased over time.

Figure 4 shows the training and testing loss models. The models exhibited high training and testing loss initially at epoch 1, which gradually decreased as the number of epochs increased. This trend indicates that the models were effectively trained and performed well on the testing dataset, as the loss decreased over time. In this case, the ensemble model stands out with the highest average values, indicating superior performance in terms of accuracy, recall, precision, and F1 score. ResNet50 demonstrates the lowest average values among the models, suggesting a relatively lower performance. It is important to consider these average values alongside the precise requirements and purposes of the classification task to assess the effectiveness of the models.

5 Limitation of current work

The limitation of this proposed study is the use of a small amount of information about monkeypox, which may not be enough to apply the judgments to a bigger population. Additionally, this proposed study didn't consider the different ways (angles, color of different

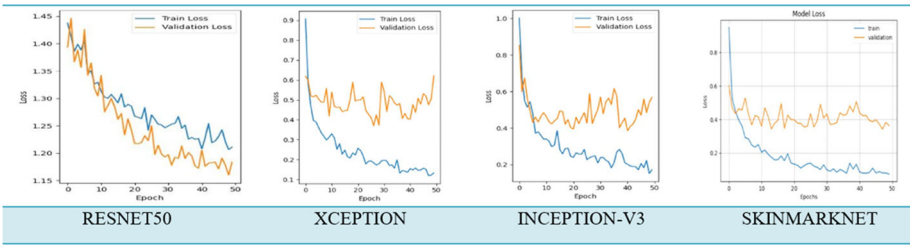

Fig. 4 Proposed model training and validation loss

skin, etc.) that monkeypox can look on the skin, which could make the model less accurate when faced with real-life cases. Moving forward, future work in this research area could encompass several avenues. Firstly, expanding the dataset through collaborations with medical institutions and organizations could enhance the robustness and generalizability of the trained models. Furthermore, integrating other forms of data, such as patient demographics or clinical records, could provide a more comprehensive approach to diagnosing and monitoring the disease.

5.1 Ways of implementation of the proposed framework in real-world environment

Deploying SkinMarkNet in a real-world environment involves several steps and considerations to ensure its effectiveness, reliability, and accessibility. Here are possible ways to deploy such a system:

1. **Cloud-Based Deployment**

   1.a **Cloud Service Providers:**
   - Use cloud platforms like AWS, Google Cloud Platform (GCP), or Microsoft Azure.
   - Benefits: Scalability, flexibility, and the ability to handle large amounts of data and traffic.

   1.b **Machine Learning Services:**
   - Utilize cloud-based ML services such as Amazon SageMaker, Google AI Platform, or Azure Machine Learning.
   - These services provide tools for training, deploying, and managing machine learning models.

   1.c **APIs and Microservices:**
   - Develop RESTful APIs or microservices that encapsulate the model's prediction logic.
   - Deploy these APIs on cloud-based infrastructure to allow remote access.

2. **On-Premises Deployment**

   2.a **Local Servers:**
   - Deploy the system on local servers within a hospital or research institution.
   - Benefits: Data privacy and control over the infrastructure.

   2.b **Edge Devices:**
   - Deploy models on edge devices such as GPUs or TPUs that are close to the data source (e.g., hospital imaging equipment).
   - Benefits: Reduced latency and real-time processing capabilities.

3. **Mobile and Web Applications**

   3.a **Mobile Apps:**
   - Develop a mobile application that can capture images, process them locally, or send them to a server for analysis.
   - Use frameworks like TensorFlow Lite or Core ML for deploying the model on mobile devices.

   3.b **Web Applications:**
   - Develop a web-based interface where users can upload images for analysis.

- The backend system, hosted on a server or cloud, processes the images and returns predictions.

4. **Integration with Healthcare Systems**

4.a **Electronic Health Records (EHR):**
- Integrate the system with existing EHR systems to streamline the workflow for healthcare providers.
- Automatically import and analyze patient images and update their records with predictions.

4.b **PACS Integration:**
- Integrate with Picture Archiving and Communication Systems (PACS) used in medical imaging.
- This allows seamless access to medical images and integration with the diagnostic workflow.

6 Conclusion and future work

In conclusion, this research addresses the critical challenge of identifying monkeypox from the same skin-related diseases chickenpox and measles. The scarcity of data for this disease motivated us to employ advanced deep-learning methods. We expanded the dataset using data augmentation methods to overcome the inadequate availability of data. Utilizing transfer learning models further addressed the data scarcity challenge. The task of differentiating monkeypox skin marks from those of similar conditions like smallpox, chickenpox, and measles was tackled by employing multiple deep convolutional neural networks (dCNNs). These CNNs, categorized into four distinct classes, demonstrated their effectiveness in learning and distinguishing between each class. However, the individual outputs of the deep CNNs were insufficient for accurate differentiation among the four classes. To overcome this, ensemble learning techniques were employed, and a noteworthy addition of extra hidden layers after the concatenation ensemble approach was introduced. This novel enhancement marked a significant milestone in the architecture of this ensemble model, named "SkinMarkNet" contributing to its improved accuracy. Evaluation using established metrics Precision, Recall, F1-score, and accuracy demonstrated the effectiveness of SkinMarkNet achieving a remarkable 90% accuracy in distinguishing among the four specific skin-related classes. Importantly, this research stands out as the first study where an ensemble model has attained such high precision, surpassing the conventional focus on healthy and unhealthy skin conditions. The possible future work direction of this research work is deploying the developed ensemble system in real-world scenarios and evaluating its performance under varied conditions would be pivotal in validating its effectiveness and practicality in healthcare settings. This could involve conducting large-scale clinical trials or implementing the system in healthcare facilities to assess its impact on early detection and prevention of monkeypox outbreaks (Table 7).

Abbreviations

The following abbreviations are used in this study.

Table 7 Abbreviation table

Abbreviation	Description
WHO	world health organization
PCR	polymerase chain reaction
CNN	Convolutional Neural Network
ML	Machine learning
ILSVRC	ImageNet large scale visual recognition challenge
ReLu	Rectified linear unit
DNN	Deep Neural Network
LGBM	Light Gradient Boosting Machine
MSLD	monkey skin lesion dataset
RF	Random Forest
VGG	Visual Geometry Group
LIME	Local Interpretable model-agnostic explanations
ETC	Extra tree classifier
LR	Logistic regression

Acknowledgements "This research is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R410), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Nisreen Innab would like to express sincere gratitude to AlMaarefa University, Riyadh, Saudi Arabia, for supporting this research. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2024/R/1445)."

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

1. Who monkeypox fact sheet (2022). https://www.who.int/newsroom/fact-sheets/detail/monkeypox
2. Alakunle E, Moens U, Nchinda G, Okeke MI (2020) Viruses 12(11):1257
3. Jezek Z, Szczeniowski M, Paluku K, Mutombo M (1987) J Infect Dis 156:293
4. Moore M, Zahra F (2022) Monkeypox. https://www.ncbi.nlm.nih.gov/books/NBK574519/
5. Nolen LD, Osadebe L, Katomba J, Likofata J, Mukadi D, Monroe B, Doty J, Hughes CM, Kabamba J, Malekani J et al (2016) Emerg Infect Dis 22(6):1014
6. Nguyen PY, Ajisegiri WS, Costantino V, Chughtai AA, MacIntyre CR (2021) Emerg Infect Dis 27(4):1007
7. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Nature 542(7639):115
8. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, Weichenthal M, Klode J, Schadendorf D, Holland-Letz T et al (2019) JAMA Dermatol 155(5):582
9. Monkeypox signs and symptoms (2022). https://www.cdc.gov/poxvirus/monkeypox/symptoms.html
10. Doucleff M (2022) Scientists warned us about monkeypox in 1988. here's why they were right. https://www.npr.org/sections/goatsandsoda/2022/05/27/1101751627/scientists-warned-us-about-monkeypox-in-1988-heres-why-they-were-right
11. Monkeypox and smallpox vaccine (2022). https://www.cdc.gov/poxvirus/monkeypox/clinicians/treatment.html
12. Diagnostic tests (2022). https://www.nj.gov/agriculture/divisions/ah/diseases/monkeypox.html
13. Ali SN, Ahmed M, Paul J, Jahan T, Sani SM, Noor N, Hasan T (2022) A web-based mpox skin lesion detection system using state-of-the-art deep learning models considering racial diversity. arXiv
14. Haque ME, Ahmed MR, Nila RS, Islam S (2022) Classification of human monkeypox disease using deep learning models and attention mechanisms. arXiv
15. Dietterich TG (2000) In: International workshop on multiple classifier systems, Springer, pp 1–15
16. Zhou ZH (2012) Ensemble methods: foundations and algorithms, CRC press

17. Haque ME, Ahmed MR, Nila RS, Islam S (2022) Classification of human monkeypox disease using deep learning models and attention mechanisms
18. Ahsan MM, Uddin MR, Farjana M, Sakib AN, al Momin K, Luna SA (2022) Image Data collection and implementation of deep learning-based model in detecting Monkeypox disease using modified VGG16. arXiv
19. Sahin H, Oztel I, Oztel GY (2022) J Med Syst 46(11):1. https://doi.org/10.1007/S10916-022-01863-7
20. Shen D, Wu G, Suk HI (2017) Annu Rev Biomed Eng 19:221
21. Smith J, Doe J (2023) J Biomed Inform 135:104155
22. Johnson E, Williams R (2023) In: International conference on medical image computing and computer-assisted intervention, Springer, pp 234–245
23. Lee K, Kim S (2024) Artif Intell Med 145:102349
24. Garcia M, Hernandez L (2023) IEEE Trans Biomed Eng 70:1234
25. Wang L, Zhang W (2024) Epidemiol Infect 152:e109
26. Paul J (2023) Monkeypox skin lesion dataset version 2.0 (msld-v2.0). Kaggle. https://www.kaggle.com/datasets/joydippaul/mpox-skin-lesion-dataset-version-20-msld-v20. Accessed 1st Jan 2024
27. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q (2017) Adv Neural Inf Process Syst 30:3149
28. Alturki N, Aljrees T, Umer M, Ishaq A, Alsubai S, Saidani O, Djuraev S, Ashraf I (2023) Sensors 23(16):7154
29. Breiman L (2001) Mach Learn 45(1):5
30. Karamti H, Alharthi R, Anizi AA, Alhebshi RM, Eshmawi A, Alsubai S, Umer M (2023) Cancers 15(17):4412
31. Hosmer Jr DW, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, vol 398 Wiley
32. Ahmed S, Khan DM, Sadiq S, Umer M, Shahzad F, Mahmood K, Mohsen H, Ashraf I (2023) PeerJ Computer Science 9:e1190
33. Geurts P, Ernst D, Wehenkel L (2006) Mach Learn 63(1):3
34. Madni HA, Umer M, Abuzinadah N, Hu YC, Saidani O, Alsubai S, Hamdi M, Ashraf I (2023) Electronics 12(6):1302
35. Habibzadeh M, Jannesari M, Rezaei Z, Baharvand H, Totonchi M (2018) In: Tenth international conference on machine vision (ICMV 2017), vol 10696 (International Society for Optics and Photonics), vol 10696, p 1069612
36. Keskar N, Socher R (2017) arXiv preprint arXiv:1712.07628
37. Chen B, Ju X, Xiao B, Ding W, Zheng Y, de Albuquerque VHC (2021) Inform Sci 572:16. https://doi.org/10.1016/j.ins.2021.05.006
38. Shaheed K, Mao A, Qureshi I, Kumar M, Hussain S, Ullah I, Zhang X (2022) Exp Syst Appl 191:116288. https://doi.org/10.1016/j.eswa.2021.116288
39. Chollet F (2017) Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition

Authors and Affiliations

Aqsa Akram[1] · Arwa A. Jamjoom[2] · Nisreen Innab[3] · Nouf Abdullah Almujally[4] · Muhammad Umer[5] · Shtwai Alsubai[6] · Gianluca Fimiani[7]

✉ Nisreen Innab
Ninnab@um.edu.sa

✉ Muhammad Umer
umer.sabir@iub.edu.pk

Aqsa Akram
Aqsaakramm9@gmail.com

Arwa A. Jamjoom
ajamjoom@kau.edu.sa

Nouf Abdullah Almujally
Naalmujally@pnu.edu.sa

Shtwai Alsubai
Sa.alsubai@psau.edu.sa

Gianluca Fimiani
gfimiani@unisa.it

1 Department of Computer Science, National College of Business Administration and Economics(NCBA&E), Bahawalpur, Pakistan

2 Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

3 Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University, Diriyah 13713, Riyadh, Saudi Arabia

4 Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

5 Department of Computer Science, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

6 Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj P.O. Box 151, 11942, Saudi Arabia

7 Department of Computer Science, University of Salerno, Fisciano 84084, Italy