Grenze International Journal of Engineering and Technology, January Issue

Hybrid Transformer Model for Skin Lesion Classification with Focuses on Monkeypox Diagnosis

Imran Mehdi[1], Neerendra Kumar[2] and Harnain Kour[3]
[1-3]Department of Computer Science and Information Technology, Central University of Jammu, Bagla (Rahya Suchani), Samba, 181143, Jammu and Kashmir, India.
Email: im23mcst04@cujammu.ac.in, neerendra.csit@cujammu.ac.in, harnain.cse@cujammu.ac.in

***Abstract*— Monkeypox is a serious health problem that needs immediate diagnosis and quarantine of infected individuals. Distinguishing Monkeypox from other cutaneous lesions, including Chickenpox or Measles, is important since misdiagnosis may have severe consequences. Visual differentiation among these lesions is frequently impractical. Conventional approaches such as dermoscopy or high-resolution ultrasound imaging, though helpful, are predicated on subjective analysis and take time. In this research, a combination of EfficientNetB3-Transformer-MLP is suggested for quantitative and objective classification of the six types of skin lesions: Chickenpox, Cowpox, HFMD, Healthy, Measles, and Monkeypox. The model combines EfficientNetB3, ImageNet pretrained as a feature extractor, with a Vision Transformer (ViT) to capture interpatch interactions and an MLP for ultimate classification, utilizing transfer learning and fine-tuning. The hybrid model contrasted with single EfficientNetB3 and Transformer models, along with other widely used pretrained deep learning networks, utilizing a transfer learning framework. The proposed hybrid model presented a better performance in terms of accuracy. The outcomes reflect competitive performance against existing studies, specifically in having balanced metrics against all classes, even with the drawbacks of class imbalance. The proposed model provides a trustworthy approach for researchers and dermatologists to speedily and effectively diagnose skin lesions, particularly in urgent situations such as Monkeypox outbreaks.**

***Keywords*— Monkeypox, Deep Learning, Multi-Layer Perceptron, Vision Transformer, Machine Learning.**

I. INTRODUCTION

Monkeypox, a zoonotic viral infection, was initially discovered in the laboratory of monkeys in 1958 while they were labouring on pock-like disease. Rodents, however, such as Gambian pouched rats and dormice, are the usual reservoir for monkeypox virus [1]. Human cases first were described in 1970 in the Democratic Republic of Congo (DRC), in immunodepleted districts smallpox following the withdrawal of vaccination following the eradication of smallpox from the world [2]. Eradication of smallpox. Monkeypox is a infection caused by the monkeypox virus, which is part of the Orthopoxvirus genus. It is structurally akin to the variola virus (smallpox) but less fatal. The monkeypox outbreaks were previously confined to Central and West African rainforests [3]. The pandemic since 2022, however, has placed the reality of the worldwide spread of the disease in focus, and fears regarding its epidemic and pandemic potential. The global situation of monkeypox has considerably developed over the last decade.

Grenze ID: 01.GIJET.12.1.552

The 2022–2024 outbreaks were highly endemic, and the disease spread to over 122 countries, many of which had no prior history of monkeypox [4]. In August 2024, the WHO declared a state of global health emergency in response to the pox epidemic that surfaced in the Democratic Republic of Congo and went on to diversify to a more large number of African nations. A total of more than 13000 mpox cases and 49 deaths so far in the DRC, with children being the main victims [5]. The new epidemic is being fuelled by a new generation of clade 1, also known as clade 1b, which looks to transmit widely by close contact. For the first time, clade one mpox is also transmitted through sexual contact [6].

In India, 33 cases have been reported between 2022 and 2025, and 1 death, mostly among travellers or those in close contact with laboratory-confirmed cases [7]. Indian health authorities, learning from COVID-19, took stringent Surveillance and containment interventions for the control of such cases and the prevention of its widespread. Interventions in public health involved the creation of standardized diagnostic laboratories and the promotion of awareness of the disease [8]. Monkeypox is spread by several routes that include direct contact with the lesions, body fluids, or respiratory droplets close to an infected individual over a long period of time. Indirect spread is through contaminated materials like clothing or bedding [9]. The concern over animal-to-human persists and is most particularly in those areas with frequent human interaction with wildlife shown in Fig. 1. Unlike smallpox, monkeypox can also spread through sexual contact, as evidenced by atypical outbreaks in non-endemic countries [10].

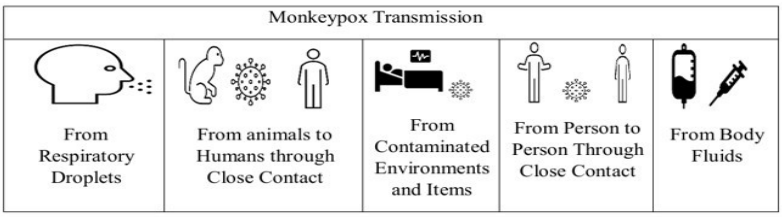

Fig. 1 Transmission of monkeypox virus from air, from animals, from person

Monkeypox has stages, with the initial incubation period ranging from 5–21 days, followed by prodromal symptoms like fever, headache, muscle aches, and lymphadenopathy. The hallmark of monkeypox is a rash that evolves through macules, papules, vesicles, pustules, and scabs [1]. The outbreak of monkeypox recognizes important gaps in global preparedness to meet new infectious disease threats. WHO declared monkeypox as a Public Health Emergency in 2022, raised International Concern, and identified the need for concerted action [4]. The key contributions of this study include the following points (i.-iii.):

- A hybrid Transformer-MLP model is created that uses EfficientNetB3 to extract spatial features from MSLD v2.0 skin lesion images, the Transformer encoder to model interpatch relationships for detailed contextual features (e.g., lesion distribution), and the MLP classifies into different classes.
- Optimized the hybrid EfficientNetB3-Transformer-MLP model's architecture through hyperparameter tuning, adjusting the Transformer encoder (8 heads, 512 FFN units, 0.1 dropouts) for efficient capturing of patterns (e.g., lesion patterns in MSLD v2.0).
- An evaluation of the proposed hybrid model in comparison to the Existing studies.

II. Related Work

A. Traditional Techniques

The traditional ML algorithms such as SVM and Decision Tree initially addressed the issue of skin lesion classification. While efficient computationally, traditional techniques were poor at handling high-dimensional image data with immense feature engineering needed. Studied ensemble-based ML approaches with 89% accuracy [18]. Nevertheless, the models lacked scalability and flexibility owing to their reliance on hand-engineered feature selection and extraction methods [19]. As such, DL techniques, particularly CNNs, resolved most limitations of classical ML. DL models make use of hierarchical fea-ture representations of raw image data and hence increase the accuracy and robustness of skin lesion classifiers. For instance, Author has deployed VGG19 on MSID with 83% accuracy. Their work also underscored the advantage of using other metadata, such as patient demographics and history of the lesion, that can enhance model performance [20].

Clinical uptake of DL models is constrained by their black box nature, regardless of high accuracy. Explainable AI (XAI) methods such as Grad-CAM and saliency maps have been suggested as means to render model

predictions visualizable by highlighting the most critical regions within an image. Heatmap visualizations of the CNN-based classifiers were employed to make model decisions with accuracy [17]. The method aimed at explainability diagnostics accuracy tradeoff and increased clinician confidence in AI systems.
To enhance feature extraction, the Convolutional Block Attention Module (CBAM) is incorporated into VGG19, Xception, DenseNet121, EfficientNetB3, and MobileNetV2. The Xception-CBAM-Dense model achieved the highest validation accuracy of 83.89%, confirming the effectiveness of attention processes in enhancing classification performance [21]. The number of studies trying to detect monkeypox lesions from skin images has grown in recent years due to the increased human monkeypox disease. The authors developed a new human monkeypox image database in [1] They thereafter labelled the images using DL techniques.
Further, in recent works, some of the monkeypox classification research has utilized the Vision transformer model. Authors of [22] had employed a binary classification of monkeypox versus not monkeypox using a well hyperparameter Vision Transformer model. The study in [23] had used a monkeypox versus not monkeypox binary classification using Vision Transformer and few other machine learning models. Afterwards, the models were compared. The binary classification with Vision Transformer and three other transfer learning models modified as such VGG and ResNet [24]. Table 1 depicts the comparative analysis of existing DL techniques.

TABLE I: COMPARATIVE ANALYSIS OF DL BASED TECHNIQUES, PERFORMANCE, DATASETS AND LIMITATIONS

Author	Technique Used	Dataset	Performance	Limitations
S N Ali, et al. 2022 [1]	DL (VGG-16, ResNet50, Inception V3)	MSID consists of 228 images among which 102 belong to monkey pox and 126 belong to others (chickenpox and measles)	Accuracy 82.96%	The dataset is very small, leading to overfitting, especially for models like VGG-16 and ResNet50, which are made for larger datasets.
T Islam, et al. 2022 [20]	ResNet50, Inception-V3, DenseNet121, MnasNet-A1, MobileNet-V2, ShuffleNet-V2, SqueezeNet	MSLD V2.0 consists of 755 images	Accuracy 79%	Accuracy is low, indicating the model's inability to efficiently learn the discriminative features from the dataset.
Haque, et al. 2022 [21]	MobileNet- V2, DenseNet121, VGG-19	MSID	Accuracy 84%, Precision 90%, Recall 85%	Binary classification and models like VGG-19 and DenseNet121 have a large number of parameters and are prone to overfitting.
Sitaula, et al. 2022 [25]	EffectNet, Inception, ResNet, VGG-16, MobileNet	MSID	Accuracy 87%	A lesser number of classes do not fully capture the diversity of skin lesions.
Gozde Yolcu Oztel, 2023 [26]	Vision transformer with CNN	MSLD	Accuracy 81.91%, Precision 87%, Recall 74%	Recall is very low, causing the models to struggle to identify all positive cases.

B. Research gaps

Based on the analysis of existing studies, following research gaps are identified (i.-iv.):

- Visual Similarity Among Diseases: Monkeypox shares overlapping features with Chickenpox and Measles in terms of lesion texture and color, while intra-class differences like lesion stages and patient skin tone further complicate consistent classification [1].
- Lack of Multimodal Integration: Most models rely solely on images, missing out on complementary data such as clinical symptoms, molecular diagnostics, and exposure history, which are crucial for accurate Monkeypox identification [2].
- Limited Use of Hybrid AI Models: There is a research gap in combining deep learning with evolutionary algorithms (e.g., genetic algorithms) for optimized feature selection, which could enhance model precision for visually similar diseases [27].
- Dataset Limitations: Small sample sizes, class imbalance, and variability in image quality and skin tones in datasets like MSID-v2.0 hinder the training of fair and generalizable classification models [27].

III. PROPOSED METHODOLOGY

The proposed framework depicted in Fig. 2 details the pipeline for training a hybrid EfficientNetB3-Transformer-MLP model on the Mpox Skin Lesion Dataset Version 2.0 (MSLD v2.0), which consists of 6 classes: Monkeypox, Cowpox, Chickenpox, Measles, HFMD, and Healthy. It starts with image preprocessing,

where MSLD v2.0 images are resized to 224x224 and normalized to be compatible with EfficientNetB3's ImageNet weights. Data augmentation comes next, adding rotation, flipping, and brightness changes to increase model resilience. Originally, the dataset contained 755 images, and after augmentation, the size increased to 20800 images. The data is then divided into a training, test, and Validation set. The hybrid model, which includes EfficientNetB3, Transformer Encoders, and an MLP Classifier, is trained for more than 50 epochs. Performance metrics include calculating training and test accuracy and loss, printing a classification report, a confusion matrix, and a loss plot comparing training and test loss. This pipeline guarantees the model learns to classify MSLD v2.0 images well, with evaluation metrics reflecting its generalization over the six classes. Fig. 4 demonstrate how hybrid model works

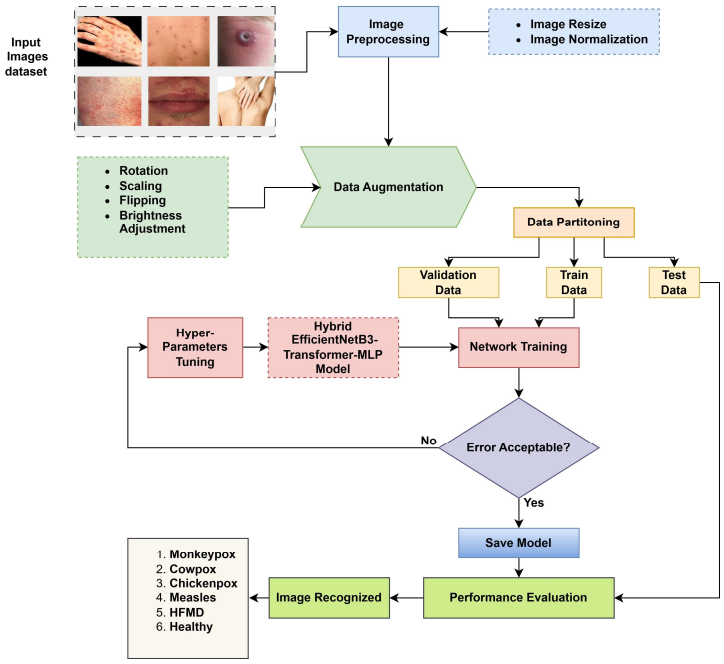

Fig. 2 Proposed framework

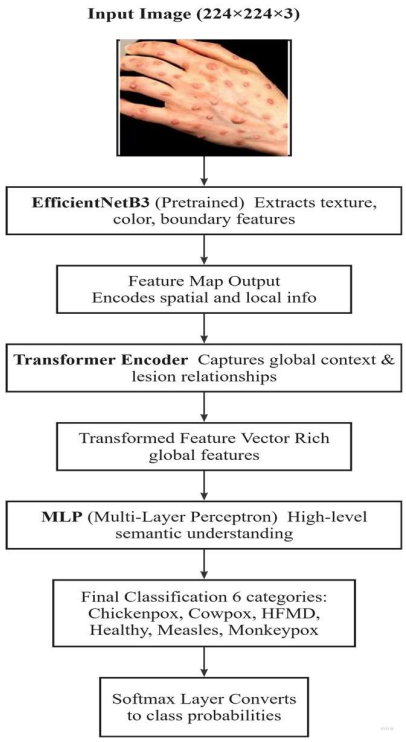

Fig. 3 Hybrid Model working

The hybrid Transformer Model, taking EfficientNetB3 as its CNN backbone, processes the dataset MSLD v2.0 (including images of skin lesions belonging to 6 categories: Monkeypox, Cowpox, Chickenpox, Measles, HFMD, Healthy) through sequential layers, which pull out and filter out special features specific to differentiate these complicated lesion patterns in Fig. 5. It starts with the input layer, taking 224x224x3 RGB images from the dataset MSLD v2.0, which maintains similarity in size as well as shape. The EfficientNetB3 CNN then extracts hierarchical features, outputting an 8x8x1536 feature map in Fig. 4. The feature map (Fc) from the CNN is obtained using the EfficientNetB3 model, as shown in (1):

$$F_c = \text{EfficientNetB3}(x; \theta_{cnn}) \qquad (1)$$

where, $x \in R^{224\times224\times3}$ is the input image, and θ_{cnn} are frozen ImageNet weights, producing $F_c \in R^{8\times8\times1536}$. EfficientNetB3, with its compound scaling of depth, width, and resolution, is especially well-suited for MSLD v2.0 since its initial layers (convolutional blocks) pick up low-level features such as edges and color gradients (e.g., redness in Monkeypox lesions or whiteness of vesicles in Chickenpox), mid-level layers detect textural patterns (e.g., roughness of pustules in Monkeypox), and deeper layers detect high-level semantic features (e.g., shapes of lesions distinguishing Measles' flat rashes from HFMD's vesicular clusters). The 1536 channels improve the capacity of the model to capture fine-grained details of lesions, such as the very subtle differences between MSLD v2.0 classes.

The reshape layer (P) reshapes this feature map into a 64x1536 patch sequence utilizing (2):

$$P = \text{Reshape}\,(F_c) + E\text{pos} \quad (2)$$

where, E_{pos} = Embedding(k,1536), with k = 0,1,...,63, ensuring the Transformer can distinguish between, for example, a cluster of top-left pustules in Monkeypox and a bottom-right rash in Measles.

The Transformer encoder, consisting of two blocks, processes these patches with multi-head attention and feed-forward networks (FFNs). The multihead attention mechanism (with 8 heads and d_k = 1536/8 = 192) calculates attention scores for every pair of patches in utilizing (3):

$$\text{Attention}\,(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \quad (3)$$

where, Q(*Query*) = The query matrix represents the input patches in a form that asks, K(*Key*) = The key matrix represents the input patches in a form that answers, V (*Value*) = The value matrix contains the actual information from the patches that will be combined based on attention scores.

IV. Results and Discussion

A. Dataset Description

The dataset used in this study obtained from Kaggle MSLD v2.0 comprises images from six distinct classes, namely Mpox (284 images), Chickenpox (75 images), Measles (55 images), Cowpox (66 images), Hand-foot-mouth disease or HFMD (161 images), and Healthy (114 images). The dataset includes 755 original skin lesion Images [30].

B. Experimental setup and evaluation metrics

The training and testing were performed by utilizing Nvidia RTX 4060 GPU with a capacity of 8GB in VS Code via a Jupyter Notebook. The training parameters used while training the hybrid EfficientNetB3-Transformer-MLP model are given below: learning rate of 1e-5, batch size of 64, and number of epochs equal to 50, with the evaluation strategy being epoch-wise validation. Within the classification frameworks, performance is evaluated based on a number of criteria indices, which allow users and researchers to examine the performance of the hybrid EfficientNetB3-TransformerMLP model for skin lesion classification in-depth. The indices offer information about various aspects of the system's performance, with each having strengths and weaknesses. Thus, their use together gives a more extensive assessment. In this research, four performance metrics Accuracy (a), Recall (r), F1-Score (F), and Precision (p) are used to evaluate results utilizing (4)-(7) [31].

$$a = \frac{T_P T_N}{T_P + F_N + T_N + F_P} \quad (4)$$

$$p = \frac{T_P}{T_P + F_P} \quad (5)$$

$$r = \frac{T_P}{T_P + F_N} \quad (6)$$

$$F = \frac{2pr}{p + r} \quad (7)$$

where, T_P = True Positive, T_N= True Negative, F_P= False Positive, and F_N=False Negative

Fig. 6 demonstrates the hybrid EfficientNetB3-Transformer-MLP confusion matrix for the classification of skin lesions. We use different hyperparameters for testing our hybrid model and the values where we find highest accuracy are highlighted and are shown in Table 3.

Although the system has a few misclassifications, in general, it provides encouraging performance with an average accuracy of 88%.

The best performance is experienced for the class Monkeypox, where there are 82 out of 93 samples accurately classified, thereby proving the strength of the model in detecting the condition, whose pustular characteristics are particularly unique. In contrast, the worst performance is found for the chickenpox class, with six samples classified as Monkeypox, potentially because of a visual resemblance between lesion patterns (e.g., pustules) in these two diseases.

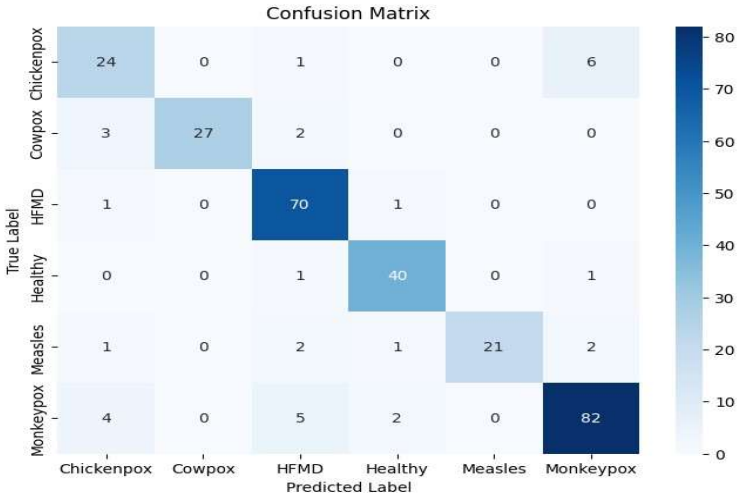

Fig. 4 Confusion matrix of hybrid EfficientNetB3-Transformer-MLP model

TABLE III: HYPERPARAMETER TESTING FOR EFFICIENTNETB3-TRANSFORMER-MLP MODEL

Hyperparameter	Value	Accuracy (%)	Precision	Recall	F1 Score
Batch size	32 **64**	85 88	0.84 0.875	0.83 0.87	0.84 0.872
	128	79	0.82	0.81	0.82
Optimizer	SGD **Adam**	80 88	0.78 0.875	0.773 0.87	0.79 0.872
Activation Function	ReLU **GELU**	84 88	0.81 0.875	0.77 0.87	0.83 0.872
Transformer layer	1 **2**	81 88	0.83 0.875	0.81 0.87	0.84 0.872
	3	83	0.822	0.74	0.77

C. Comparative analysis

The hybrid EfficientNetB3-Transformer-MLP model is contrasted with standalone EfficientNetB3 and Transformer models to gauge skin lesion classification performance based on accuracy, precision, recall, and F1-score on the test set of the dataset. Table 4 shows that the hybrid EfficientNetB3-Transformer-MLP model has higher accuracy, precision, recall, and F1-score, and their performance is shown in Fig. 5.

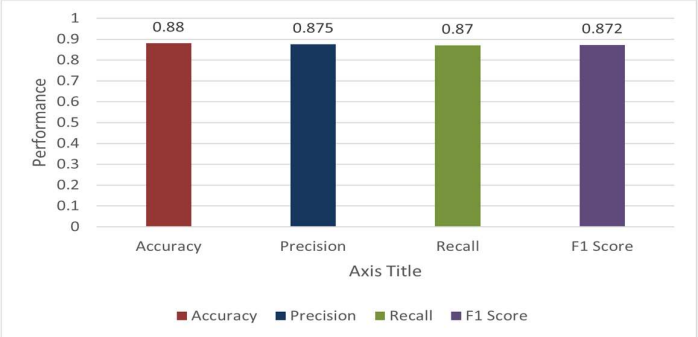

Fig. 5 Performance evaluation of the proposed framework

When the proposed model is compared to standalone EfficientNetB3 and Transformer models. The hybrid approach suggested in this paper is to enhance overall accuracy with equilibrated precision and recall scores to

offer valid skin lesion classification, such as monkeypox, without sacrificing important diagnoses. It is evident from Table 4 that the single Transformer model possesses poor classification accuracy compared to EfficientNetB3 and the hybrid EfficientNetB3 Transformer-MLP model.

TABLE IV: COMPARISON OF PROPOSED METHOD WITH OTHER STUDIES

Model/Technique	Accuracy (%)	Precision	Recall	F1-score
EfficientNetB3 [1]	74.61 ± 3.94	0.75 ± 0.06	0.71 ± 0.06	0.72 ± 0.06
T Islam, et al. [20]	83	0.85	-	-
Haque, et al. [21]	83.89	-	-	-
C Sitaula, TB Shahi [25]	85.44	0.85	0.85	0.84
Vision Transformer [26]	80.66	0.80	0.80	0.80
Bagging-Ensemble Vision Transformer With DenseNet201 [26]	81.91	0.87	0.74	0.78
Proposed Model	**88**	**0.875**	**0.87**	**0.872**

This is mainly due to the poor ability of the Transformer to extract fine-grained spatial features from images when standing alone since it is patch-based processing dependent without the convolutional feature extraction capability that is crucial in skin lesion analysis. While Transformers excel in modeling long-range dependencies, they perform poorly on local feature extraction from image data without a CNN backbone. EfficientNetB3 as a solo model is better than the Transformer due to its deep convolutional structure, which successfully extracts the spatial features of lesion textures and patterns. EfficientNetB3 alone lacks the global context modeling of the Transformer.

This work employs a hybrid approach by mixing EfficientNetB3's powerful feature extraction with the ability of the Transformer to model inter-patch relations and then applying an MLP for final classification. The Transformer component overcomes the use of individual Transformers by leveraging on the robust features extracted by EfficientNetB3 and the MLP, giving robust classification by mapping such features to the 6 skin lesion classes (Chickenpox, Cowpox, HFMD, Healthy, Measles, Monkeypox). The best result is obtained when the deep feature extraction capability of EfficientNetB3 is merged with the global context modeling of the Transformer and the classification potency of the MLP, as established by the enhanced performance of the hybrid model.

V. CONCLUSION

This paper proposed a hybrid EfficientNetB3-Transformer-MLP model for the classification of skin lesions, e.g., Monkeypox, into 6 classes: Chickenpox, Cowpox, HFMD, Healthy, Measles, and Monkeypox. For this, the Vision Transformer (ViT) state-of-theart was mixed with EfficientNetB3 and an MLP. Vision Transformer models perform better on larger datasets, but since the comparatively small dataset size was involved, transfer learning and fine-tuning were applied. The pretrained EfficientNetB3 module, trained on ImageNet (a dataset of over 14 million labeled images), was used as the feature extractor with frozen weights while the ViT and MLP were fine-tuned for skin lesion classification. For comparative purposes, the hybrid model was contrasted with standalone Efficient-NetB3 and Transformer models. In all experiments, the hybrid model outperformed the single models in terms of Accuracy, Precision, Recall, and F1-Score, with a balanced outcome in all four parameters. Additionally, some known pre-trained networks were trained on skin lesion classification using transfer learning and compared in their test results. EfficientNetB3 had the highest standalone model accuracy. The highest total performance was obtained from the hybrid model with feature extraction by EfficientNetB3, global context modeling by ViT, and MLP classification. Results are demonstrated in tables within the Comparative analysis section. The hybrid EfficientNetB3-Transformer-MLP model achieved 88.0% Accuracy, 87.5% Precision, 87.0% Recall, and 87.2% F1-Score for the 6-class classification task. Depending on these values, the model rivals the state-of-the-art methods and outperforms most of them, particularly concerning symmetrical performance for every metric. The proposed model possesses promising potential in classifying skin lesions, particularly in diseases like monkeypox, where recognition needs to be done quickly and accurately.

In future work, incorporating multimodal solutions by integrating patient history, molecular diagnostics, and epidemiological data alongside image analysis can significantly improve the model's diagnostic accuracy. Leveraging this contextual information will enable more precise differentiation between Monkeypox and

mimicking skin diseases, such as Chickenpox or Measles, and also expand the dataset with more diverse and balanced samples.

REFERENCES

[1] Ali, S.N., Ahmed, M.T., Paul, J., Jahan, T., Sani, S.M.S., Noor, N., Hasan, T.: Monkeypox skin lesion detection using deep learning models: A feasibility study. IEEE Journal of Biomedical and Health Informatics (2022) https://doi.org/10.48550/arXiv.2207.03342

[2] Sahin, V.H., Oztel, I., Yolcu Oztel, G.: Human monkeypox classification from skin lesion images with deep pre-trained network using mobile application. Journal of Medical Systems 46(11), 79 (2022) https://doi.org/10.1007/s10916-022-01863-7

[3] Letafati, A., Sakhavarz, T.: Monkeypox virus: A review. Microbial Pathogenesis 176, 10961208 (2023) https://doi.org/10.1016/j.micpath.2023.106027

[4] Organization, W.H.: Mpox: WHO Health Emergency Appeal 2025. Online. [Accessed: Mar. 27, 2025] (2025). https://www.who.int/publications/m/item/ mpox--who-health-emergency-appeal-2025

[5] Data, O.W.: Mpox - Our World in Data. Online. [Accessed: Mar. 27, 2025] (2025). https://ourworldindata.org/mpox

[6] Duarte, P.M., Adesola, R.O., Priyadarsini, S., Singh, R., Shaheen, M.N.F., Ogundijo, O.A., Gulumbe, B.H., Lounis, M., Samir, M., Govindan, K., Adebiyi, O.S., Scott, G.Y., Ahmadi, P., Mahmoodi, V., Chogan, H., Gholami, S., Shirazi, O., Moghadam, S.K., Jafari, N., Shahabinejad, F., Subbaram, K., Fawzy, M., Rehman, A., Tazerji, S.S.: Unveiling the global surge of mpox (monkeypox): A comprehensive review of current evidence. The Microbe 4, 100141 (2024) https://doi.org/10.1016/j.microb.2024.100141

[7] Alakunle, E., Kolawole, D., Diaz-C´anova, D., Alele, F., Adegboye, O., Moens, U., Okeke, M.I.: A comprehensive review of monkeypox virus and mpox characteristics. Frontiers in Cellular and Infection Microbiology 14, 2235–2988 (2024) https://doi.org/10.3389/fcimb.2024.1360586

[8] Naga, N.G., Nawar, E.A., Mobarak, A.A., Faramawy, A.G., Al-Kordy, H.M.H.: Monkeypox: a re-emergent virus with global health implications – a comprehensive review. Tropical Diseases, Travel Medicine and Vaccines 11, 20550936 (2025) https://doi.org/10.1186/s40794-024-00237-w

[9] Ahmadi, S., Amirzadeh, M., Ahmadi, M., Soleiman-Meigooni, S.: From outbreaks to artificial intelligence: A comprehensive review of monkeypox virus epidemiology, diagnosis, treatment, vaccination, and deep learning applications. Journal of Tropical Medicine 1, 1687–9686 (2024) https://doi.org/10.1155/jotm/6688914

[10] Nakhaie, M., Pirmoradi, Z., Bashash, D., Rezaei Zadeh Rukerd, M., Charostad, J.: Beyond skin deep: shedding light on the neuropsychiatric consequences of monkeypox (mpox). Acta neurologica Belgica 124, 1189–1197 (2024) https://doi.org/10.1007/s13760-023-02361-4

[11] Chadaga, K., Prabhu, S., Sampathila, N., Nireshwalya, S., Katta, S.S., Tan, R.S., Acharya, U.R.: Application of artificial intelligence techniques for monkeypox: A systematic review. Diagnostics 13, 20754418 (2023) https://doi.org/10.3390/diagnostics13050824

[12] Debelee, T.G.: Skin lesion classification and detection using machine learning techniques: a systematic review. Diagnostics 13(19), 3147 (2023) https://doi.org/ 10.3390/diagnostics13193147.

[13] Jaradat, A.S., Mamlook, R.E.A., Almakayeel, N., Alharbe, N., Almuflih, A.S., Nasayreh, A., Gharaibeh, H., Gharaibeh, M., Gharaibeh, A., Bzizi, H.: Automated monkeypox skin lesion detection using deep learning and transfer learning techniques. International Journal of Environmental Research and Public Health 20(4), 4422 (2023) https://doi.org/10.3390/ijerph20054422

[14] Wu, Y., Chen, B., Zeng, A., Pan, D., Wang, R., Zhao, S.: Skin cancer classification with deep learning: a systematic review. Frontiers in Oncology 12, 893972 (2022) https://doi.org/10.3389/fonc.2022.893972

[15] Bala, D., Hossain, M.S., Hossain, M.A., Abdullah, M.I., Rahman, M.M., Manavalan, B., Gu, N., Islam, M.S., Huang, Z.: Monkeynet: A robust deep convolutional neural network for monkeypox disease detection and classification. Neural Networks 161, 757–775 (2023) https://doi.org/10.1016/j.neunet.2023.02.022

[16] Azar, A.S., Naemi, A., Rikan, S.B., Mohasefi, J.B., Pirnejad, H., Wiil, U.K.: Monkeypox detection using deep neural networks. BMC Infectious Diseases 23, 14712334 (2023) https://doi.org/10.1186/s12879-023-08408-4

[17] Chen, J., Han, J.: A study on the recognition of monkeypox infection based on deep convolutional neural networks. Frontiers in Immunology 14, 16643224 (2023) https://doi.org/10.3389/fimmu.2023.1225557

[18] Yadav, A.: A study of diagnosis of skin disease using deep learning techniques. International Journal for Research in Applied Science and Engineering Technology 9, 771–774 (2021) https://doi.org/10.22214/IJRASET.2021.33310

[19] Cao, X., Ye, W., Moise, K., Quisqueya, U., Coffee, H.M.: Mpoxvlm: A visionlanguage model for diagnosing skin lesions from mpox virus infection. Proceedings of Machine Learning Research 259, 1–13 (2024) https://doi.org/10.48550/arXiv. 2411.10888

[20] Islam, T., Hussain, M.A., Chowdhury, F.U.H., Islam, B.M.R.: Can artificial intelligence detect monkeypox from digital skin images? bioRxiv (2022) https: //doi.org/10.1101/2022.08.08.503193

[21] Haque, M.E., Ahmed, M.R., Nila, R.S., Islam, S.: Human monkeypox disease detection using deep learning and attention mechanisms. In: 2022 25th International Conference on Computer and Information Technology (ICCIT), pp. 1069–1073 (2022). https://doi.org/10.1109/ICCIT57492.2022.10055870

[22] Vuran, S., Ucan, M., Akin, M., Kaya, M.: Multi-classification of skin lesion images including mpox disease using transformer-based deep learning architectures. Diagnostics 15, 20754418 (2025) https://doi.org/10.3390/diagnostics15030374
[23] Aloraini, M.: An effective human monkeypox classification using vision transformer. International Journal of Imaging Systems and Technology 34, 22944 (2024) https://doi.org/10.1002/IMA.22944
[24] Muthulakshmi, A., Prasad, C.H.V.S., Balachandran, G., Ranjith, S.: Optimized global aware siamese network based monkeypox disease classification using skin images. Biomedical Signal Processing and Control 101, 107125 (2025) https:// doi.org/10.1016/J.BSPC.2024.107125
[25] Sitaula, C., Shahi, T.B.: Monkeypox virus detection using pre-trained deep learning-based approaches. Journal of Medical Systems 46, 15736896 (2022) https://doi.org/10.1007/s10916-022-01868-2
[26] Oztel, G.Y.: Vision transformer and cnn-based skin lesion analysis: classification of monkeypox. Multimedia Tools and Applications 83, 71909–71923 (2024) https: //doi.org/10.1007/s11042-024-19757-w
[27] Asif, S., Zhao, M., Li, Y., Tang, F., Ur Rehman Khan, S., Zhu, Y.: Ai-based approaches for the diagnosis of mpox: challenges and future prospects. Archives of Computational Methods in Engineering 31(6), 3585–3617 (2024) https://doi.org/10.1007/s11831-024-10091-w
[28] Khan, S.H., Iqbal, R., Naz, S.: A recent survey of the advancements in deep learning techniques for monkeypox disease detection. arXiv preprint arXiv:2311.10754, 56–76 (2023) arXiv:2311.10754 [eess.IV]
[29] Alarfaj, A.A., Ahmad, S., Hakeem, A.M., Alabdulqader, E.A., PERO, C., Alsubai, S., Innab, N., Ashraf, I.: Novel vision transformer and data augmentation technique for efficient detection of monkeypox disease. Multimedia Tools and Applications 10, 1–9 (2024) https://doi.org/10.1007/s11042-024-20456-9
[30] Kaggle: Mpox Skin Lesion Dataset Version 2.0 (MSLD v2.0). Online. [Accessed: Mar. 27, 2025] (2025). https://www.kaggle.com/datasets/joydippaul/ mpox-skin-lesion-dataset-version-20-msld-v20/code
[31] Sun, J., Yao, K., Huang, G., Zhang, C., Leach, M., Huang, K., Yang, X.: Machine learning methods in skin disease recognition: a systematic review. Processes 11(4), 1003 (2023) https://doi.org/10.3390/pr11041003