City Research Online

City, University of London Institutional Repository

Citation: Raha, A. D., Gain, M., Saha, S. K., Rahman, M. S., Adhikary, A., Rameswar, D., Bairagi, A. K. & Biswas, S. (2025). Resource-Aware Monkeypox Diagnosis: Leveraging High-Capacity and Lightweight Models with Knowledge Distillation. Paper presented at the 17th International Conference on COMmunication Systems and NETworks (COMSNETS), 6-10 Jan 2025, Bengaluru, India. doi: 10.1109/COMSNETS63942.2025.10885586

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/34828/

Link to published version: https://doi.org/10.1109/COMSNETS63942.2025.10885586

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Resource-Aware Monkeypox Diagnosis: Leveraging High-Capacity and Lightweight Models with Knowledge Distillation

Avi Deb Raha[1,2], Mrityunjoy Gain[1,2], Sudip Kumar Saha[3], Md Shohanur Rahman[4], Apurba Adhikary[5], Rameswar Debnath[1], Anupam Kumar Bairagi[1,*], Sujit Biswas[6,*]

[1]*Computer Science and Engineering Discipline, Khulna University, Khulna, Bangladesh*

[2]*AI Innovation Lab, Khulna, Bangladesh*

[3]*United International University, Dhaka, Bangladesh*

[4]*Dept. of Mechatronics Engineering, KUET, Khulna, Bangladesh*

[5]*Dept. of ICE, Noakhali Science and Technology University, Noakhali, Bangladesh*

[6]*Computer Science Department, City, University of London, EC1V 0HB, UK*

E-mail: dev1611@cseku.ac.bd, gain1624@cseku.ac.bd, sudipsaha.kuet@gmail.com, shohan@mte.kuet.ac.bd, apurba@nstu.edu.bd, rdebnath@cseku.ac.bd , anupam@cseku.ac.bd, Sujit.biswas@city.ac.uk.

***Abstract*—Monkeypox, a novel zoonotic disease akin to smallpox, necessitates prompt and accurate diagnosis for effective treatment. Conventional diagnostic techniques, such as Polymerase Chain Reaction (PCR), offer high precision but require specialized equipment and trained personnel, rendering them impractical in resource-limited settings. Existing deep learning approaches for monkeypox diagnosis have predominantly relied on single, resource-intensive models, prioritizing accuracy over deployment feasibility across diverse computing platforms. In this study, we present a resource-conscious model deployment strategy that balances diagnostic accuracy with computational efficiency, enabling precise diagnosis in both resource-rich environments, such as hospitals, and resource-constrained contexts. We utilize a pretrained ConvNeXt-B model, trained on the extensive ImageNet-22K dataset, for deployment in resource-abundant scenarios, and a SqueezeNet model, optimized for resource-limited devices using the ImageNet-1K dataset. To enhance the performance of the lightweight SqueezeNet model without increasing computational complexity, we apply Knowledge Distillation. Experimental results demonstrate that the ConvNeXt-B model achieves an accuracy of 95.75%, which is 3.47% higher than the previous studies. Similarly, the Knowledge Distillation-enhanced SqueezeNet model attains an accuracy of 91.89%, representing a 2.32% improvement over the baseline. This dual-model approach ensures that accurate monkeypox diagnostics are accessible across a wide range of computational environments, thereby supporting more effective outbreak management and contributing to improved public health outcomes.**

***Index Terms*—Monkeypox detection, Knowledge Distillation, ConvNext-B, Lightweight Model**

I. Introduction

Monkeypox is an emerging zoonotic viral disease caused by the Monkeypox virus, a member of the Orthopoxvirus genus [1]. The disease presents symptoms similar to smallpox—such as fever, rash, and swollen lymph nodes—but is generally less severe. Recent outbreaks have highlighted its potential for widespread transmission, posing significant public health challenges [2]. Accurate and timely diagnosis of Monkeypox is crucial for effective disease management and containment. The current gold standard for diagnosis is the Polymerase Chain Reaction (PCR) test, which detects viral DNA with high specificity and sensitivity [3]. However, PCR testing faces several limitations, particularly in resource-constrained settings. The requirement for specialized laboratory equipment and trained personnel—often unavailable in remote or under-resourced areas—leads to delays as samples must be transported to centralized laboratories [3]. Additionally, the high cost and infrastructural demands limit the accessibility and scalability of PCR testing during large-scale outbreaks [4].

In light of these challenges, there is a pressing need for alternative diagnostic approaches that are both rapid and accessible [2]. Deep learning, a subset of artificial intelligence, has shown remarkable promise in medical image analysis, particularly in the classification and detection of skin lesions [5]. Convolutional Neural Networks (CNNs) are adept at learning complex patterns and features from large datasets [6], enabling them to distinguish between different dermatological conditions with high accuracy [7]. Leveraging deep learning for Monkeypox classification could facilitate the development of automated diagnostic tools that operate quickly and efficiently, reducing reliance on laboratory-based PCR tests.

Several recent studies have explored the use of deep learning architectures for the detection of Monkeypox [2], [3], [8], [9]. These works have demonstrated the feasibility of employing CNNs for automatic Monkeypox diagnosis. However, most of these studies utilize a single architecture, often pretrained on ImageNet-1K, which may not fully capture the complexity of the task or accommodate varying computational resources. In real-world scenarios, computational resources vary significantly—from high-performance servers in hospitals and cloud infrastructures to resource-constrained mobile devices like smartphones. Therefore, there is a need for multiple architectures tailored to different deployment environments to ensure broad accessibility and effectiveness.

To address this gap, we propose a resource-aware deployment strategy employing two distinct architectures: a high-

* Corresponding Author: anupam@cseku.ac.bd, Sujit.biswas@city.ac.uk

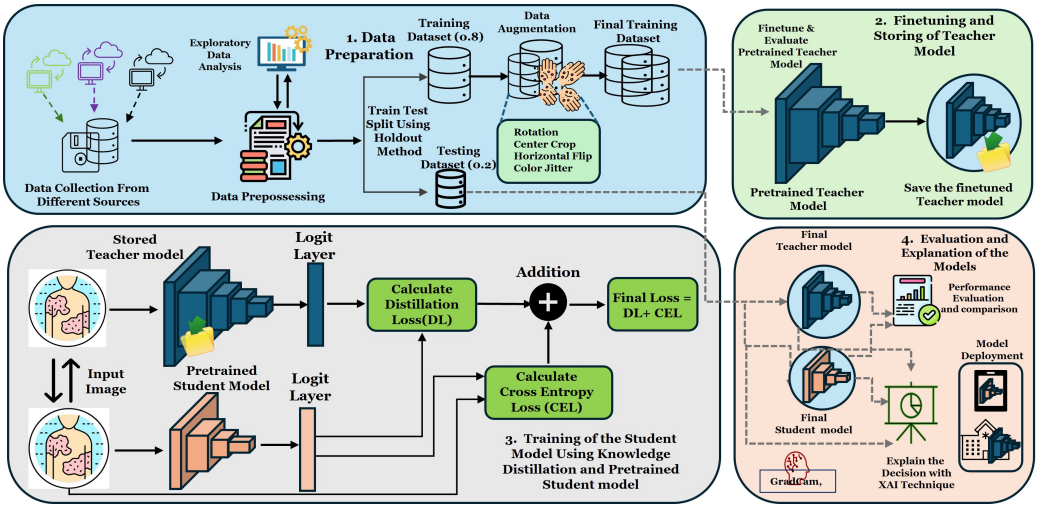


Fig. 1: System Model for Resource-Aware Approach with Knowledge Distillation for Monkeypox Diagnosis.

capacity ConvNeXt-B model pretrained on ImageNet-22K for resource-rich settings like hospitals, and a lightweight SqueezeNet model optimized for mobile devices. The ConvNeXt-B model leverages hospital computational resources to deliver reliable diagnostics, while the SqueezeNet model ensures efficient performance on devices with limited processing power, such as smartphones. This approach enables accurate diagnostics in both centralized healthcare facilities and at the point of care. To enhance the SqueezeNet model's performance without increasing computational complexity, we apply Knowledge Distillation (KD). In our framework, the ConvNeXt-B model serves as the teacher, providing informative outputs that the SqueezeNet student model leverages to improve classification performance efficiently.

This strategy is pivotal in ensuring that accurate diagnostics are accessible across diverse deployment environments. The ConvNeXt-B model delivers reliable and precise results in high-resource settings such as hospitals, where computational power is abundant. Conversely, the optimized SqueezeNet model offers swift and dependable performance on resource-constrained mobile devices like smartphones. This flexibility is especially critical during outbreaks, as it facilitates rapid and widespread deployment of diagnostic tools. By catering to both centralized and field-based settings, our approach can enhances disease control and prevention efforts, ensuring timely and effective responses to public health challenges. The main contributions of this paper are as follows:

1) **Resource Aware Deployment Strategy for Diverse Computational Environments**: We employ a resource-aware deployment strategy by leveraging two distinct deep learning architectures: the high-capacity ConvNeXt-B model, pretrained on ImageNet-22K, tailored for resource-rich environments such as hospitals, and the lightweight SqueezeNet model, optimized for edge devices like mobile phones through Knowledge Distillation. This approach ensures that accurate Monkeypox diagnostics are accessible across centralized healthcare facilities and remote, field-based settings, thereby enhancing diagnostic reach and adaptability.
2) **Performance Enhancement of Lightweight Model via Knowledge Distillation**: We apply Knowledge Distillation (KD) to distill the knowledge from the ConvNeXt-B model to the SqueezeNet model, significantly improving the SqueezeNet's classification accuracy from 89.57% to 91.89% without increasing its computational complexity.
3) **Model Interpretability with Grad-CAM Visualizations**: We incorporate Grad-CAM to provide visual explanations of the models' predictions. This enhances the interpretability and trustworthiness of the diagnostic decisions made by both the ConvNeXt-B and the enhanced SqueezeNet models.
4) **Comprehensive Evaluation on EMSID Dataset**: We conduct extensive evaluations of both models on the Extended Monkeypox Skin Images Dataset (EMSID), demonstrating their efficacy in real-world diagnostic scenarios. The ConvNeXt-B model achieves an accuracy of 95.75%, while the fine-tuned and distilled SqueezeNet model attains a competitive accuracy of 91.89%, surpassing the baseline SqueezeNet variant.

II. Related Work

The application of deep learning techniques for the detection and classification of monkeypox using skin lesion images has gained significant attention in recent years. Various convolutional neural network (CNN) architectures have been explored to address this challenge, including ResNet-50, VGG-19, InceptionV3, and MobileNetV2 [10]. These models have been employed to analyze image data with the aim of providing rapid and accurate diagnostics. Ali et al. [3] adapted and trained three CNN models including VGG-16, ResNet-50, and InceptionV3 using the Monkeypox Skin Lesion Database (MSLD). Due to the limited availability of data, they applied data augmentation techniques to expand the training set. Their results indicated that ResNet-50 achieved the highest accuracy.

Bala et al. [8] addressed the data scarcity issue by developing the Monkeypox Skin Images Dataset (MSID), segmented

into four classes: "Monkeypox," "Chickenpox," "Measles," and "Normal." They introduced 'MonkeyNet,' a customized DenseNet-201 CNN architecture. The model was rigorously tested on both the original and augmented versions of the MSID, achieving a diagnostic accuracy of 93.19% on the original dataset and 98.91% on the augmented dataset.

To enhance classification performance further, Ahsan et al. [11] proposed a Generalization and Regularization-based Transfer Learning Approach (GRA-TLA) designed for both binary and multi-class classification of monkeypox. For multi-class classification, the ResNet-101 model achieved impressive accuracy ranges from 84% to 99%. Kundu et al. [12] conducted a comparative analysis between classical machine learning models (K-NN and SVM) and deep learning models (ViT and ResNet-50) on a dataset comprising 1,300 images. The K-NN model achieved an accuracy of 84%, while the ViT outperformed the others with an accuracy of 93%.

Attention mechanisms have also been applied to enhance model performance in monkeypox detection. Raha et al. [2] aimed to develop a more generalized model suitable for edge deployment by utilizing an extended version of the MSID dataset (EMSID). They employed a lightweight deep learning architecture based on MobileNetV2, enhanced with both spatial and channel attention mechanisms. This approach aimed to balance resource efficiency with high diagnostic accuracy. The attention-based MobileNetV2 model demonstrated impressive results, achieving 92.28% accuracy on the EMSID dataset, 98.19% on the original MSID dataset, and 93.33% on the MSLD dataset. However, despite its lightweight design, MobileNetV2 still comprises approximately 3.2 million parameters, which can be burdensome for deployment on devices with very limited computational capabilities. Most of the previous studies have concentrated on single-model approaches, optimizing either for high accuracy or computational efficiency but not both. Additionally, many models rely on pretraining with the ImageNet-1K dataset, which may not fully capture the complexity of Monkeypox detection tasks. The variability in computational resources across different deployment environments—from high-performance servers in hospitals to resource-constrained mobile devices—further necessitates the development of multiple architectures tailored to these diverse settings.

III. Methodology

In this section, we describe the architecture of our proposed system, including the models used, the dataset, and the training procedures. Fig. 1 shows the overall mechanism of this study.

A. Dataset

We utilized the Extended Monkeypox Skin Images Dataset (EMSID) [2], which combines the original Monkeypox Skin Images Dataset (MSID) [8] with images of similar skin conditions from DermNet. EMSID includes eight classes: Monkeypox, Chickenpox, Measles, Eczema, Lupus, Molluscum contagiosum, Scabies, and Normal. Each condition presents distinct skin features—such as Monkeypox's pustular rash, Chickenpox's itchy blisters, and Lupus's butterfly-shaped rash—that

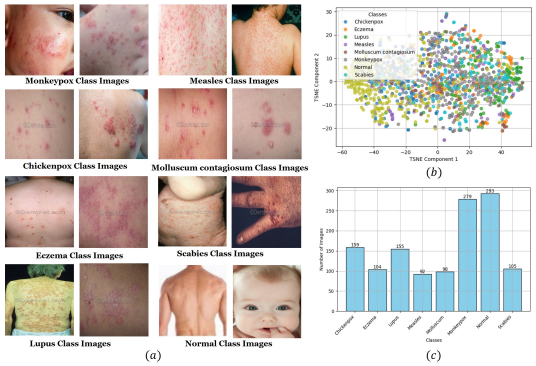


Fig. 2: EMSID Dataset visualization. (a) Instances of EMSID (b) t-SNE visualization of feature embeddings (c) Data Distribution in the EMSID.

can easily be confused by the untrained eye, increasing the risk of misdiagnosis. The dataset comprises 1,285 images, with Monkeypox and Normal classes representing approximately 21.7% and 22.8% of the instances, respectively. Lupus and Chickenpox each account for around 12%, while Eczema, Scabies, Measles, and Molluscum contagiosum range between 8% and 12%. Fig. 2 illustrates the dataset, where Fig. 2 (a) shows sample images from EMSID, Fig. 2(b) presents a t-SNE visualization of feature embeddings extracted from the images, and Fig. 2 (c) provides a visual summary of the class distribution.

B. Data Preprocessing

For data preprocessing, we first divided the EMSID dataset into training and testing sets using the hold-out method with an 80:20 ratio. To enhance the diversity of the training data and improve model robustness, we applied data augmentation techniques, including random resized cropping, horizontal flipping, rotation, and color jittering. The augmentation process generated six additional versions for each original image, effectively increasing the dataset size. After augmentation, each class comprised the following number of instances: Chickenpox (889), Eczema (1,205), Lupus (1,798), Measles (511), Molluscum contagiosum (1,134), Monkeypox (1,561), Normal (1,638), and Scabies (1,218).

C. Model Architectures

1) High-Performance Deployment: To achieve superior performance, we explored the ConvNeXt-B and Swin-B models [13]. ConvNeXt-B, in particular, has demonstrated exceptional capability in handling complex datasets across various scenarios [14]. Pre-trained on the extensive ImageNet-22K dataset, ConvNeXt-B's deep architecture and substantial capacity have proven highly effective for processing complex, high-resolution images. In our specific settings, ConvNeXt-B also outperformed other models, solidifying its suitability for our task. The architecture of a ConvNeXt-B block is represented as follows:

$$\mathbf{Y} = \mathbf{X} + \mathcal{F}(\mathbf{X}) \tag{1}$$

$$\mathcal{F}(\mathbf{X}) = \mathrm{Cv}_{1\times1}\left(\mathrm{GELU}\left(\mathrm{Cv}_{1\times1}\left(\mathrm{LN}\left(\mathrm{DC}_{7\times7}(\mathbf{X})\right)\right)\right)\right) \tag{2}$$

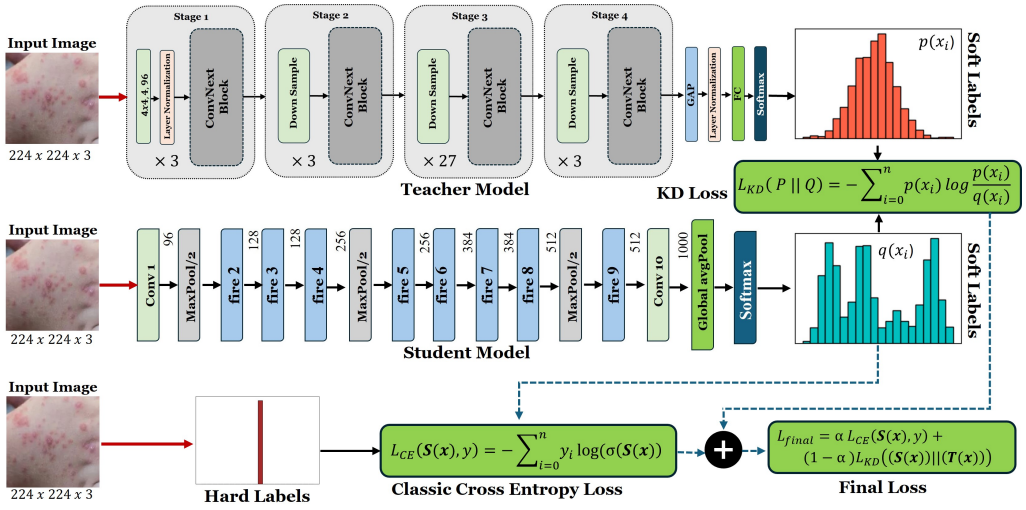


Fig. 3: Knowledge Distillation Framework: Distilling Knowledge from ConvNeXt-B to SqueezeNet.

Here, $\mathrm{DC}_{7\times7}$ denotes a depthwise convolution layer utilizing a 7×7 kernel, which captures spatial dependencies in the input feature map. LN refers to layer normalization, a technique that stabilizes and accelerates the training process by normalizing the inputs across the features. $\mathrm{Cv}_{1\times1}$ signifies a pointwise convolution, which adjusts the number of feature channels without altering the spatial dimensions and GELU stands for Gaussian Error Linear Unit, an activation function that introduces non-linearity into the model. The output feature map $\mathbf{Y}$ incorporates a residual connection from the input $\mathbf{X}$, facilitating the flow of information and mitigating issues related to vanishing gradients during training.

2) Resource Efficient Deployment: To address deployment in environments with limited computational resources, we incorporate the SqueezeNet architecture [15] into our framework. SqueezeNet is a lightweight convolutional neural network designed to achieve high accuracy with significantly fewer parameters compared to traditional models, making it ideal for deployment on mobile and edge devices. The fundamental building block of SqueezeNet, known as the Fire module, operates based on the following formulation:

$$\mathbf{Y} = \mathbf{X} + \mathcal{F}(\mathbf{X}) \tag{3}$$

$$\mathcal{F}(\mathbf{X}) = \text{Expand}\left(\text{Squeeze}(\mathbf{X})\right). \tag{4}$$

In these equations. $\text{Squeeze}(\mathbf{X})$ represents a convolutional layer with 1×1 kernels that reduces the number of input channels, effectively compressing the feature maps. $\text{Expand}(\cdot)$ consists of two parallel convolutional layers with 1×1 and 3×3 kernels, respectively. These layers increase the number of channels to capture a diverse set of features.

The output feature map $\mathbf{Y}$ is obtained by adding the input $\mathbf{X}$ to the processed output from the Fire module $\mathcal{F}(\mathbf{X})$, thereby incorporating a residual connection. This residual connection facilitates better gradient flow during training and enhances the network's ability to learn complex representations without a substantial increase in computational overhead. To further enhance the performance of the lightweight SqueezeNet model, we employ Knowledge Distillation (KD).

$$\mathcal{L}_{\text{KD}}(\mathbf{S}(\mathbf{x}), \mathbf{T}(\mathbf{x})) = \text{KL}\left(\sigma\left(\frac{\mathbf{T}(\mathbf{x})}{T}\right) \middle\| \sigma\left(\frac{\mathbf{S}(\mathbf{x})}{T}\right)\right) \times T^2 \tag{5}$$

D. Knowledge Distillation

KD is a model compression technique introduced by Hinton et al. [16], which enables a smaller student model to learn from a larger, well-trained teacher model. In our framework, we leverage KD to enhance the performance of the lightweight SqueezeNet model, denoted as $\mathbf{S}$, by transferring knowledge from the high-capacity ConvNeXt-B model, denoted as $\mathbf{T}$. The core idea of KD is to train the student model using a combination of the ground truth labels and the soft target distributions produced by the teacher model. This allows the student model to capture nuanced information learned by the teacher, including inter-class similarities that are not evident from the hard labels alone. The training objective for the student model involves minimizing a total loss function $\mathcal{L}_{\text{total}}$, which is a weighted sum of the standard cross-entropy loss $\mathcal{L}_{\text{CE}}$ and the distillation loss $\mathcal{L}_{\text{KD}}$:

$$\mathcal{L}_{\text{total}} = \alpha\, \mathcal{L}_{\text{CE}}(\mathbf{S}(\mathbf{x}), \mathbf{y}) + (1-\alpha)\, \mathcal{L}_{\text{KD}}(\mathbf{S}(\mathbf{x}), \mathbf{T}(\mathbf{x})) \tag{6}$$

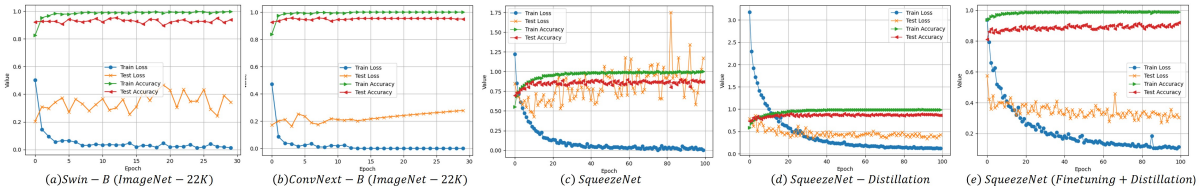
In this equation, $\mathbf{x}$ represents the input data, and $\mathbf{y}$ denotes the true labels corresponding to $\mathbf{x}$. The parameter $\alpha \in [0, 1]$ is a weighting factor that balances the importance of the two loss terms. The functions $\mathbf{S}(\mathbf{x})$ and $\mathbf{T}(\mathbf{x})$ denote the logits (pre-softmax outputs) of the student and teacher models, respectively. The cross-entropy loss $\mathcal{L}_{\text{CE}}$ is defined as:

$$\mathcal{L}_{\text{CE}}(\mathbf{S}(\mathbf{x}), \mathbf{y}) = -\sum_i y_i \, \log \sigma_i(\mathbf{S}(\mathbf{x})) \tag{7}$$

Here, y_i is the ground truth label for class i, and $\sigma_i(\mathbf{S}(\mathbf{x}))$ is the softmax probability of class i produced by the student model, computed as:

$$\sigma_i(\mathbf{S}(\mathbf{x})) = \frac{\exp(S_i(\mathbf{x}))}{\sum_j \exp(S_j(\mathbf{x}))} \tag{8}$$

In this expression, $S_i(\mathbf{x})$ is the logit output of the student model for class i, and the denominator sums over all classes j to normalize the probabilities. The distillation loss $\mathcal{L}_{\text{KD}}$ measures the divergence between the softened output distributions



(a)Swin − B (ImageNet − 22K) (b)ConvNext − B (ImageNet − 22K) (c) SqueezeNet (d) SqueezeNet − Distillation (e) SqueezeNet (Finetuning + Distillation)

Fig. 4: Performance curves for training and testing across multiple models.

(a) Swin − B (ImageNet − 22K) (b) ConvNext − B (ImageNet − 22K) (c) SqueezeNet (d) SqueezeNet − Distillation (e) SqueezeNet (Finetuning + Distillation)

Fig. 5: Confusion matrix for different models.

of the teacher and student models, using the Kullback-Leibler (KL) divergence:

In this context, T is the temperature parameter that controls the softness of the probability distributions. Higher values of T produce softer probabilities, allowing the student model to learn from the relative probabilities assigned to incorrect classes. The softened softmax function $\sigma\left(\frac{\mathbf{z}}{T}\right)$ is computed as:

$$\sigma_i\left(\frac{\mathbf{z}}{T}\right) = \frac{\exp\left(\frac{z_i}{T}\right)}{\sum_j \exp\left(\frac{z_j}{T}\right)} \tag{9}$$

where $\mathbf{z}$ represents the logits vector (either $\mathbf{T}(\mathbf{x})$ for the teacher or $\mathbf{S}(\mathbf{x})$ for the student), and z_i is the logit for class i.

The KL divergence $\mathrm{KL}(P\|Q)$ between two probability distributions P and Q is defined as:

$$\mathrm{KL}(P\|Q) = \sum_i P_i \log\left(\frac{P_i}{Q_i}\right) \tag{10}$$

This divergence quantifies how one probability distribution Q diverges from a reference distribution P. The temperature scaling factor T^2 in the distillation loss accounts for the gradients' magnitude during backpropagation, ensuring that the contribution of $\mathcal{L}_{\mathrm{KD}}$ remains appropriately scaled relative to $\mathcal{L}_{\mathrm{CE}}$ [16]. By minimizing the total loss $\mathcal{L}_{\mathrm{total}}$, the student model learns to mimic the teacher's output distribution while also fitting the true labels. This dual training objective enables the student model to generalize better and achieve higher accuracy without increasing its complexity. In our implementation, we set the temperature $T = 5$ and the weighting factor $\alpha = 0.3$ based on empirical studies and task-specific tuning. Fig. 3 shows the distillation procedure from ConvNext-B to SqueezeNet model.

IV. Experimental Setup and Result Analysis

A. Experimental Setup

In our experimental setup, we utilized the PyTorch framework (version 2.0.1) for model implementation and training, running on an NVIDIA RTX 3080 Ti GPU with 12 GB of memory and an Intel Core i9 processor. The Adam optimizer was employed with a learning rate of 0.0001 and a batch size of 32 to ensure efficient memory utilization and optimal convergence speed. All input images were resized to 224×224 pixels to match the dimensions required for models pretrained on the ImageNet dataset. For training protocols, larger models were trained for 30 epochs, whereas distillation experiments were extended to 100 epochs to ensure adequate learning. Specifically, for the SqueezeNet variants in this study, the standard SqueezeNet and SqueezeNet-Distillation were each trained for 100 epochs, while the SqueezeNet-Distillation-Finetuning variant underwent an initial 30 epochs of fine-tuning followed by 100 epochs of distillation.

B. Numerical Results

Fig. 4 shows the training and testing loss and accuracy for six different models. In Fig. 4 (c), SqueezeNet demonstrates notable fluctuations in the test loss, suggesting instability in generalization during training. However, with the introduction of KD, the training and testing curves become more stable, particularly for the distilled versions of SqueezeNet(Fig. 4 (d), Fig. 4 (e)). This stability can be attributed to the distillation process, where the student model (SqueezeNet) learns from the softer, more informative predictions of the teacher model (ConvNeXt-B). This reduces overfitting by guiding the student model to focus on broader patterns in the data, rather than being overly influenced by noise in the training data. Fig. 5 highlights the confusion matrices for each model, providing a detailed view of class-level performance. As summarized in Table I, ConvNeXt-B, pretrained on ImageNet-22K, consistently outperforms other models with an accuracy of 95.75%, precision of 96.22%, and a specificity of 99.42%. Swin-B performs similarly, though slightly behind ConvNeXt-B. SqueezeNet, despite its smaller size, achieves competitive results, especially when enhanced with distillation, reaching an accuracy of 91.89% and a specificity of 98.82%. This demonstrates that Knowledge Distillation effectively enhances lightweight models like SqueezeNet, enabling them to perform well in resource-constrained environments while maintaining computational efficiency. Figure 6 showcases the Grad-CAM visualizations for both the ConvNeXt-B and SqueezeNet-Distillation-Balance models, illustrating the specific regions of the input images that each model prioritizes during pre-

TABLE I: Performance of deep learning models on the EMSID dataset for Monkeypox detection.

Model	Pretrained Dataset	Parameters	Accuracy	Precision	Recall	F1 Score	Specificity
ResNet-152 [2]	ImageNet-1K	60.2M	0.9147	0.8959	0.9040	0.8983	0.9880
Attention-MobileNetv2 [2]	ImageNet-1K	3.7M	0.9228	0.9048	0.8942	0.8984	0.9890
SqueezeNet (This Study)	ImageNet-1K	0.7M	0.8957	0.8791	0.8918	0.8837	0.9852
SqueezeNet-Distillation (This Study)	ImageNet-1K	0.7M	0.9034	0.9027	0.8858	0.8876	0.9859
SqueezeNet-Distillation-Balance (This Study)	ImageNet-1K	0.7M	0.9189	0.9151	0.9090	0.9096	0.9882
Swin-B (This Study)	ImageNet-22K	87M	0.9536	0.9521	0.9528	0.9506	0.9931
ConvNext-B (This Study)	ImageNet-22K	89M	**0.9575**	**0.9622**	**0.9609**	**0.9608**	**0.9942**

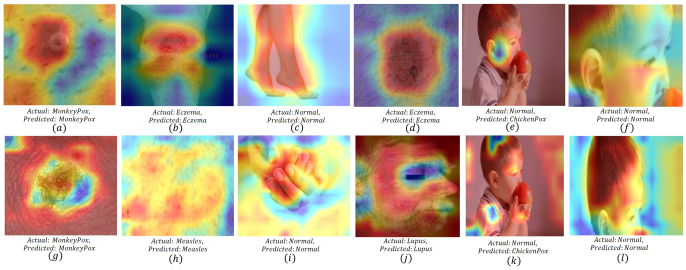


Fig. 6: GradCam Visualization For ConvNext-B and SqueezeNet Model.

diction. The first row displays the ConvNeXt-B model, while the second row presents the SqueezeNet-Distillation-Balance model. By utilizing these Grad-CAM visualizations, users can verify whether the models are concentrating on the relevant areas of the images, thereby ensuring the reliability of the diagnostic results. For example, in Fig. 6(e) and (k), both models erroneously focus on incorrect regions, leading to misclassifications. However, after adjusting the image angle, the models successfully predict the correct class, demonstrating that proper image capture techniques significantly enhance model performance. This capability allows users to assess and confirm that the models are making informed decisions based on appropriate image features, thereby providing assurance of the diagnostic accuracy in various imaging conditions.

V. Conclusion

This study presents a dual-model deployment strategy to enhance Monkeypox diagnosis across diverse computational environments. The high-capacity ConvNeXt-B model delivers superior diagnostic accuracy in resource-rich settings, achieving 95.75% accuracy, making it suitable for hospitals and centralized healthcare facilities. The lightweight SqueezeNet model, optimized for mobile devices and enhanced via Knowledge Distillation, attains a competitive accuracy of 91.89%, effectively operating in resource-constrained settings. Evaluations on the EMSID dataset confirm the efficacy of this approach in accurately diagnosing Monkeypox and distinguishing it from similar skin conditions. By providing a flexible and scalable diagnostic solution, this work has significant implications for public health, enabling rapid and accurate disease detection both centrally and in the field. Future research could explore integrating this diagnostic tool into telemedicine platforms and applying federated learning to enhance data privacy and security, further improving healthcare outcomes in resource-limited settings.

References

[1] World Health Organization, "Monkeypox," https://www.who.int/news-room/fact-sheets/detail/monkeypox, 2022, [Online; accessed 26-10-2024].

[2] A. D. Raha, M. Gain, R. Debnath, A. Adhikary, Y. Qiao, M. M. Hassan, A. K. Bairagi, and S. M. S. Islam, "Attention to monkeypox: An interpretable monkeypox detection technique using attention mechanism," *IEEE Access*, vol. 12, pp. 51 942–51 965, 2024.

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

[4] S. Chakraborty, "Democratizing nucleic acid-based molecular diagnostic tests for infectious diseases at resource-limited settings–from point of care to extreme point of care," *Sensors & Diagnostics*, vol. 3, no. 4, pp. 536–561, 2024.

[5] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[6] M. Gain, A. D. Raha, and R. Debnath, "Ccc: Color classified colorization," *arXiv preprint arXiv:2403.01476*, 2024.

[7] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *International Journal of Computer Vision*, vol. 129, no. 6, pp. 1789–1819, 2021.

[8] D. Bala, M. S. Hossain, M. A. Hossain, M. I. Abdullah, M. M. Rahman, B. Manavalan, N. Gu, M. S. Islam, and Z. Huang, "Monkeynet: A robust deep convolutional neural network for monkeypox disease detection and classification," *Neural Networks*, vol. 161, pp. 757–775, 2023.

[9] F. J. Dihan, S. A. Murad, A. J. M. Muzahid, K. Uddin, M. J. Alenazi, A. K. Bairagi, and S. Biswas, "Mpoxsldnet: A novel cnn model for detecting monkeypox lesions and performance comparison with pre-trained models," *arXiv preprint arXiv:2405.21016*, 2024.

[10] N. Nazmee, M. S. Ali, S. Mahmud, K. Alam, A. Chakrabarty, and M. Fahim-Ul-Islam, "Enhancing monkeypox diagnosis: A machine learning approach for skin lesion classification," in *2023 26th International Conference on Computer and Information Technology (ICCIT)*, 2023, pp. 1–6.

[11] M. M. Ahsan, M. R. Uddin, M. S. Ali, M. K. Islam, M. Farjana, A. N. Sakib, K. Al Momin, and S. A. Luna, "Deep transfer learning approaches for monkeypox disease diagnosis," *Expert Systems with Applications*, vol. 216, p. 119483, 2023.

[12] D. Kundu, U. R. Siddiqi, and M. M. Rahman, "Vision transformer based deep learning model for monkeypox detection," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2022, pp. 1021–1026.

[13] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 976–11 986.

[14] A. D. Raha, A. Adhikary, M. Gain, Y. Qiao, and C. S. Hong, "Boosting federated domain generalization: The role of advanced pre-trained architectures," *arXiv preprint arXiv:2409.13527*, 2024.

[15] Q. Huang, H. Ding, and M. Effatparvar, "Breast cancer diagnosis based on hybrid squeezenet and improved chef-based optimizer," *Expert Systems with Applications*, vol. 237, p. 121470, 2024.

[16] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.