XAI IN DERMATOLOGY:SKIN LESION CLASSIFICATION USING DL MODELS AND XAI TECHNIQUES

Rakesh Krishna R B,Vellore Institute of Technology,Chennai
Sruthi Y,Vellore Institute of Technology,Chennai
Mohammed Farman S,Vellore Institute of Technology,Chennai

Abstract

Skin cancer is one of the most common types of cancer, and its early detection can be life-saving. In this project, we built a system that uses deep learning to classify skin lesions as either benign or malignant. We applied three well-known models - InceptionV3, Xception, and EfficientNet to achieve this. While these models are highly accurate, they often operate as "black boxes," making it hard for doctors to understand how they reach their decisions. To solve this, we integrated explainability techniques like Grad-CAM and SmoothGrad, which visually show which areas of the image influenced the model's decision. This not only improves trust in the system but also makes it easier for healthcare professionals to verify its results. Our experiments showed that InceptionV3 performed best in terms of both accuracy and explanation quality. This work brings us closer to making AI-driven tools more trustworthy and transparent in real-world medical applications.

Keywords:Skin Lesion Classification, Deep Learning, InceptionV3, Xception, EfficientNet, Explainable AI, Grad-CAM, SmoothGrad, Medical Image Analysis, Skin Cancer Detection.

Introduction

Skin cancer is a growing health concern, with millions of cases diagnosed every year. Early detection is key, but identifying skin cancer from images can be challenging, even for experienced dermatologists. Mistakes in diagnosis can be costly, so having reliable tools that assist doctors is crucial. This is where artificial intelligence (AI), particularly deep learning, comes in. By training models on large datasets of skin images, AI systems can help detect skin cancer with impressive accuracy.

However, despite their potential, AI models often operate in a "black box" manner, meaning they make decisions without offering any insight into how those decisions were made. This lack of transparency can be a barrier to their adoption in healthcare, where trust and understanding are essential. Doctors need to know not just the outcome, but why a certain result was reached.

To address this, we used explainable AI (XAI) techniques in our project. Specifically, we employed Grad-CAM and SmoothGrad, which provide visual explanations by highlighting the parts of the image that influenced the AI's decision. These techniques help make the model's thought process more transparent, so that doctors can see and understand the rationale behind the predictions.

In this project, we focused on three popular deep learning models: InceptionV3, Xception, and EfficientNet. We evaluated their performance in classifying skin lesions and analysed how well Grad-CAM and SmoothGrad helped explain their predictions. Our goal is to create

a system that is not only accurate but also understandable, making it more useful for real-world medical applications.

Literature Review

1.Dagnaw, G. H. et al. (2024) proposed a multimodal approach using Vision Transformers (ViTs), CNNs, and XAI for skin cancer classification. Their work demonstrated how integrating these models improved diagnostic accuracy and transparency. They emphasized the role of XAI in building trust for AI-based systems in clinical settings.

2.Gamage, L. et al. (2024) developed a novel Saliency Mask-Guided Vision Transformer (SM-ViT) for early melanoma detection. They compared various deep learning models and highlighted SM-ViT's superior performance. The study also introduced a web application for real-time melanoma detection, showing the practical impact of AI.

3.Mosquera-Zamudio et al. (2022) examined deep learning's ability to analyze whole-slide images of melanocytic tumors. They found that AI models performed on par with expert pathologists in diagnosing melanomas. The study showed the promise of deep learning in dermatopathology by efficiently identifying key diagnostic features.

4.Nigar, N. et al. (2022) discussed the importance of improving diagnostic accuracy for skin cancer. They highlighted the "black-box" nature of deep learning models and how it limits clinical adoption. The study advocated for XAI as a solution to make AI models more interpretable and trusted in healthcare.

5.Attallah, O. (2024) explored the global impact of skin cancer and the critical need for early detection. The study introduced XAI, especially LIME, to make AI-driven predictions more transparent. Attallah emphasized that explainability would help clinicians trust AI systems in real-world settings.

6.Lucieri, A. et al. (2021) highlighted early melanoma detection as key to improving patient survival rates. They critiqued the lack of explanations in AI diagnosis and called for systems that combine both accuracy and transparency. Their work stressed the need for AI tools that provide clear, understandable decisions for clinical use.

7.Mridha, K. et al.(2023) underscored the necessity of early, accurate skin cancer diagnosis to reduce healthcare workloads. Their approach using CNNs and XAI like Grad-CAM achieved an 82% classification accuracy. The study also demonstrated how XAI techniques can make model decisions more interpretable for clinicians.

8.Alche, M. N. et al. (2021) worked on improving early detection of neglected tropical skin diseases (NTDs) in resource-limited settings. They used EfficientNet-B3 and achieved a 91.53% accuracy rate on their dataset. Their findings highlighted the potential of machine learning to help medical staff in low-income regions.

9.Azlaan, H., & Oluwaseyi, J. (2024) focused on early detection of neglected tropical skin diseases (NTDs) using EfficientNet-B3. They achieved a 91.53% classification accuracy, showcasing the model's effectiveness in under-resourced areas. Their research emphasized the importance of AI tools for frontline healthcare workers in developing countries.

10.Owen, J., & Olaoye, G. O. (2024) investigated how XAI techniques can improve the trustworthiness of AI models in diagnosing skin cancer and blood cell diseases. They evaluated the performance and explainability of various XAI methods on benchmark datasets.The study highlighted the need for collaboration between AI researchers and medical professionals to ensure ethical use of XAI in healthcare.

11.Gautam, Y. et al. (2024) explored the use of AI in medical image analysis for skin cancer detection. They focused on XAI techniques like Grad-CAM and SHAP to make AI model decisions more interpretable. Their study emphasized the importance of explainability for clinical workflows and real-world adoption of AI systems.

12.Mohan, J. et al. (2024) examined deep learning's role in automating skin disease classification across 31 classes. Using transformer-based architectures and transfer learning, they achieved high accuracy and improved upon existing benchmarks. Their work was supported by XAI techniques like Grad-CAM and SHAP to aid dermatologists in early detection and diagnosis.

13.Rezk, E. et al. (2023) tackled the "black-box" issue of AI in skin cancer diagnosis by integrating lesion taxonomy into the model development process. This approach improved classification accuracy for skin lesions. Their use of XAI methods helped visualize decision-making, making AI predictions more transparent for healthcare providers.

14.Pintelas, E. et al.(2021) addressed the transparency challenges of CNNs in image classification. They developed an explainable image classification framework for skin cancer and plant disease prediction. Their model combined traditional machine learning with XAI techniques to make decision-making more interpretable.

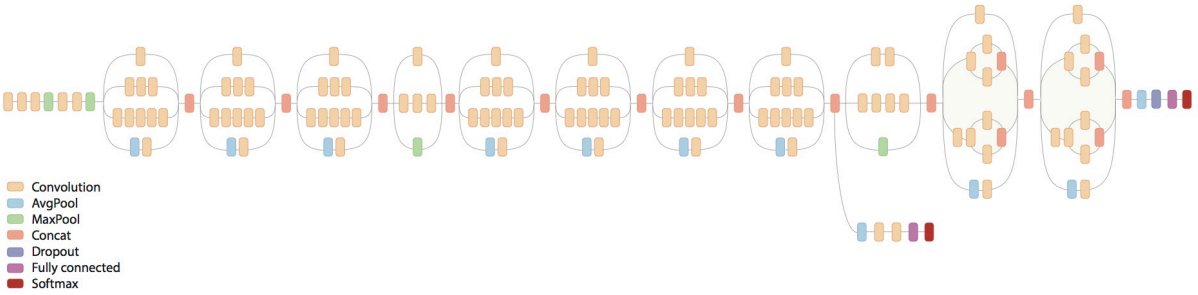
15.Fontes, M. et al. (2024) reviewed example-based XAI techniques in medical imaging, highlighting their potential in clinical practice. They examined the contributions of these methods to improving accuracy and transparency in AI-driven medical applications. Their study identified research gaps and called for more patient-centered approaches in XAI.

Summary Of Findings

Recent advancements in skin cancer classification have been driven by the integration of state-of-the-art deep learning models, such as Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs), alongside Explainable AI (XAI) techniques, to improve both diagnostic accuracy and clinical transparency. Novel architectures, including Saliency Mask-Guided ViTs and multimodal approaches that combine ViTs, Swin Transformers, and CNNs, have shown significant promise. XAI methods such as Grad-CAM, SHAP, and LIME are increasingly being employed to address the "black-box" nature of AI models, offering clinicians interpretable visual insights that enhance trust in automated systems. These advancements are not merely technical but are positioned to revolutionise dermatological care by enabling earlier detection, more accurate prognosis, and ultimately better patient outcomes. As AI continues to evolve, integrating explainability into machine learning workflows remains crucial for fostering clinical adoption and ensuring that these tools can be seamlessly integrated into real-world healthcare settings. This body of work sets the stage for further exploration of explainable and transparent AI, paving the way for a more human-centred approach to AI-driven healthcare.

Model Architectures

InceptionV3:



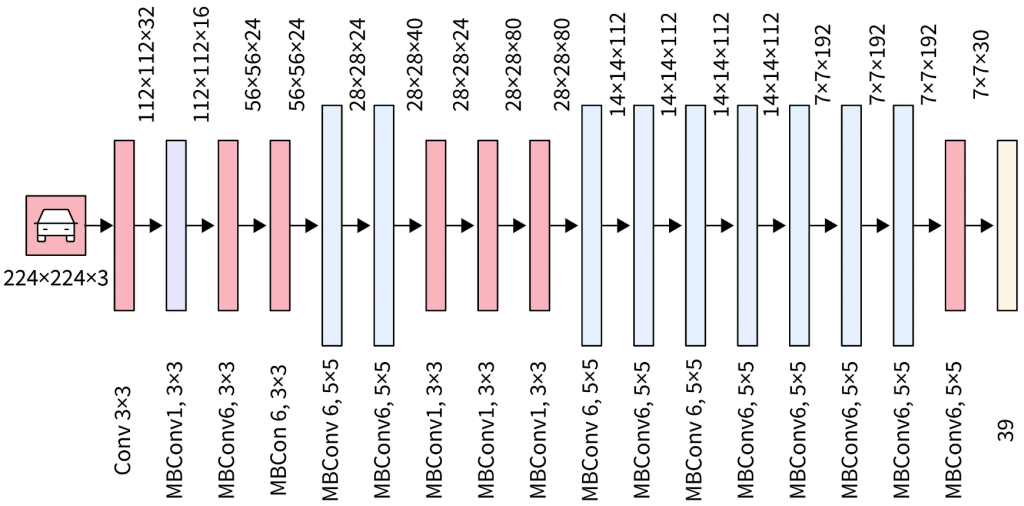
InceptionV3 is a deep convolutional neural network (CNN) that extends the original Inception model, designed to enhance computational efficiency and improve performance. Its key innovations are the use of multiple kernel sizes within the same layer (Inception modules), factorized convolutions, and efficient grid size reduction techniques.

- **Inception Modules**: Combines multiple convolution operations with different kernel sizes (1x1, 3x3, 5x5) in parallel and concatenates the results, allowing the model to capture information at multiple scales.
- **Factorized Convolutions**: Factorizes larger convolutions (e.g., a 5x5 convolution) into smaller, more efficient ones (e.g., two 3x3 convolutions) to reduce computational cost.
- **Auxiliary Classifiers**: InceptionV3 includes auxiliary classifiers to improve convergence by inserting additional loss functions in intermediate layers.
- **Key Layers**:
    - Convolutional layers with different filter sizes.
    - MaxPooling and AveragePooling for dimensionality reduction.
    - Fully connected (Dense) layers at the end.
- **Total Parameters**: ~23 million

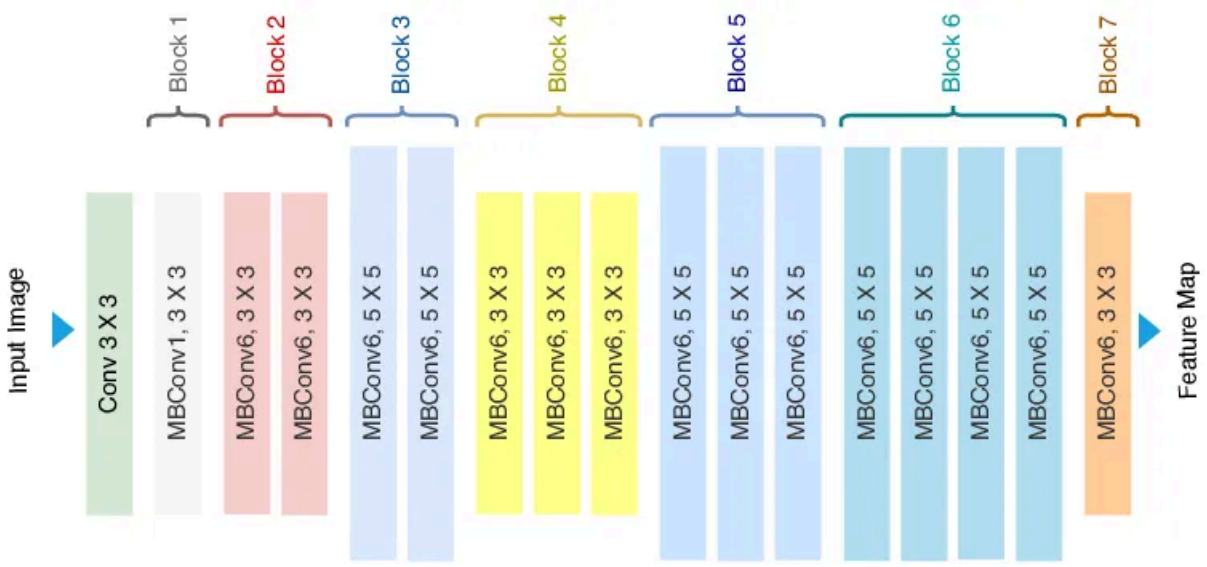
Architecture Flow:

- Input image (299x299x3)
- Stack of Inception Modules
- Global Average Pooling (reduces spatial dimensions)
- Fully Connected Layer
- Softmax for classification

EfficientnetB0:



EfficientNet:



EfficientNetB0 is part of the EfficientNet family, which uses a method called "compound scaling" to uniformly scale depth, width, and resolution of the network, resulting in a more efficient model compared to other architectures.

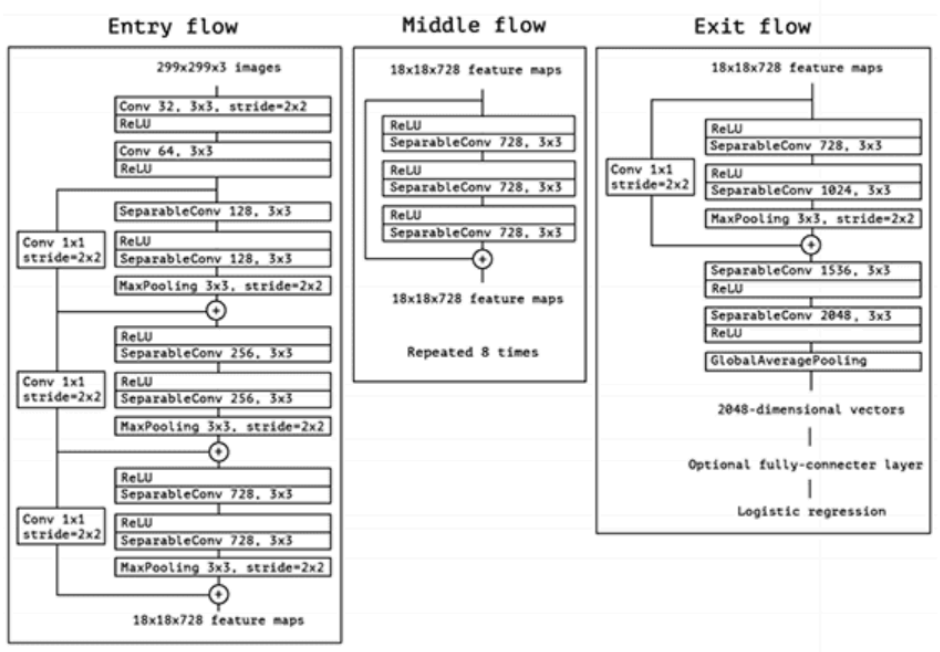
- **Compound Scaling**: EfficientNet scales three dimensions—network depth, network width (number of channels), and input resolution—uniformly with a scaling factor.

- **Mobile Inverted Bottleneck Convolution (MBConv)**: This block is derived from MobileNet and uses depthwise separable convolutions along with an inverted residual structure.
    - **Squeeze and Excitation (SE)**: SE blocks are included to recalibrate feature maps and improve model performance.
- **Key Layers**:
    - MBConv blocks for feature extraction.
    - Global Average Pooling to reduce dimensions.
    - Fully connected layers at the end for classification.
- **Total Parameters**: ~5.3 million (EfficientNetB0)

Architecture Flow:

- Input image (224x224x3 for B0) and (128x128x3 for EfficientNet)
- Multiple MBConv blocks with squeeze-and-excitation
- Global Average Pooling
- Fully Connected Layer
- Softmax for classification

Xception model:



Xception (Extreme Inception) is based on the Inception architecture but replaces the traditional Inception modules with depthwise separable convolutions. This modification makes the model more efficient and powerful for feature extraction.

- **Depthwise Separable Convolutions**: Breaks a standard convolution into two separate layers:
    - **Depthwise Convolution**: Applies a single convolution filter per input channel.

    - **Pointwise Convolution**: Follows it with a 1x1 convolution to combine the channels.
- **Residual Connections**: Like ResNet, Xception employs skip connections between layers to prevent gradient vanishing and allow deeper architectures.
- **Key Layers**:
    - Multiple depthwise separable convolution layers.
    - Residual blocks with skip connections.
    - Global Average Pooling at the end before the fully connected layer.
- **Total Parameters**: ~22.9 million

Architecture Flow:

- Input image (299x299x3)
- Entry Flow (convolutions and pooling)
- Middle Flow (multiple depthwise separable convolutions)
- Exit Flow (more convolutions and global average pooling)
- Fully Connected Layer
- Softmax for classification

Results and Discussion

Dataset

For this project, we used a subset of the **ISIC (International Skin Imaging Collaboration) Archive**, which is a large collection of skin images aimed at improving skin cancer diagnosis with AI. Specifically, we focused on classifying skin lesions into two categories: **malignant (cancerous)** and **benign (non-cancerous)**. This binary classification task helps in distinguishing potentially dangerous skin lesions from harmless ones, which is critical for early cancer detection.

The images in the dataset are high-quality, and each one has been labeled by expert dermatologists. Along with the images, there is also useful information such as the patient's age, gender, and where the lesion is located on the body. This extra information helps provide context and variety, making the dataset diverse and well-suited for training deep learning models to recognize different types of skin conditions.

By focusing on whether lesions are malignant or benign, our project aims to provide an accurate and trustworthy AI system that can assist doctors in making quick and reliable decisions about skin cancer.

Evaluation Metrics Comparison:

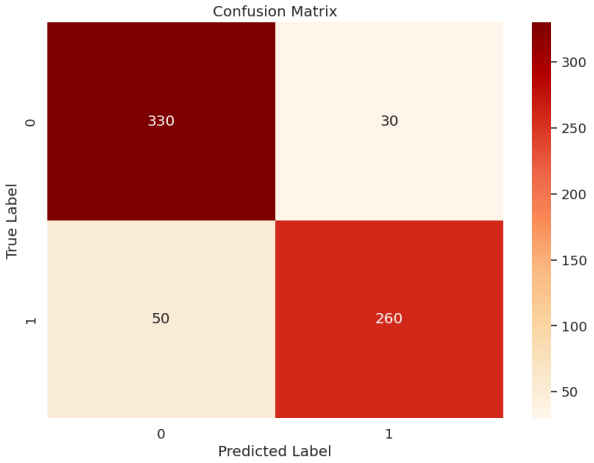
Model	Benign Precision	Benign Recall	Benign F1-Score	Malignant Precision	Malignant Recall	Malignant F1-Score	Overall Accuracy
Xception	0.87	0.92	0.89	0.9	0.84	0.87	88
EfficientNet (128x128)	0.91	0.75	0.82	0.76	0.92	0.83	83
EfficientNet B0 (224x224)	0.97	0.62	0.76	0.69	0.98	0.81	79
InceptionV3	0.86	0.79	0.82	0.77	0.85	0.81	82

Analysis and Understanding:

Xception Model:

Strengths: This model offers a well-balanced performance between benign and malignant classifications. It has high precision (0.87 for benign, 0.90 for malignant) and recall (0.92 for benign, 0.84 for malignant), resulting in a strong F1-score (0.89 and 0.87, respectively). The overall accuracy is 0.88, which is the highest among all models.

Reasoning: The Xception architecture benefits from depthwise separable convolutions, which allow it to efficiently capture complex features while keeping computational costs low. This likely contributes to its well-rounded performance in both precision and recall.

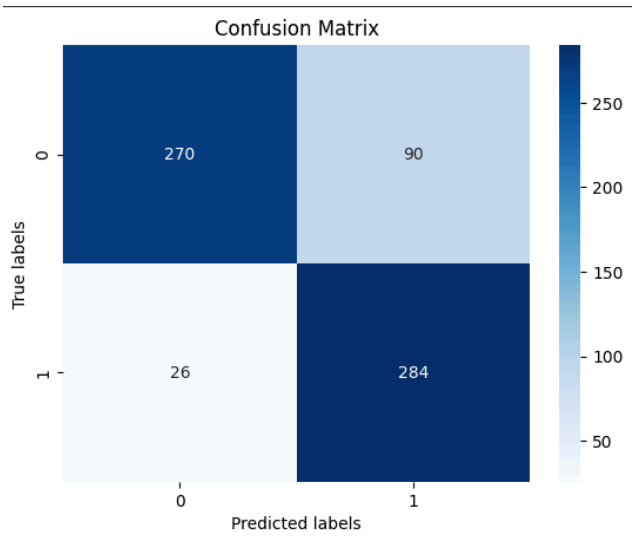


EfficientNet (128x128 Input Size):

Strengths: This model has a strong performance in benign precision (0.91), but its recall for benign (0.75) is lower compared to other models. Its malignant classification metrics (precision: 0.76, recall: 0.92) show that it is better at identifying malignant lesions.

Weaknesses: The drop in benign recall (0.75) suggests the model might be missing some benign cases, which could impact its clinical usefulness where false negatives are undesirable. However, its high malignant recall (0.92) makes it a reliable option for cancer detection.

Reasoning: EfficientNet scales both depth and width, which can result in more generalizable models. The smaller input size of 128x128 may limit its ability to capture fine details in the image, which might explain its lower benign recall.

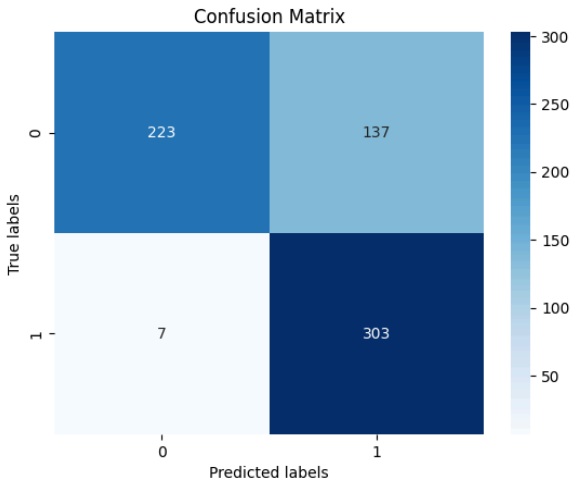


EfficientNetB0 (224x224 Input Size):

Strengths: The model excels in detecting malignant cases with an excellent recall of 0.98, ensuring that almost all malignant lesions are detected. Its precision for benign lesions is also very high (0.97), but the recall is significantly lower (0.62).

Weaknesses: The low recall for benign lesions means this model misses many benign cases, which might lead to unnecessary follow-ups or over-diagnosis.

Reasoning: The larger input size (224x224) helps capture more detailed information, which enhances the model's ability to detect malignant lesions. However, this focus on malignant cases might come at the expense of benign classification, which explains the low benign recall.

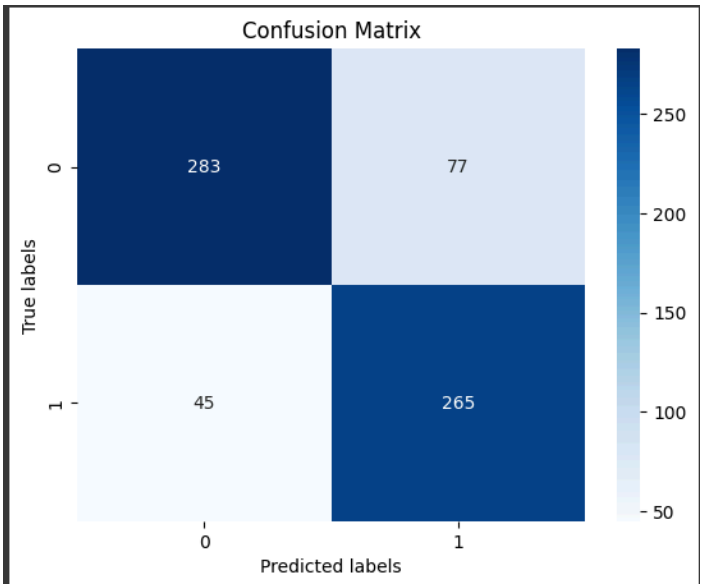


InceptionV3:

Strengths: InceptionV3 offers a balanced performance across the board, with an accuracy of 0.82. Its precision and recall for both benign (0.86 and 0.79) and malignant (0.77 and 0.85) are reasonably balanced, resulting in an overall reliable model.

Weaknesses: Although InceptionV3's performance is consistent, it doesn't excel in any one area compared to the other models.

Reasoning: InceptionV3 uses a multi-path architecture that allows it to capture different spatial resolutions simultaneously, which might be why its performance remains steady but doesn’t outperform in any specific metric.



Best Overall Model: ***The Xception model*** emerges as the most balanced, offering strong precision and recall for both benign and malignant classifications, making it ideal for clinical applications where both sensitivity and specificity are crucial.

Best for Malignant Detection: ***EfficientNetB0*** with a larger input size is excellent for malignant lesion detection, achieving a recall of 0.98. However, its lower benign recall could be a limitation.

Best for Balanced Performance: ***InceptionV3*** provides a middle ground with solid metrics across both benign and malignant classifications, but it doesn't lead in any specific category.

Explainable AI (XAI) Techniques

In medical applications, especially in areas like skin cancer detection, it is crucial for AI models to provide not just accurate predictions but also explanations that can be understood by clinicians. **Explainable AI (XAI)** addresses the "black box" nature of deep learning models by highlighting how a model arrives at its decision. In this project, we used two widely recognized XAI techniques—**Grad-CAM** and **SmoothGrad**—to improve the interpretability of the predictions made by InceptionV3, Xception, and EfficientNet models.

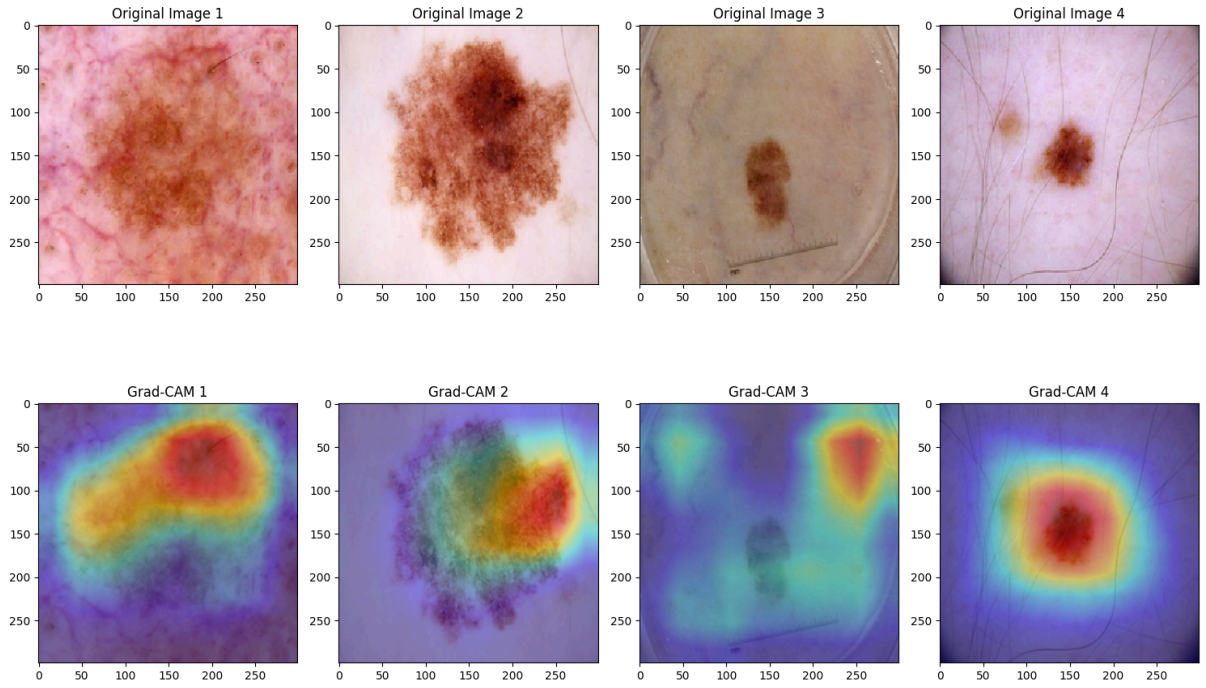
1. Gradient-weighted Class Activation Mapping (Grad-CAM)

Grad-CAM is an XAI technique that produces a coarse localization map of the important regions in an image by computing the gradient of the predicted class score with respect to the feature maps of the convolutional layers. It highlights the areas in the image that the model focused on to make its classification decision.

How Grad-CAM Works:

- The Grad-CAM algorithm computes the gradients of the class score (e.g., benign or malignant) with respect to the feature maps of the last convolutional layer.
- These gradients are then pooled and used to weight the importance of each feature map.
- Finally, a heatmap is generated, showing which parts of the input image were most relevant for the model's decision.

For each skin lesion image, Grad-CAM was applied to visualize which regions of the image contributed most to the classification as either benign or malignant. These heatmaps were overlaid on the original images to help dermatologists understand which areas the model considered important.



Benefits:

- Grad-CAM provided insight into the model's decision-making process, showing that certain areas of a lesion were consistently highlighted when classified as malignant.
- Clinicians can use these heatmaps as a second opinion, increasing their trust in the AI system.

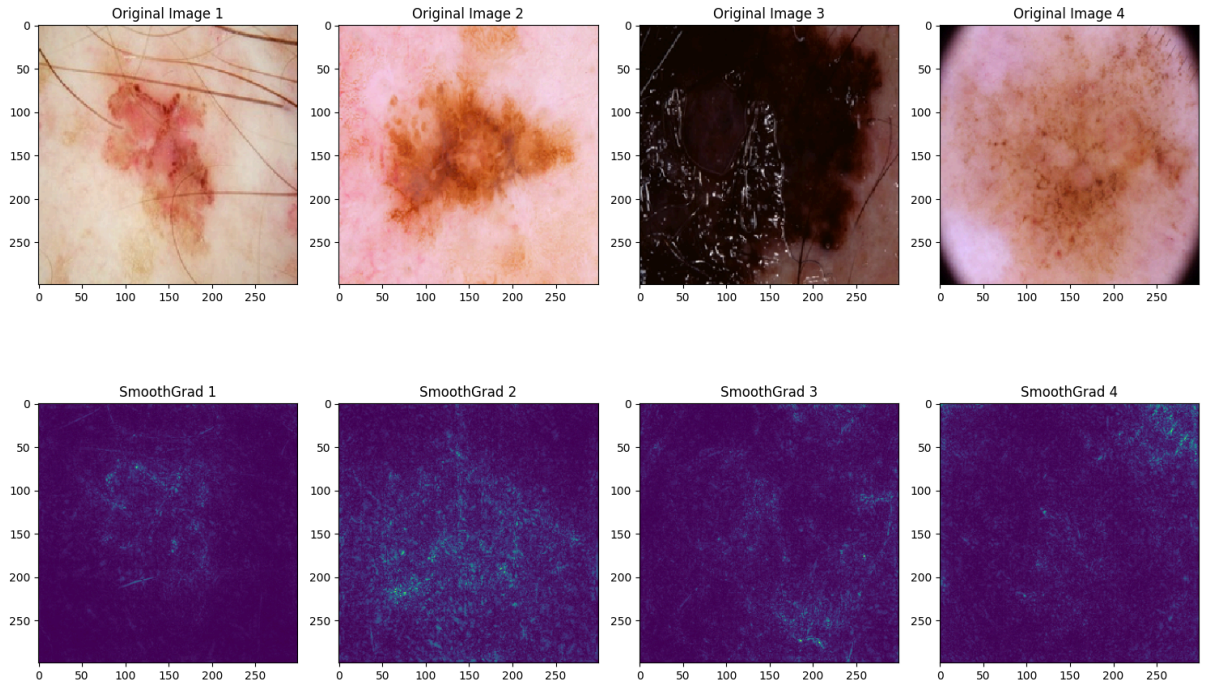
2. SmoothGrad

SmoothGrad enhances the interpretation of deep learning models by visualizing the input features that contribute the most to the model's predictions. It does this by generating noisy versions of the input image, running these through the model, and averaging the gradients to reduce noise and sharpen the explanations.

How SmoothGrad Works:

- SmoothGrad takes multiple noisy copies of the input image by adding random Gaussian noise.
- It computes the gradient of the class score with respect to the input image for each noisy copy.
- The gradients are averaged to produce a more stable and visually interpretable gradient map.

SmoothGrad was used to generate detailed saliency maps for each input image, highlighting the pixels that were most influential in the model's decision. These saliency maps helped provide a clearer understanding of how the model differentiated between benign and malignant lesions.



Benefits:

- SmoothGrad produces clearer and more focused explanations compared to basic gradient-based methods.
- It reduces visual noise in the explanations, making it easier for clinicians to interpret which parts of a lesion image were influential in the model's decision.

Comparing Grad-CAM and SmoothGrad

- **Grad-CAM** provides localized explanations by highlighting regions of the image, which is particularly useful when working with convolutional layers in image classification tasks.
- **SmoothGrad** offers pixel-level insights, helping to highlight the fine details that contribute to a model's decision.

Both techniques complement each other in terms of interpretability:

- **Grad-CAM** excels at providing high-level region-based explanations.
- **SmoothGrad** excels at detailed, pixel-level explanations.

Results of XAI Techniques

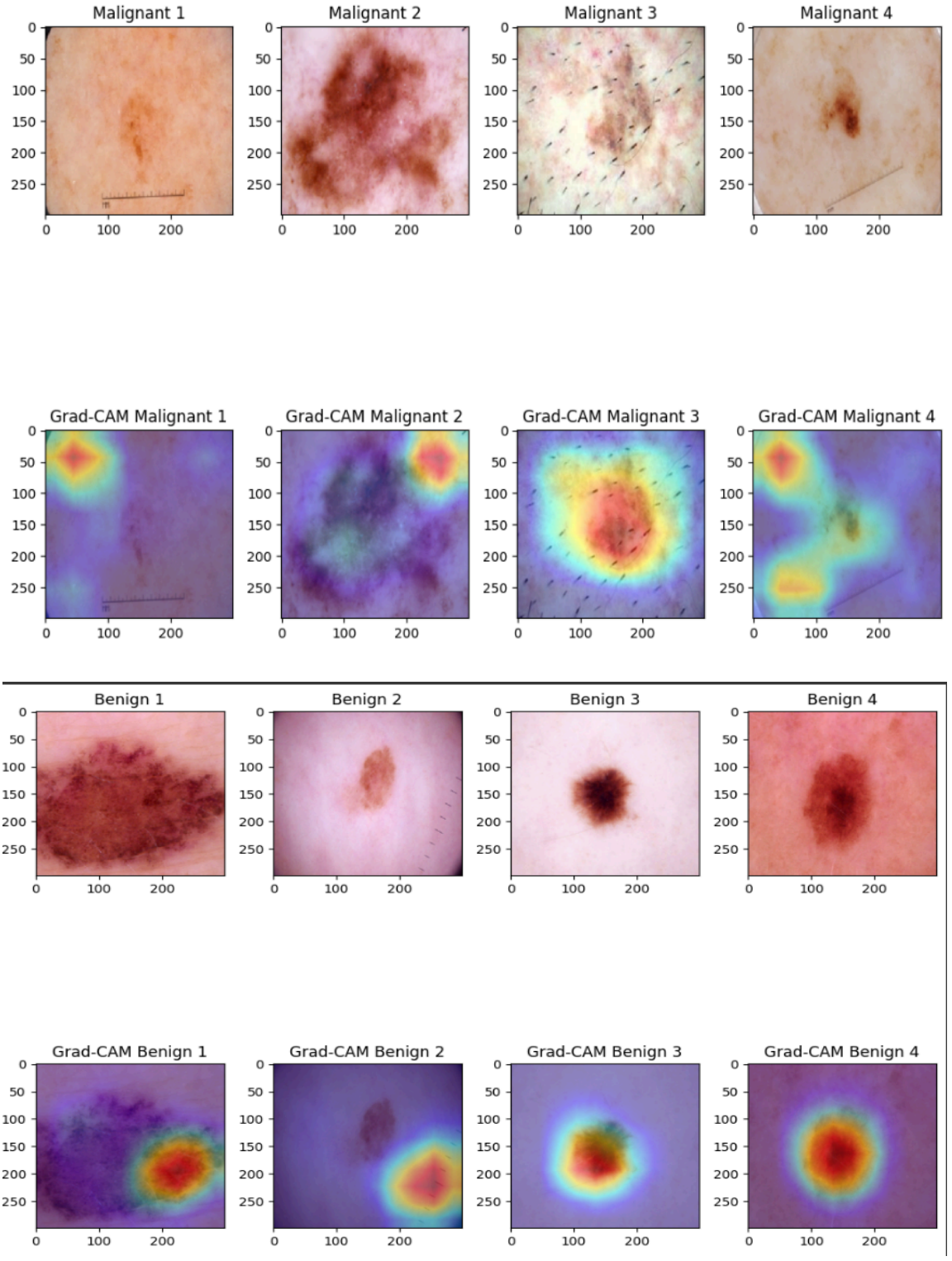
By applying Grad-CAM and SmoothGrad to our models (InceptionV3, Xception, and EfficientNet), we were able to:

- **Highlight critical regions** in the skin lesion images that influenced the model's decisions.
- **Increase model transparency**, making it easier for healthcare professionals to trust the AI system.

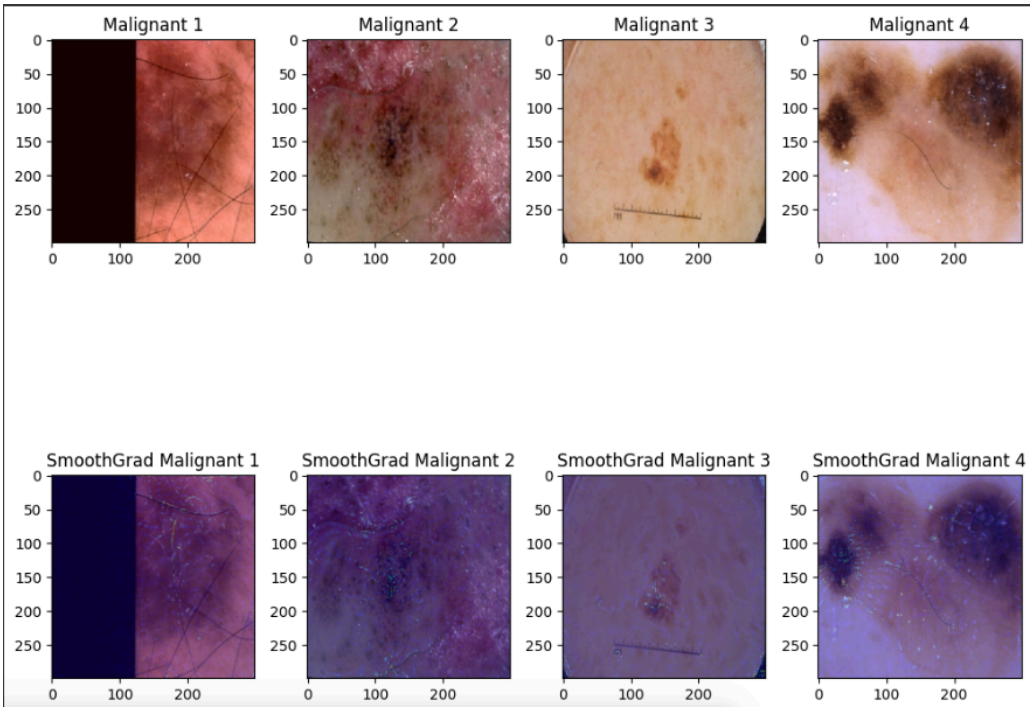
- **Validate model behavior**, ensuring that the AI is focusing on medically relevant parts of the lesion (e.g., color changes, irregular borders).

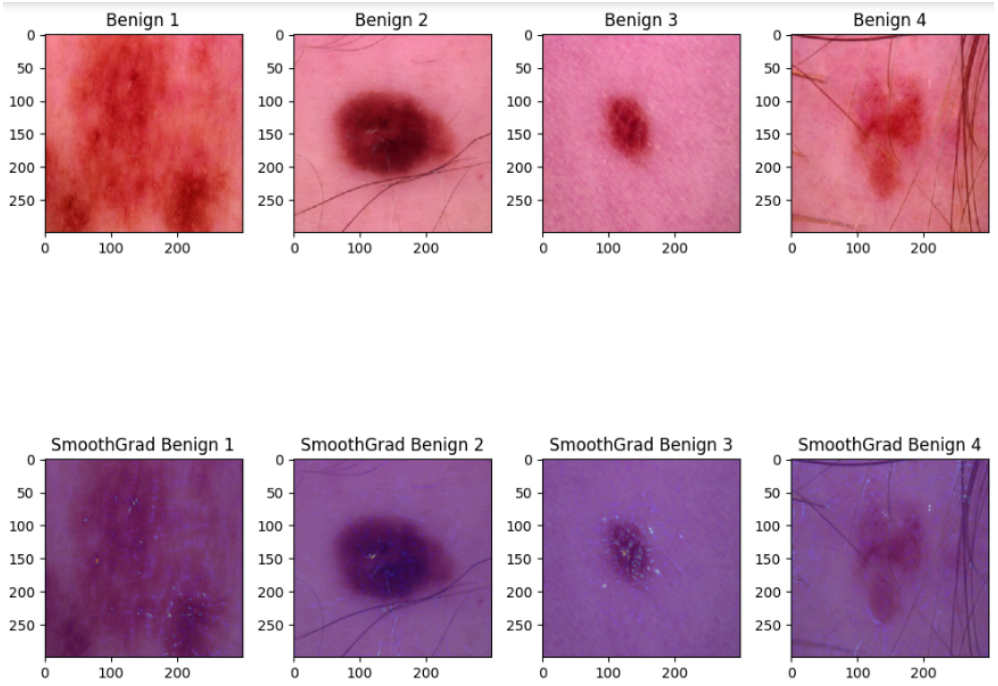
Example Visualizations:

- In cases classified as malignant, Grad-CAM heatmaps often highlighted the irregular edges and darker spots on the lesion, which are key features used in human diagnosis.



- SmoothGrad visualizations showed the fine-grained features in the lesion that had the highest influence on the model's decision, such as texture and minor color variations.





Conclusion and Future Works

In this project, we applied deep learning models—InceptionV3, Xception, and EfficientNet—to classify skin lesions, while using Grad-CAM and SmoothGrad for explainability. The Xception model stood out for its balanced performance in detecting both benign and malignant lesions, making it a strong candidate for real-world medical use. By using explainable AI techniques, we were able to highlight important areas in the images that influenced the models' decisions, ensuring the focus was on medically relevant features. This not only improves accuracy but also builds trust in AI-driven diagnostics, making them more transparent and easier for healthcare professionals to rely on.

Looking ahead, we aim to fine-tune these models to further improve their accuracy and ensure even better performance, particularly in more complex cases. We also plan to explore other explainable AI methods, like SHAP and LRP, to provide more detailed insights into the models' decisions. Additionally, we envision developing a real-time application that integrates these AI models with an easy-to-use explainability dashboard for clinical use. Expanding the dataset to include more diverse skin lesions and exploring hybrid models that combine traditional machine learning and deep learning will be key steps in enhancing the models' effectiveness and applicability in healthcare.

References

1. Dagnaw, G. H., El Mouhtadi, M., & Mustapha, M. (2024). Skin cancer classification using vision transformers and explainable artificial intelligence. Journal of Medical Artificial Intelligence.
2. Gamage, L., Isuranga, U., Meedeniya, D., De Silva, S., & Yogarajah, P. (2024). Melanoma skin cancer identification with explainability utilizing mask guided technique. Electronics, 13(4), 680.
3. Mosquera-Zamudio, A., Launet, L., Tabatabaei, Z., Parra-Medina, R., Colomer, A., Oliver Moll, J., ... & Naranjo, V. (2022). Deep learning for skin melanocytic tumors in whole-slide images: A systematic review. Cancers, 15(1), 42.
4. Nigar, N., Umar, M., Shahzad, M. K., Islam, S., & Abalo, D. (2022). A deep learning approach based on explainable artificial intelligence for skin lesion classification. IEEE Access, 10, 113715-113725.
5. Attallah, O. (2024). Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning. Computers in Biology and Medicine, 178, 108798.
6. Lucieri, A., Dengel, A., & Ahmed, S. (2021). Deep learning based decision support for medicine–a case study on skin cancer diagnosis. arXiv preprint arXiv:2103.05112.
7. Mridha, K., Uddin, M. M., Shin, J., Khadka, S., & Mridha, M. F. (2023). An interpretable skin cancer classification using optimised convolutional neural network for a smart healthcare system. IEEE Access, 11, 41003-41018.
8. Alche, M. N., Acevedo, D., & Mejail, M. (2021). EfficientARL: improving skin cancer diagnoses by combining lightweight attention on EfficientNet. Proceedings of the IEEE/CVF International Conference on Computer Vision, 3354-3360.
9. Azlaan, H., & Oluwaseyi, J. (2024). Explainable AI (XAI) for Skin Cancer Detection. ResearchGate.

10. Owen, J., & Olaoye, G. O. (2024). A Comparative Study of Explainable AI in Skin Cancer and Blood Cell Detection: Building Trust in Medical Diagnosis. ResearchGate.
11. Gautam, Y., Gupta, P., Kumar, D., Kalra, B., Kumar, A., & Hemanth, J. D. (2024). FusionEXNet: an interpretable fused deep learning model for skin cancer detection. International Journal of Computers and Applications, 1-11.
12. Mohan, J., Sivasubramanian, A., Sowmya, V., & Vinayakumar, R. (2024). Enhancing Skin Disease Classification Leveraging Transformer-based Deep Learning Architectures and Explainable AI. arXiv preprint arXiv:2407.14757.
13. Rezk, E., Eltorki, M., & El-Dakhakhni, W. (2023). Interpretable skin cancer classification based on incremental domain knowledge learning. Journal of Healthcare Informatics Research, 7(1), 59-83.
14. Pintelas, E., Liaskos, M., Livieris, I. E., Kotsiantis, S., & Pintelas, P. (2021). A novel explainable image classification framework: Case study on skin cancer and plant disease prediction. Neural Computing and Applications, 33(22), 15171-15189.
15. Fontes, M., De Almeida, J. D. S., & Cunha, A. (2024). Application of example-based explainable artificial intelligence (XAI) for analysis and interpretation of medical imaging: a systematic review. IEEE Access, 12, 26419-26427.