iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

EARLY DETECTION OF DIABETES MELLITUS VIA FOOT THERMAL IMAGING USING CONVNEXT AND EFFICIENTNETB3 ARCHITECTURES: A COMPARATIVE STUDY

P. Vishnuvardhan Reddy, P. Uday Kiran,
N. Hemanth Padma Thribhuvan, M. Manikanta
Department of Electronics and Communication Engineering
R.V.R. & J.C. College of Engineering, Guntur, Andhra Pradesh, India

Dr. N. Sudheer Kumar, M.Tech., Ph.D.
Department of Electronics and Communication Engineering
R.V.R. & J.C. College of Engineering, Guntur, Andhra Pradesh, India

ABSTRACT:
Diabetes Mellitus (DM) is a chronic metabolic disorder that can lead to severe complications such as diabetic foot syndrome if not diagnosed early. Traditional diagnostic methods, though reliable, are often invasive and impractical for large-scale screening or continuous monitoring. This study proposes a non-invasive framework for early DM detection using thermal foot imaging combined with deep learning models. Specifically, it compares the performance of two advanced Convolutional Neural Network (CNN) architectures—ConvNeXt and EfficientNetB3—both optimized using data augmentation techniques.

A variety of augmentation strategies, including geometric and intensity-based transformations, were ap- plied to improve the model's generalization. ConvNeXt, with its transformer-inspired enhancements, demonstrated superior accuracy, while EfficientNetB3 offered balanced performance with fewer compu- tational resources. Evaluation metrics such as accuracy, sensitivity, precision, and F1-score were used to assess model effectiveness. Additionally, Grad-CAM heatmaps were employed to improve interpretabil- ity. A web-based interface was also developed using Flask to support real-time image classification. The results confirm the feasibility of AI-powered thermal imaging for early and efficient diabetes screening.

Keywords
Diabetes Mellitus, Foot Thermal Imaging, ConvNeXt, EfficientNetB3, Deep Learning, Data Augmenta- tion, Grad-CAM, Web Deployment.

INTRODUCTION
Diabetes Mellitus (DM) is a major global health issue, often leading to severe complications such as diabetic foot syndrome, which may result in ulcers, infections, and amputations. Early detection of these complications is crucial for preventing long-term damage.

Traditional diagnostic methods are invasive and not suited for frequent screening. Thermal imaging offers a non-invasive, cost-effective alternative by capturing temperature patterns that may indicate early diabetic abnormalities. However, manual interpretation of these images is inconsistent and inefficient.

Deep learning, particularly Convolutional Neural Networks (CNNs), enables automated feature extrac- tion from thermal images, making them ideal for medical classification tasks. Yet, limited medical datasets often hinder performance. To address this, various data augmentation techniques are employed to increase dataset diversity and robustness.

This study compares two advanced CNN models—ConvNeXt and EfficientNetB3—on diabetic foot ther- mal images. It applies multiple augmentation strategies, evaluates performance using classification met- rics, and visualizes model focus using Grad-CAM. A Flask-based web interface is also developed to support real-time deployment, improving accessibility for clinical use.

OBJECTIVES OF THE STUDY
The primary objectives of this study are as follows:

To develop an automated deep learning framework for early detection of diabetic complications using foot thermal images.

To compare the performance of two state-of-the-art CNN models—ConvNeXt-Tiny and Efficient- NetB3—under various data augmentation techniques.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

To enhance model interpretability using Grad-CAM heatmaps to identify the key regions influencing classification.
To evaluate model performance using metrics such as Accuracy, Precision, Recall (Sensitivity), and F1 Score.
To implement a Flask-based web application for real-time thermal image classification and visual- ization for clinical usage.

MATERIALS AND METHODS

Dataset Description

This study uses the ThermoDataBase, a publicly available dataset for diabetic foot thermal image clas- sification. The dataset consists of infrared foot thermal images categorized into two classes: diabetic (DM Group) and non-diabetic (Control Group). It includes:
Training Set: 1,444 images (722 DM, 722 Control)
Validation Set: 442 images (220 DM, 220 Control)
The dataset is balanced to avoid classification bias and to ensure robust performance evaluation across both classes.

Data Preprocessing

Before training the CNN models, thermal images are preprocessed as follows:
Rescaling: Pixel values normalized to $[0, 1]$ using rescale = 1./255 to improve model conver- gence.
Resizing: Images resized to architecture-specific dimensions:
ConvNeXt: 224×224 pixels
EfficientNetB3: 300×300 pixels
Tensor Conversion: Converted to float32 tensors compatible with Keras models.

Data Augmentation

To address the limitations of small datasets in medical imaging, we applied data augmentation techniques to artificially increase data variability and improve generalization. Augmentations were performed using the ImageDataGenerator module in Keras. Techniques included:
Geometric Transformations: Horizontal and vertical flip, random rotation, width and height shifts, and shear transformations.
Zoom and Brightness Adjustments: Simulate different capture conditions and thermal inten- sity variations.
Advanced Augmentations: Techniques like Principal Component Analysis (PCA), Kernel PCA, Non-negative Matrix Factorization (NMF), Independent Component Analysis (ICA), and Factor Analysis were simulated through parameter-tuned transformations.
These augmentations were applied individually to evaluate their individual impact on model performance. A comparative analysis of metrics across different augmentations is discussed in the results section.

METHODOLOGY

ConvNeXt Architecture Overview

ConvNeXt is a convolutional neural network (CNN) designed to match the performance of transformer- based architectures while preserving the efficiency of convolution operations. It builds upon the ResNet foundation by incorporating modern design principles such as large kernel sizes, Layer Normalization, and GELU activation. It maintains a stage-wise structure and processes images hierarchically.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

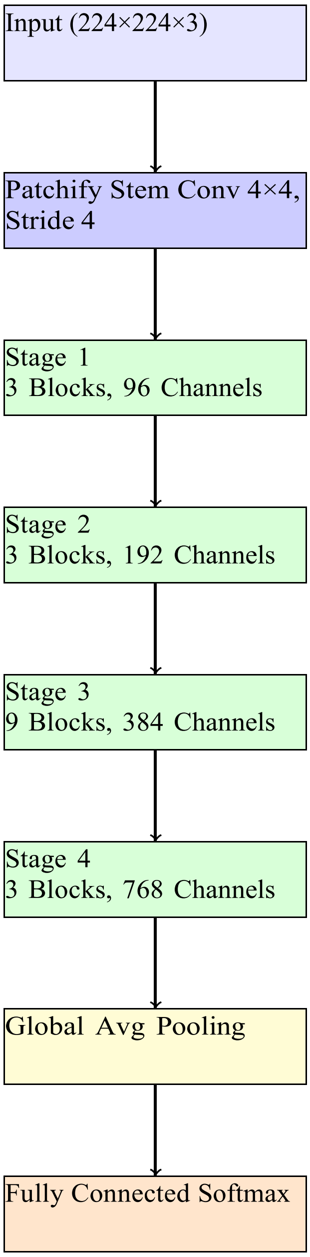


Figure 1: ConvNeXt-Tiny Block Diagram Showing Stage-Wise Architecture

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

EfficientNetB3 Architecture Overview

EfficientNetB3 is part of the EfficientNet family, which employs compound scaling to uniformly scale the network's depth, width, and resolution. It is based on the Mobile Inverted Bottleneck Convolution (MBConv) with Squeeze-and-Excitation (SE) modules to improve channel interdependencies.

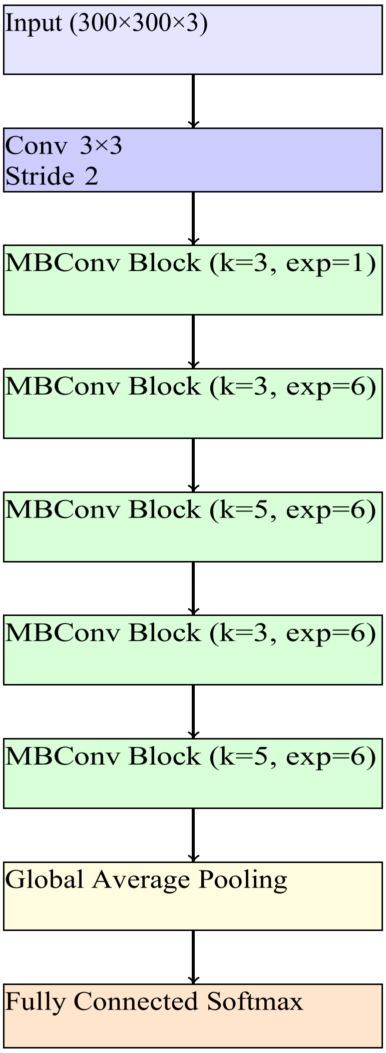


Figure 2: Block Diagram of EfficientNetB3 Architecture

Design Principles of ConvNeXt

Key elements of ConvNeXt's design include:
Patchify Stem: Replaces patch embedding with 4 × 4 convolution.
Depthwise Convolutions: Reduce computation while preserving receptive field.
Layer Normalization: Enhances stability.
GELU Activation: Smooth and effective non-linearity.
Residual Blocks: Allow deeper networks to train effectively.

Design Principles of EfficientNetB3

EfficientNetB3 integrates:
Compound Scaling: Balances width, depth, and resolution.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

MBConv Blocks: Efficient feature extraction with expansion layers.
SE Attention: Re-calibrates feature maps.
Swish Activation: Improves gradient-based optimization.
Transfer Learning and Implementation
Both models were initialized with ImageNet pretrained weights. Custom heads were added:
Global Average Pooling
Dense Layer (256 units, ReLU)
Dropout (0.4)
Sigmoid Output
Initially, base layers were frozen for transfer learning, and gradually unfrozen for fine-tuning.

Training Configuration and Setup
Training was done on Google Colab using GPU. Shared configuration:
Loss Function: Binary Crossentropy
Optimizer: AdamW
Batch Size: 32
Epochs: 25
Callbacks: EarlyStopping, ModelCheckpoint

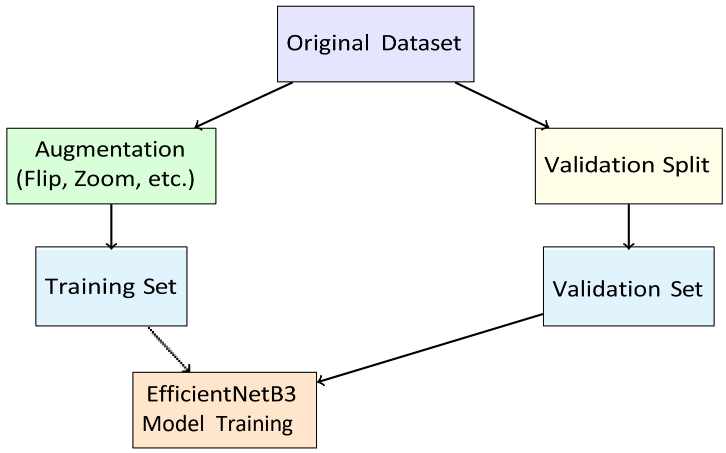


Figure 3: Training Pipeline with Augmentation for EfficientNetB3

IJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

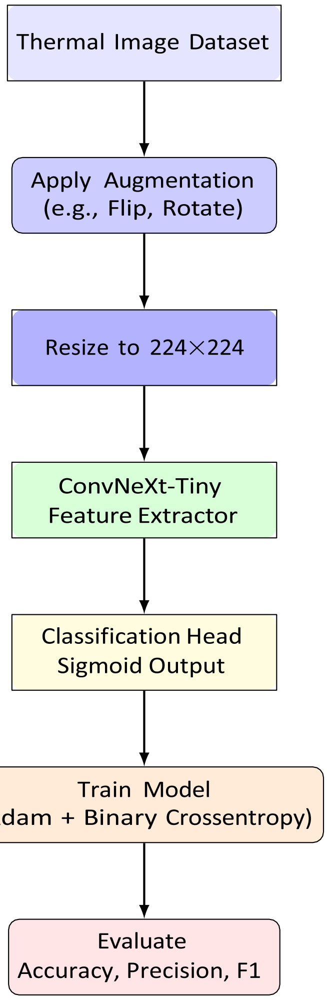


Figure 4: ConvNeXt Model Training Pipeline with Augmentation Techniques

RESULTS AND EVALUATION

ConvNeXt-Tiny Evaluation Metrics

Table **??** shows the classification performance of the ConvNeXt-Tiny model across various augmentation techniques. The model achieved the highest accuracy of 97.89% using Principal Component Analysis (PCA), followed by Dictionary Learning and Incremental PCA. These methods enhanced generalization by introducing meaningful variance to the training set. On the other hand, augmentations like horizontal and vertical flips reduced accuracy—likely due to distortion in anatomical alignment between the left and right foot, which is critical in diabetic diagnosis.

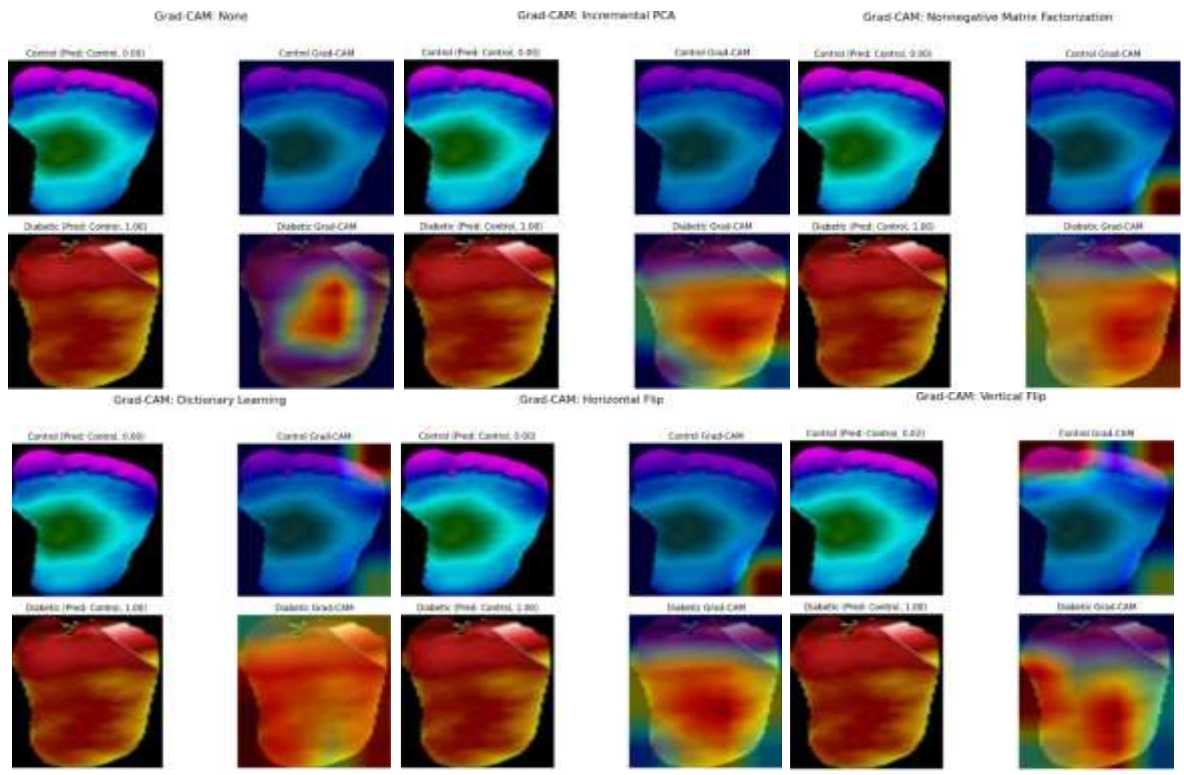
iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Table 1: ConvNeXt-Tiny Metrics Across Augmentation Techniques

Augmentation	Accuracy (%)	Sensitivity (%)	Precision (%)	F1 Score (%)
None	89.57	88.10	89.96	89.64
Incremental PCA	97.18	98.25	97.19	97.18
NMF	92.20	96.27	92.17	92.13
Dictionary Learning	97.42	95.87	97.71	97.46
PCA	**97.89**	96.27	**98.18**	**97.93**
Horizontal Flip	86.75	96.43	88.68	85.79
Vertical Flip	87.93	92.30	87.86	87.87
Gaussian Filter	92.68	92.70	92.88	92.73
Factor Analysis	94.57	96.27	94.57	94.57
LDA	93.86	92.70	94.22	93.93
Kernel PCA	93.63	90.32	94.49	93.72
ICA	96.47	97.86	96.47	96.47
Change of Color Space	96.47	93.89	97.00	96.53

Visual Attention via Grad-CAM (ConvNeXt)

Figure 5 illustrates Grad-CAM heatmaps from the ConvNeXt-Tiny model across all applied augmentations. The model consistently focused on localized thermal anomalies—especially asymmetric hotspots—corresponding to diabetic abnormalities. These observations confirm that ConvNeXt makes decisions based on clinically relevant cues.



iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

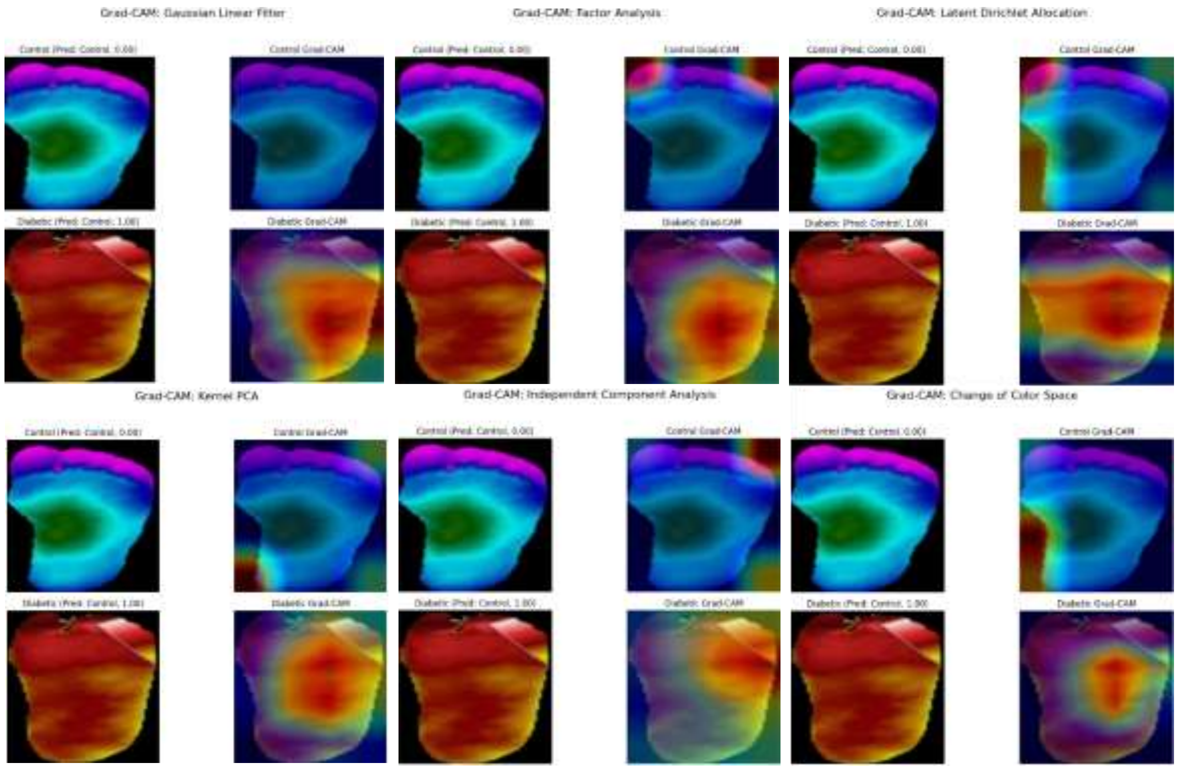


Figure 5: Grad-CAM Visualizations for ConvNeXt-Tiny Across Augmentations

EfficientNetB3 Evaluation Metrics

EfficientNetB3, while slightly behind ConvNeXt in overall accuracy, exhibited stronger sensitivity under some augmentations. Table **??** shows that the model achieved top sensitivity scores using Factor Analysis and LDA augmentations, making it highly effective in detecting true diabetic cases.

Table 2: EfficientNetB3 Metrics Across Augmentation Techniques

Augmentation	**Accuracy (%)**	**Sensitivity (%)**	**Precision (%)**	**F1 Score (%)**
None	89.57	88.10	89.96	89.64
Incremental PCA	92.18	93.25	92.19	92.18
NMF	92.20	91.27	92.17	92.13
Dictionary Learning	92.42	90.87	92.71	92.46
PCA	92.89	91.27	93.18	92.93
Horizontal Flip	93.39	96.83	94.06	93.12
Vertical Flip	93.86	94.68	93.94	93.89
Gaussian Filter	93.15	91.90	93.55	93.22
Factor Analysis	88.65	**97.62**	90.69	87.83
LDA	**94.10**	95.24	94.32	93.95
Kernel PCA	90.05	89.52	91.54	90.14
ICA	94.10	95.24	94.32	93.95
Change of Color Space	91.47	92.46	91.73	91.45

Visual Attention via Grad-CAM (EfficientNetB3)

Figure 6 shows Grad-CAM heatmaps from EfficientNetB3 across all applied augmentations. Unlike ConvNeXt, the attention regions are broader and sometimes include peripheral areas. This suggests that EfficientNetB3 captures a wider region of interest but may occasionally highlight irrelevant zones.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

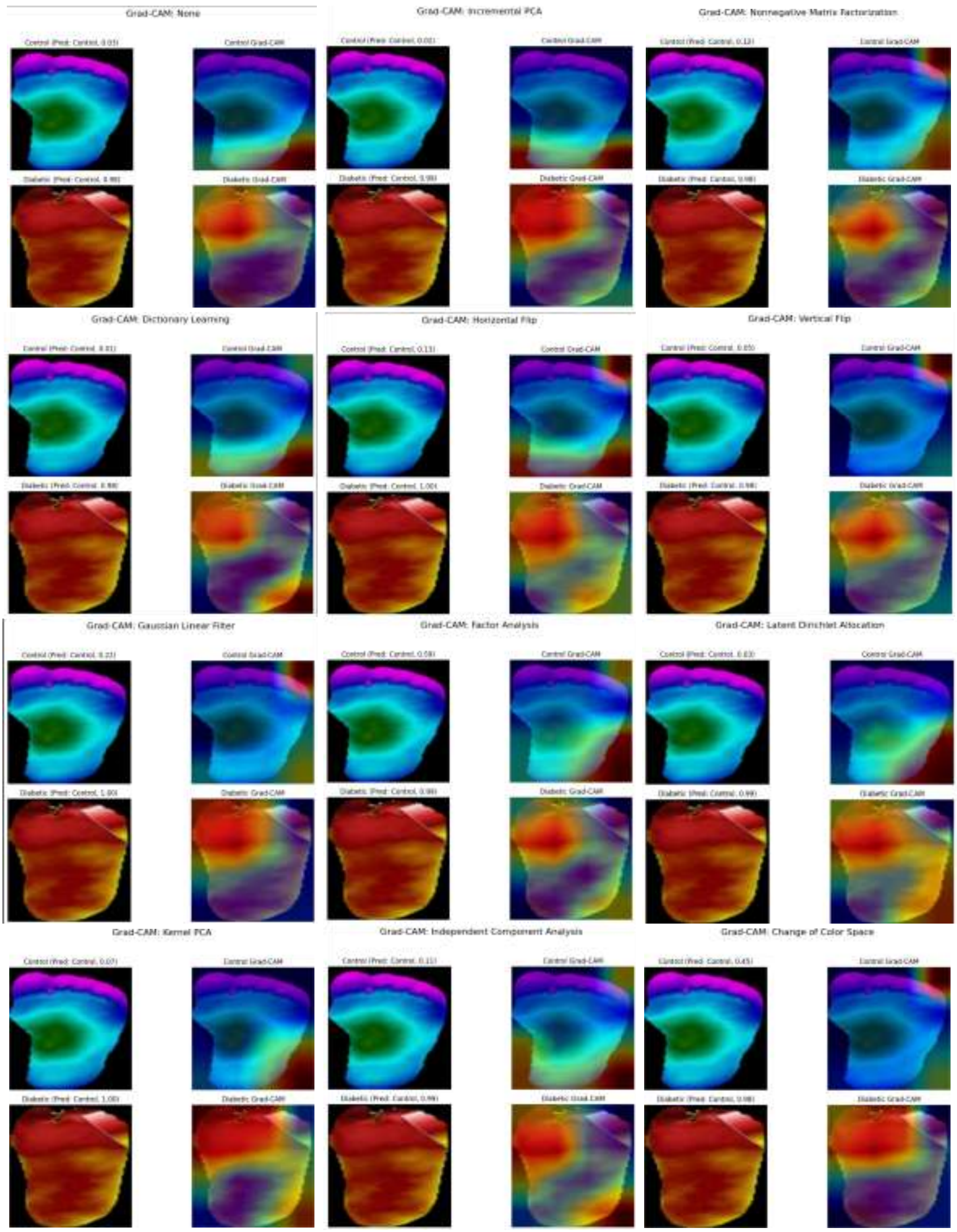
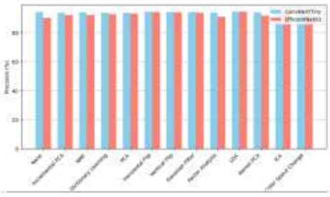
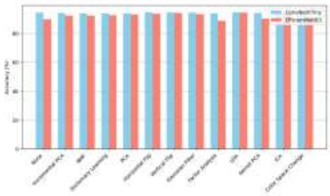


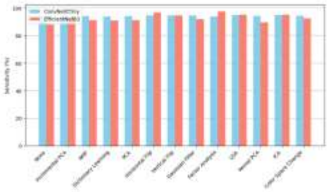
Figure 6: Grad-CAM Visualizations for EfficientNetB3 Across Augmentations

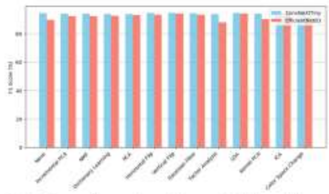
iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

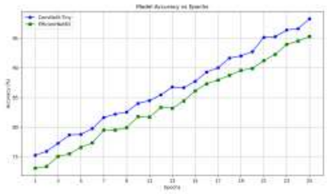
Comparison of Metrics & Web Interface


(a) Precision Comparison Across Models and Augmentations


(b) Accuracy Comparison Across Models and Augmentations


(c) Sensitivity Comparison Across Models and Augmentations


(d) F1 Score Comparison Across Models and Augmentations


(e) Accuracy Progression Over Epochs (1)

(f) Accuracy Progression Over Epochs (2)

(g) Accuracy Progression Over Epochs (3)

(h) Accuracy Progression Over Epochs (4)

Comparative Graphical Analysis

The bar plots in Figures 5 through 8 compare both models across key performance metrics—Accuracy, Sensitivity, Precision, and F1 Score—across all augmentation types.
- Figure 6 (Accuracy): ConvNeXt consistently outperformed EfficientNetB3, especially under PCA and Dictionary Learning augmentations. - Figure 7 (Sensitivity): EfficientNetB3 had better sensitivity in several cases, making it reliable in identifying actual diabetic cases. - Figure 5 (Precision): ConvNeXt showed higher precision across most augmentations, indicating fewer false positives. - Figure 8 (F1 Score): ConvNeXt maintained a strong balance between sensitivity and precision, which is reflected in higher F1 Scores.
These comparisons reinforce that ConvNeXt is more precise and stable, while EfficientNetB3 leans toward being more sensitive. In a clinical context, ConvNeXt would be preferred for minimizing misclassifica- tions, while EfficientNetB3 might serve well in screening settings where missing a true case would have more serious consequences.

Accuracy Over Epochs

Figure 9 illustrates how the models' accuracy improved over the course of 25 training epochs. ConvNeXt showed a sharper learning curve, reaching a peak accuracy of 98.3%, whereas EfficientNetB3 had a slower but steady progression, peaking at 95.3%. This shows that ConvNeXt not only learns faster but also generalizes better under the same training conditions, making it more suitable for deployment when rapid model convergence is essential.
The smooth convergence of both models indicates that the training was well-configured, with no over- fitting or erratic spikes in performance. However, ConvNeXt's architectural advantages allowed it to extract more discriminative features early on, leading to more rapid performance gains.

Web Interface for Model Deployment

A web-based interface was built using Flask to facilitate the practical deployment of the trained models. As shown in Figures 10 to 12, users can upload thermal foot images, apply various augmentations such as PCA, ICA, and LDA, and

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

view real-time visual outputs. This interactive platform makes the system accessible to non-technical users, such as clinicians, allowing them to explore model predictions and augmentation effects without requiring any coding expertise.
To further interpret model decisions, Grad-CAM visualizations were employed to compare the attention mechanisms of ConvNeXt-Tiny and EfficientNetB3. These visual heatmaps revealed that ConvNeXt- Tiny consistently focused on key asymmetric thermal regions, whereas EfficientNetB3 displayed broader, less specific activations. This comparative analysis reinforces ConvNeXt's stronger precision and clinical interpretability, while also highlighting EfficientNetB3's strength in capturing broader patterns for higher sensitivity.

CONCLUSION

This study successfully established a deep learning-based framework for the early detection of Diabetes Mellitus using foot thermal imaging. By leveraging thermal patterns indicative of diabetic complications, the project compared two advanced convolutional neural network architectures—ConvNeXt-Tiny and EfficientNetB3—each evaluated under multiple data augmentation techniques to assess generalization and classification performance.
Experimental results demonstrated that the ConvNeXt-Tiny model consistently outperformed Efficient- NetB3 across most evaluation metrics. Notably, with Principal Component Analysis (PCA) augmenta- tion, ConvNeXt-Tiny achieved:
Accuracy: 97.89%
Precision: 98.18%
Sensitivity (Recall): 96.27%
F1 Score: 97.93%
Conversely, EfficientNetB3, while slightly behind in overall accuracy and precision, demonstrated stronger sensitivity in some settings. Its highest performance under LDA and ICA augmentations was:
Accuracy: 94.10%
Precision: 94.32%
Sensitivity (Recall): 95.24%
F1 Score: 93.95%
These findings underscore a trade-off: EfficientNetB3's higher sensitivity makes it suitable for screening scenarios where minimizing false negatives is critical, whereas ConvNeXt-Tiny's balanced performance makes it preferable for precision-focused diagnostic tasks.
Grad-CAM visualizations further supported these conclusions. ConvNeXt-Tiny consistently localized rel- evant asymmetric temperature anomalies with higher precision, while EfficientNetB3 exhibited broader, occasionally less specific attention maps.
To ensure clinical accessibility, a Flask-based web interface was developed, enabling real-time image classification and augmentation visualization without requiring programming knowledge. This practical tool bridges research and application, facilitating potential use in clinical or remote healthcare settings.
In summary, ConvNeXt-Tiny with PCA augmentation offers the most robust solution for diabetic foot thermal image classification in this study. Its superior accuracy, interpretability, and adaptability mark it as a promising candidate for early screening tools. Future work will focus on extending this framework to mobile platforms, incorporating ensemble models, and validating with larger and more diverse datasets.

REFERENCES

A. Anaya-Isaza and M. Zequera-Diaz, Detection of Diabetes Mellitus With Deep Learning and Data Augmentation Techniques on Foot Thermography, IEEE Access, vol. 10, pp. 59564–59584, 2022.
Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, A ConvNet for the 2020s, in Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
M. Tan and Q. V. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, in Proceedings of International Conference on Machine Learning (ICML), 2019.
J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database, in Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2009.
S. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, Grad-CAM: Visual Expla- nations from Deep Networks via Gradient-based Localization, in Proceedings of IEEE International Conference on Computer Vision (ICCV), 2017.
A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in Advances in Neural Information Processing Systems (NeurIPS), 2012.
I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition, in Proceedings of IEEE

International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
D. P. Kingma and J. Ba, Adam: A Method for Stochastic Optimization, in Proceedings of Interna- tional Conference on Learning Representations (ICLR), 2015.
H. Huang, J. Liu, B. Lv, and L. Zhang, Thermal Imaging for Early Screening of Diabetes Foot Ulcers: A Review, Journal of Healthcare Engineering, 2021.
M. Abadi et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, arXiv preprint arXiv:1603.04467, 2016.
O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional Networks for Biomedical Image Seg- mentation, in Proceedings of International Conference on Medical Image Computing and Computer- Assisted Intervention (MICCAI), 2015.
R. Acharya, S. Mookiah, N. Kumaravel, S. Sree, and L. Martis, Diagnosis of Diabetes Mellitus Based on Thermal Images Using Texture Features and Artificial Neural Networks, Journal of Medical Systems, vol. 36, pp. 2553–2561, 2012