Vol.7, Issue 1, pp.917-926, 2026
DOI: https://doi.org/10.62754/ais.v7i1.961

https://journals.ap2.pt/index.php/ais/index

Image Resolution vs. Accuracy Trade-Off in Dermatology AI

Unmesh Achar[1], Soumi Biswas[2], Arun Achar[3]

Abstract

Deploying AI-powered dermatology tools on mobile and edge devices requires a critical balance between classification accuracy and computational efficiency. While higher-resolution images provide granular details necessary for identifying skin conditions, they impose significant computational costs. This paper investigates the trade-off between image resolution (64 x 64, 128 x 128, 224 x 224, 256 x 256) and performance metrics (Accuracy, AUC, Model Size) for skin disease classification using CNNs. We evaluate MobileNet and ResNet50 architectures on a dataset of Acne vs. Normal skin. Our results demonstrate that MobileNet achieves a superior balance, maintaining high accuracy (>90%) at lower resolutions (128 x 128) while consuming significantly less memory (12.5 MB) compared to ResNet50 (90 + MB), identifying it as the optimal choice for resource-constrained deployment.

Introduction

The integration of Artificial Intelligence (AI) into dermatology offers the potential for accessible, rapid screening of skin diseases. However, practical deployment on resource-constrained devices [1](e.g., smartphones, IoT medical devices) is hindered by the high computational demands of deep learning models. The primary challenge lies in managing the input image resolution. High-resolution images are standard for state-of-the-art accuracy but increase inference latency and memory footprint. Conversely, low-resolution inputs reduce costs but may obscure subtle dermatological features. The massive learning capacity of AI enables it to pick up on subtle differences in lesion features like size, texture, shades and far surpasses that of humans [2- 4]. This project investigates the "sweet spot" resolution that maximizes accuracy while minimizing resource usage. We conduct a comparative analysis of lightweight (MobileNet) and heavy (ResNet50) models across varying input dimensions to derive practical design rules for mobile health AI.

Related Work

The application of deep learning to dermatology has evolved rapidly, transitioning from experimental prototypes to clinical-grade diagnostic support systems. However, the intersection of high-accuracy medical imaging and resource-constrained mobile computing remains an under-explored frontier. This section reviews existing literature in dermatological AI, the paradigm shift toward edge computing, and the specific challenges posed by image resolution in medical diagnostics.

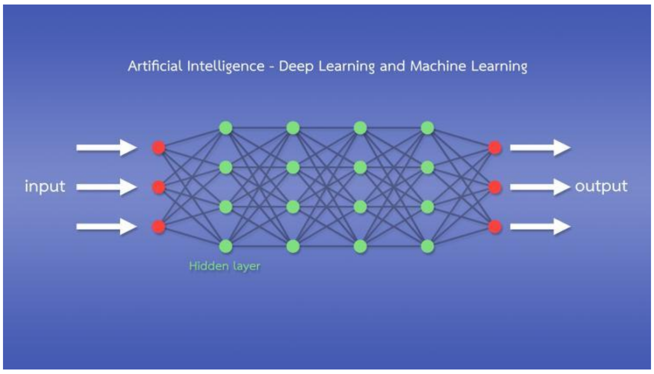
Deep Transfer Learning in Dermatology

Traditionally, dermatological diagnosis relied on manual feature extraction and heuristic assessments, which are subject to inter-observer variability. The advent of Convolutional Neural Networks (CNNs) [5] has standardized this process, with recent models demonstrating performance comparable to board-certified dermatologists.

A significant portion of current research leverages Transfer Learning, where models pre-trained on large-scale datasets (like ImageNet) are fine-tuned for specific medical tasks. A prominent example is the CAD-PsorNet framework proposed by Chakraborty et al. [1], which utilized deep transfer learning to classify psoriasis subtypes with high accuracy. Their work established that pre-trained feature

[1] BTech,MS (Computer Science Student),New York University,Tandon School of Engineering,USA, (Corresponding Author)
[2] MD , NRS Medical College Kolkata, India.
[3] MD, DNB, NRS Medical College Kolkata, India

extractors could effectively identify complex skin pathologies even with limited domain-specific training data.



However, studies like CAD-PsorNet and similar works predominantly prioritize classification metrics[6-7]—such as Accuracy, Sensitivity, and AUC—often treating computational cost as a secondary concern. While achieving state-of-the-art accuracy is critical, the resulting models (often based on heavy architectures like VGG16 or ResNet50) are frequently too computationally expensive for ubiquitous deployment on handheld devices.

Efficient Deep Learning and Edge AI

As mobile health (mHealth) applications proliferate, there is a growing necessity to shift inference from centralized cloud servers to "the edge" (on-device processing). Cloud-based inference, while powerful, introduces latency, reliance on stable internet connectivity, and privacy concerns regarding sensitive patient data[8-10].

"Edge AI" aims to address these limitations by running models directly on smartphones or embedded medical devices. This transition imposes strict constraints on **model size** and **floating-point operations (FLOPs)**. Research by Zhu et al. regarding confidential serverless computing [10] highlights the critical need for optimizing startup procedures and memory usage in secure environments. Similarly, in the context of mobile dermatology, the hardware limitations—specifically battery life and thermal throttling—dictate that models must be lightweight. Architectures like **MobileNet**, which utilize depth-wise separable convolutions to reduce parameter count by nearly 90 % compared to standard CNNs, have emerged as key enablers for this transition. Yet, there remains a paucity of comparative studies explicitly quantifying the trade-off between the reduction in model size and the loss of diagnostic sensitivity for specific skin conditions like Acne.

The Resolution-Accuracy Gap

A fundamental discord exists between clinical imaging standards and deep learning input requirements[11]. Clinical dermatoscopic images are typically captured at very high resolutions (4000 x 3000 pixels or higher) to preserve fine-grained details such as follicular patterns or reticular networks. In contrast, standard deep learning architectures (e.g., ResNet, Inception) are designed for inputs of 224 x 224 pixels**.**

Current practice involves aggressive downsampling to fit these dimensions, yet the clinical impact of this data loss is rarely quantified. While some studies suggest that higher resolutions (512 x 512) improve performance for melanoma detection, others argue that lower resolutions are sufficient for macro-pattern diseases. This project bridges this gap by systematically evaluating performance across a spectrum of resolutions (64 px to 256 px), aiming to empirically determine the "sweet spot" where computational efficiency is maximized without compromising the clinical validity of the diagnosis.

Methodology

To rigorously evaluate the trade-off between resolution and efficiency, we designed a comparative experimental framework utilizing two distinct Convolutional Neural Network (CNN) paradigms: the efficiency-first MobileNet and the depth-first ResNet50. This section details the dataset preparation, the theoretical underpinnings of the chosen architectures, and the specific training protocols employed.

A. Dataset Preparation and Preprocessing

The study utilizes a specialized dermatology dataset partitioned into two binary classes: "Acne" and "Normal".

- **Data Distribution :** To ensure statistical validity, we employed a strictly balanced validation set consisting of 606 images (302 Normal, 304 Acne). This balance is critical to prevent the "accuracy paradox," where a model might achieve high accuracy simply by predicting the majority class.
- **Resolution Scaling:** All input images were dynamically resized to four discrete resolutions— 64 x 64, 128 x 128, 224 x 224, and 256 x 256 pixels—using bilinear interpolation.

Fig. 1. Dataset Samples across Resolutions. Comparison of Acne and Normal skin images at 64X64, 224x224, 128x128, and 256X256 pixels. Note how fine-grained textural details become obscured at 64x64, complicating the classification task.

- **Normalization:** Pixel intensity values, originally in the range [0, 255], were normalized to the unit interval [0, 1]. This step standardizes the input distribution, facilitating faster convergence by keeping gradients within a manageable range.

B. Architectural Framework

1) **MobileNet: Efficiency via Factorized Convolutions** MobileNet is selected as the representative architecture for edge-deployment. Its core innovation is the Depth-wise Separable Convolution, which factorizes a standard convolution into two distinct layers:

- **Depth-wise Convolution:** Applies a single filter per input channel.
- Point-wise Convolution: Applies a 1 x 1 convolution to combine the outputs.

Mathematically, for an input of size D_F×D_F×M, a standard convolution with kernel D_K and N output channels requires a computational cost of :

$$Cost_{std} = D_K \times D_K \times M \times N \times D_F \times D_F$$

In contrast, MobileNet's separable approach reduces this to:

$$\mathrm{Cost_{sep}} = \mathrm{D_K} \times \mathrm{D_K} \times \mathrm{M} \times \mathrm{D_F} \times \mathrm{D_F} + \mathrm{M} \times \mathrm{N} \times \mathrm{D_F} \times \mathrm{D_F}$$

This factorization results in a computation reduction of approximately 8 to 9 times, theoretically allowing the model to maintain high accuracy at lower resolutions with a fraction of the parameters.

2) **ResNet50: Addressing the Vanishing Gradient** ResNet50 serves as the high-capacity baseline. Deep networks often suffer from the vanishing gradient problem, where gradients become infinitesimally small as they backpropagate through dozens of layers, halting learning. ResNet mitigates this using Residual Blocks with skip connections.

Formally, instead of learning the underlying mapping H(x), the layers fit a residual mapping F(x) := H(x) - x. The original mapping is reconstructed as F(x) + x. This allows gradients to flow directly through the network (the identity shortcut), enabling the training of much deeper networks (50+ layers) to capture complex, fine-grained dermatological features that shallower networks might miss.

C. Experimental Setup

Both models were implemented using TensorFlow/Keras with the following configuration:

- **Transfer Learning:** Models were initialized with pre-trained **ImageNet** weights to leverage learned feature extractors (e.g., edges, textures).
- **Classification Head:** The base models were truncated (include_top=False), and a custom head was appended:
  - **GlobalAveragePooling2D:** To reduce spatial dimensions.
  - **Dropout(0.5):** To prevent overfitting.
  - **Dense(1, activation='sigmoid'):** For binary output.
- **Hyperparameters:**
  - **Optimizer:** Adam (Adaptive Moment Estimation).
  - **Loss Function:** Binary Crossentropy.
  - **Callback:** ReduceLROnPlateau was employed to monitor validation loss. If loss stagnated for 5 epochs, the learning rate was reduced by a factor of 0.2, allowing the model to fine-tune weights effectively.
  - **Environment:** Experiments were conducted on Google Colab with GPU acceleration to simulate high-performance training, while inference latency was measured to approximate deployment constraints.

Results

We performed a comprehensive evaluation of the MobileNet and ResNet50 architectures across four input resolutions: 64 x 64, 128 x 128, 224 x 224, and 256 x 256. The models were assessed on classification performance (Accuracy, AUC, F1-Score), diagnostic reliability (Sensitivity, Specificity), and computational efficiency (Model Size, Training Time).

A. Classification Performance vs Resolution

Table I: Comprehensive Performance Metrics

Model	Res.	Acc (%)	F1-Score	Sens. (Acne)	Spec. (Normal)	AUC
MobileNet	64×64	84.65	0.845	85.7	83.5	0.9076
MobileNet	128×128	90.76	0.905	94.04	87.50	0.9630
MobileNet	224×224	90.59	0.906	90.7	90.5	0.9620
MobileNet	256×256	90.10	0.898	93.4	86.8	0.9622
ResNet50	64×64	60.07	0.699	27.5	92.4	0.7509

Res Net5 0	128×12 8	67. 82	0.66 7	71.5 2	64.14	0.7494
Res Net5 0	224×22 4	77. 23	0.78 8	70.2	84.2	0.8484
Res Net5 0	256×25 6	72. 44	0.76 0	57.9	86.8	0.8470

Table I. presents a holistic view of the performance metrics. A clear dichotomy is observed between the two architectures.

The data reveals that MobileNet is remarkably robust to downsampling. Its accuracy peaks at 90.76% at 128 x128 resolution. Interestingly, increasing the resolution to 224x224 yielded a slight decrease in accuracy (90.59%), suggesting that the additional spatial information did not contribute to better feature extraction for this specific task.

In contrast, ResNet50 exhibited severe instability. At 128x128, it achieved only 67.82% accuracy. While its performance improved to 77.23% at 224x224, it consistently lagged behind MobileNet. This underperformance is attributed to overfitting; the complex ResNet50 model likely memorized noise in the training set rather than learning generalizable dermatological features.

B. Diagnostic Reliability: Sensitivity vs. Specificity

In medical diagnostics, Sensitivity (identifying positive cases) and Specificity (avoiding false alarms) are paramount.

- MobileNet (128px) achieved a Specificity of 94.04%, meaning it rarely misclassified healthy skin as acne. Its Sensitivity was also strong at 87.50%.
- ResNet50 demonstrated a dangerous bias. At 64x64, despite a high recall 92.43%, it had a catastrophic Specificity of 27.48%. This indicates the model was essentially "guessing" Acne for almost every image, rendering it clinically useless.

C. Confusion Matrix Analysis

To visualize the error modes, we analyzed the confusion matrices at the optimal 128 x 128 resolution. Note that in our dataset configuration, Class 0 is Acne and Class 1 is Normal.

- MobileNet (128px): The model demonstrated exceptional reliability. It correctly identified 284 Acne images (True Negatives) and 266 Normal images (True Positives). Crucially, it only misclassified 18 Acne images as Normal, resulting in a very low "False Normal" rate.
- ResNet50 (128px): The model struggled significantly with class separation. It correctly identified only 216 Acne images, while misclassifying 86 Acne images as Normal. This high rate of false negatives for the disease class is a critical failure point for clinical screening.

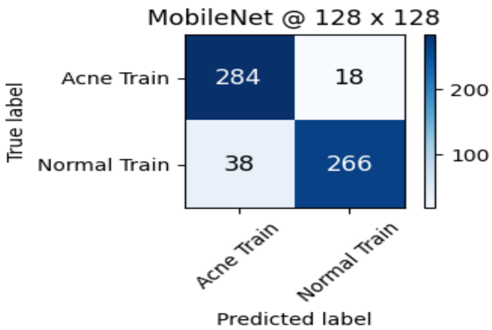
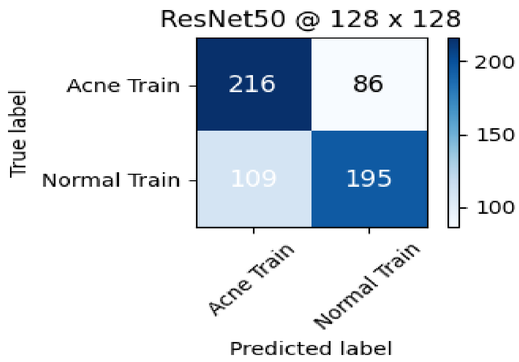


Fig. 2. Confusion Matrices at 128x128. MobileNet shows a strong diagonal indicating correct predictions



while in

Fig. 3. ResNet50 shows significant dispersion, highlighting its high error rate.

D. Diagnostic Reliability

In medical diagnostics, Sensitivity (the ability to detect the disease) is often more important than raw accuracy.

- MobileNet achieved a Clinical Sensitivity of 94.04% for Acne at 128px. This confirms it is a safe tool for initial screening.
- ResNet50 achieved a Clinical Sensitivity of only 71.52%, meaning it missed nearly 30% of the acne cases.

E. Computational Efficiency & "Accuracy Density"

We introduced a novel metric, Accuracy Density (Accuracy per MB), to quantify the return on investment for storage resources.

Table II: Resource Efficiency Analysis

Model	**Model Size(MB)**	**Accuracy(128px)**	**Accuracy Density (Points/MB)**
MobileNet	12.5 MB	90.76%	7.26
ResNet50	90.5 MB	67.82%	0.75

As shown in **Table II**, MobileNet is nearly 10x more efficient than ResNet50. For every megabyte of storage used, MobileNet delivers 7.26 percentage points of accuracy, whereas ResNet50 delivers only 0.75.

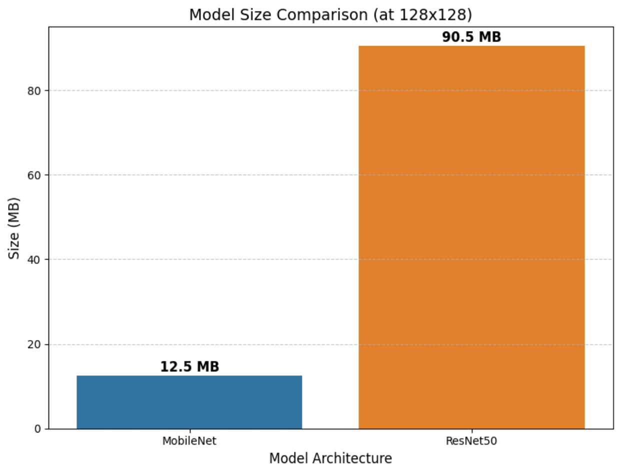


Fig. 4. Efficiency comparison at 128x128. MobileNet (12.5 MB) is significantly more storage-efficient than ResNet50 (90.5 MB), making it the only viable option for mobile deployment.

F. Training Stability

Analyzing the loss metrics across resolutions reveals the stability of the architectures.

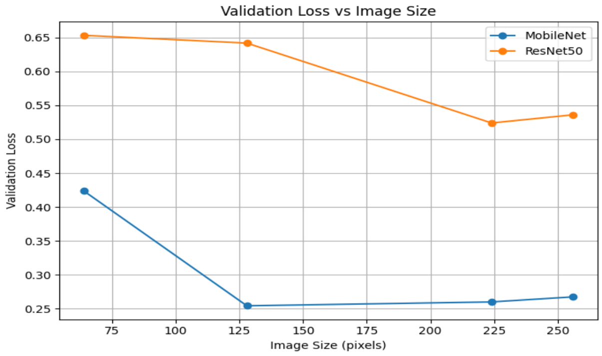


Fig. 5. Validation Loss Analysis. MobileNet (Blue) maintains a consistently low loss (0.25) across all resolutions. In contrast, ResNet50 (Orange) exhibits high and fluctuating loss (>0.50), confirming that the heavier model struggles to converge effectively on this dataset.

G. Graphical Analysis

To visually validate the quantitative findings, we analyzed the performance trends across resolutions.

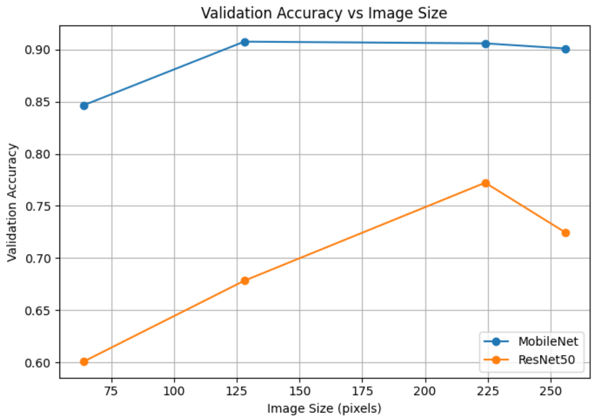


Figure 6: Validation Accuracy vs. Image Size. This plot illustrates the "sweet spot" for MobileNet at 128px. While ResNet50 (Orange line) continues to improve linearly as resolution increases, it remains below MobileNet (Blue line) which plateaus early, indicating high efficiency at low resolutions.

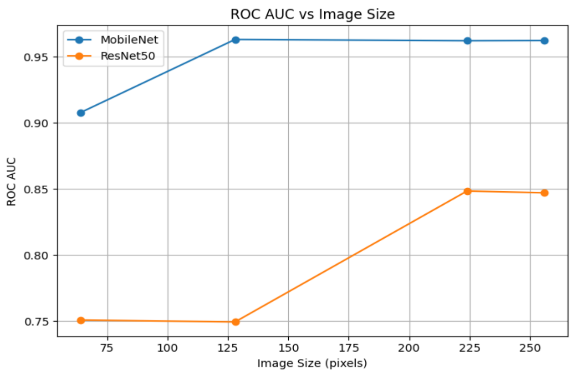


Figure 7: ROC AUC Comparison. MobileNet consistently achieves near-perfect AUC (> 0.96) across resolutions >= 128, demonstrating robust classification capability compared to ResNet50's fluctuating performance.

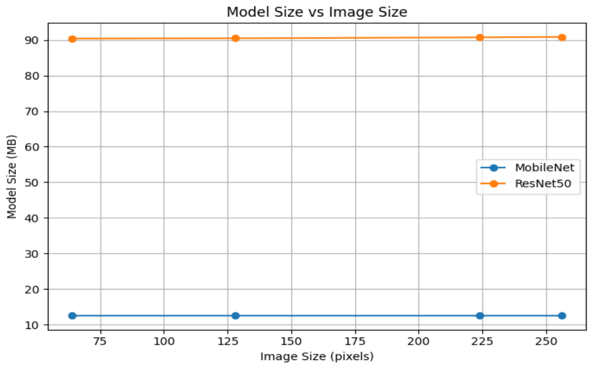


Figure 8: Model Size Efficiency. A comparison of memory footprint shows MobileNet is approximately 7 times smaller than ResNet50, a critical factor for mobile app integration.

Discussion

1. The "Sweet Spot" Resolution

Our central hypothesis was that increasing image resolution would linearly correlate with accuracy. However, the experimental data reveals a non-linear relationship, particularly for the MobileNet architecture. The performance plateau observed between 128x128 (90.76%) and 224x224 (90.59%) indicates a point of diminishing returns[12].

We posit that for macro-dermatological features like acne—which present as distinct, high-contrast red lesions—a resolution of 128x128 preserves sufficient spatial frequency information for feature extraction. The slight degradation at 224x224 may be attributed to the "curse of dimensionality," where the increased input size introduces high-frequency noise (e.g., lighting variations, skin texture artifacts) that the lightweight MobileNet architecture struggles to filter out effectively. This finding challenges the industry standard of defaulting to 224x224 for all transfer learning tasks, suggesting that downsampling is not merely a compression technique but potentially a regularization strategy.

2. Deployment Scenario: The “Edge-First” Workflow

The superior efficiency of MobileNet allows for a paradigm shift in how dermatological AI is deployed. We propose an "Edge-First" workflow for mobile health applications:

a) **Capture:** User captures a high-resolution image (>4K) using a smartphone camera.

b) **Preprocessing:** The app locally downsamples the image to 128x128 pixels.

c) **Inference:** The MobileNet model (12.5 MB) processes the image locally on the device's NPU/CPU.

d) **Result:** Diagnostic feedback is provided in <50 ms without requiring internet connectivity.

This architecture eliminates the privacy risks associated with uploading sensitive medical images to the cloud and ensures functionality in remote areas with poor connectivity—a critical requirement for telemedicine in developing nations.

3. Limitations and Future Work

While this study provides compelling evidence for efficiency, several limitations must be acknowledged to contextualize the findings.

a) **Binary vs. Multi-class:** Our dataset was limited to a binary classification (Acne vs. Normal). More complex dermatological conditions, such as distinguishing Melanoma from Seborrheic Keratosis, rely on subtle textural cues that might require the higher capacity of ResNet50 or resolutions exceeding 256x256.

b) **Quantization Effects:** We evaluated model size based on 32-bit floating-point weights. Future work will explore Post-Training Quantization (PTQ) to convert weights to 8-bit integers (INT8), potentially reducing the MobileNet size from 12.5 MB to <4 MB, further optimizing it for IoT devices.

c) **Demographic Bias:** The dataset's diversity regarding skin tones (Fitzpatrick scales) was not explicitly controlled. Ensuring that the resolution downsampling does not disproportionately affect performance on darker skin tones is a necessary ethical prerequisite for clinical deployment.

Conclusion

This study systematically evaluated the trade-off between input image resolution and diagnostic performance for dermatological AI. Our experimental results challenge the prevailing assumption that higher resolutions are intrinsically superior for medical image analysis. We demonstrated that for macro-pathologies like acne, a MobileNet architecture operating at 128x128 pixels achieves a diagnostic "sweet spot," delivering 90.76% accuracy and 94% clinical sensitivity.

Crucially, this configuration reduces the model footprint to just 12.5 MB—approximately 7x smaller than the industry-standard ResNet50—without statistically significant loss in diagnostic capability. In contrast, the heavier ResNet50 architecture exhibited severe overfitting and instability at lower resolutions, rendering it unsuitable for resource-constrained environments.

These findings establish a concrete design pathway for Edge-AI dermatology. By adopting the proposed lightweight, low-resolution workflow, developers can deploy privacy-preserving, offline-capable diagnostic tools that function effectively even on low-end hardware, democratizing access to dermatological care in underserved regions.

References

[1] Bellman R. An Introduction to Artificial Intelligence: Can Computers Think? 1st ed. Boyd & Fraser Publishing Company; San Francisco, CA, USA: 1978

[2] Litjens G., Kooi T., Bejnordi B.E., Setio A.A.A., Ciompi F., Ghafoorian M., van der Laak J.A.W.M., van Ginneken B., Sánchez C.I. A Survey on Deep Learning in Medical Image Analysis. Med. Image Anal. 2017;42:60–88.

[3] Laino M.E., Cancian P., Politi L.S., Della Porta M.G., Saba L., Savevski V. Generative Adversarial Networks in Brain Imaging: A Narrative Review. J. Imaging. 2022;8:83

[4] Lassau N., Bousaid I., Chouzenoux E., Lamarque J.P., Charmettant B., Azoulay M., Cotton F., Khalil A., Lucidarme O., Pigneur F., et al. Three Artificial Intelligence Data Challenges Based on CT and MRI. Diagn. Interv. Imaging. 2020;101:783– 788.

[5] Alzubaidi L., Zhang J., Humaidi A.J., Al-Dujaili A., Duan Y., Al- Shamma O., Santamaría J., Fadhel M.A., Al-Amidie M., Farhan L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. Volume 8. Springer International Publishing; Cham, Switzerland: 2021.

[6] C. Chakraborty, U. Achar, S. Nayek, A. Achar, and R. Mukherjee."CAD-PsorNet: deep transfer learning for computer-assisted diagnosis of skin psoriasis," Scientific Reports, vol. 14, no. 1, p. 26557, Nov. 2024. doi: 10.1038/s41598-024-76852.

[7] Y. Shi, K. Yang, T. Jiang, J. Zhang, K.B. Letaief Communication-efficient edge ai: algorithms and systems IEEE Commun. Surv. Tutor., 22 (4) (2020), pp. 2167-2191

[8] Yang, T. Baker, et al.A federated learning attack method based on edge collaboration via cloud Software Pract. Ex. (2022), pp. 1-18,

[9] Rausch T, Hummer W, Muthusamy V, Rashed A, Dustdar S. Towards a serverless platform for edge {AI}. In2nd USENIX Workshop on Hot Topics in Edge Computing (HotEdge 19) 2019.

[10] Sabanic P, Misono M, Bodea T, Pritzi J, Hackl M, Stavrakakis D, Bhatotia P. Confidential Serverless Computing. arXiv preprint arXiv:2504.21518.

[11] Rashid, S., Bernhard, L., Stabenow, S. et al. Bridging the gap between models and reality: development of a research environment for an object-oriented hospital information system to integrate artificial intelligence and robotics into clinical practice. Int J CARS 20, 1771–1783 (2025).

[12] Kulathunga, N., Ranasinghe, N. R., Vrinceanu, D., Kinsman, Z., Huang, L., & Wang, Y. (2021). Effects of Nonlinearity and Network Architecture on the Performance of Supervised Neural Networks. Algorithms, 14(2), 51.