International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

SKIN CANCER DETECTION USING DERMOSCOPIC IMAGES USING MACHINE LEARNING

SAFRIN BANU J[1], ATCHAYAALAKSHMI D[2], GURU AADITHYA S[3], SRI SAASHWAUTH N[4]

[1]*Student, Dept. of Information Science and Engineering, Bannari Amman Institute of Technology, IN*
[2]*Student, Dept. of Artificial Inteligence and Data Science Engineering, Bannari Amman Institute of Technology, IN*
[3]*Student, Dept. of* Electronics and Communication *Engineering, Bannari Amman Institute of Technology, IN*
[4]*Student, Dept. of* Electronics and Communication *Engineering, Bannari Amman Institute of Technology, IN*

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 Xception 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.

1. INTRODUCTION

Skin cancer is one of the most common cancers globally, affecting millions of people every year. The early detection of skin cancer, particularly in its malignant forms, is essential as it can significantly improve treatment outcomes and save lives. Traditional diagnostic approaches involve visual examination and, if necessary, biopsy by dermatologists. However, even for experienced practitioners, distinguishing between benign and malignant lesions based solely on appearance can be challenging, leading to potential diagnostic errors. In recent years, artificial intelligence (AI) has emerged as a powerful tool to assist medical professionals by offering accurate and rapid diagnostics. This project leverages AI and, more specifically, deep learning, to classify skin lesions, enhancing the efficiency and accuracy of skin cancer diagnosis.

1.1 Problem Statement: The Need for Explainable AI

While deep learning models have demonstrated remarkable accuracy in image classification tasks, a significant challenge in using these models for clinical applications is their "black box" nature. Deep learning algorithms are often highly complex, relying on numerous layers and vast amounts of data to make predictions. As a result, they can make decisions without providing explanations, leaving healthcare professionals uncertain about the reasoning behind specific diagnoses. This lack of transparency poses a barrier to their adoption in the medical field, where trust, understanding, and interpretability are crucial for successful implementation. Therefore, this project aims to address this gap by incorporating explainability techniques, known as Explainable AI (XAI), to make the models' decision-making process more transparent and accessible to healthcare professionals.

1.2 Objectives

The primary objective of this project is to develop an AI-based system capable of accurately classifying skin lesions as either benign or malignant, making it a useful tool in clinical settings. To achieve this, the system integrates three widely recognized deep learning models: InceptionV3, Xception, and EfficientNet. Additionally, to enhance interpretability, we apply explainable AI techniques such as Grad-CAM and SmoothGrad. These techniques aim to provide visual explanations that highlight the areas in skin lesion images which influenced the models' predictions. By improving both accuracy and interpretability, this project seeks to bridge the gap between technical performance and clinical usability, ultimately contributing to safer and more trustworthy AI-driven diagnostics.

1.3 Scope of the project

The scope of this project includes the training, testing, and evaluation of the InceptionV3, Xception, and EfficientNet models on a labeled dataset of skin lesions. In addition to

model accuracy, we assess the explainability of each model's predictions using Grad CAM and SmoothGrad

visualizations. Our analysis includes a detailed comparison of the models in terms of both performance metrics (such as precision, recall, and F1-score) and the quality of their explanations. By providing both quantitative and qualitative insights, this project lays the groundwork for future improvements in AI-based dermatological diagnostics, with the potential for real-world applications in clinical settings.

2. MODELS AND METHODOLOGY

Deep Learning Models

In this project, we used three state-of-the-art deep learning models: InceptionV3, Xception, and EfficientNet. These models are well-regarded for their performance in image classification tasks and have been adapted to classify skin lesions as benign or malignant.

2.1 InceptionV3

2.1.1 Architecture: InceptionV3 utilizes a combination of various kernel sizes in its convolutional layers, allowing it to capture features at multiple scales. It includes Inception modules, which are sub-networks that process images in parallel using different convolutional filters (e.g., 1x1, 3x3, and 5x5).

2.2.2 Advantages: High accuracy in image classification. Capable of handling images with diverse features and scales due to its multi-path approach.

2.2.3 Disadvantages: Computationally expensive and requires a significant amount of memory. Complex architecture makes it harder to interpret.

2.2.4 Applications: Suitable for general-purpose image classification tasks where a balanced performance is required.

2.2 Xception

2.2.1 Architecture: Xception, short for "Extreme Inception," replaces traditional convolution layers with depthwise separable convolutions, which separates the spatial convolution and channel-wise convolution. It uses residual connections to maintain gradient flow.

2.2.2 Advantages: More efficient than traditional convolutional networks due to its lightweight architecture. Good at capturing complex features from images.

2.2.3 Disadvantages: 4 Requires powerful hardware for training, especially for large datasets. May be prone to overfitting if not properly regularized.

2.2.4 Applications: Ideal for tasks that need a model capable of detailed feature extraction, such as medical image analysis.

2.3 EfficientNet

2.3.1 Architecture: EfficientNet uses a novel scaling method called compound scaling, which uniformly scales the model depth, width, and resolution based on a fixed set of parameters. It is optimized for both accuracy and efficiency.

2.3.2 Advantages: Highly efficient, with fewer parameters compared to other deep learning models. Scalable for different input sizes, making it versatile for various applications.

2.3.3 Disadvantages: Can be sensitive to input resolution; performance may drop if not properly scaled. Might not perform as well on datasets with high complexity without sufficient input resolution.

2.3.4 Applications: EfficientNet is useful for resource-constrained environments where computational efficiency is a priority.

Explainable AI (XAI) Techniques

To address the "black-box" nature of deep learning models and make their predictions more interpretable, we incorporated two popular XAI techniques: Grad-CAM and SmoothGrad..

2.4 Gradient-weighted Class Activation Mapping (Grad-CAM)

2.4.1 How it works: Grad-CAM generates heatmaps that highlight the areas of an input image that the model focuses on when making a prediction. It calculates gradients of the target class score concerning the feature maps of a convolutional layer.

2.4.2 Advantages: Provides visual insights into the model's decision-making process. Helps in identifying regions of interest that contributed most to the prediction.

2.4.3 Disadvantages: The generated heatmaps can be coarse and might not focus on fine details.Limited to convolutional neural networks and might not generalize to non-CNN models.

2.4.4 Applications: Useful in medical imaging to show clinicians which parts of an image influenced a model's diagnosis, enhancing trust and understanding.

2.5 SmoothGrad

2.5.1 How it works: SmoothGrad improves the quality of gradient-based explanations by adding noise to the input image multiple times and averaging the resulting gradient maps. This reduces the noise and provides clearer visualizations of the areas that influenced the prediction.

2.5.2 Advantages: Produces sharper and more focused explanations by averaging gradients across noisy variations of the input. Reduces sensitivity to noise, making explanations more stable.

2.5.3 Disadvantages: Requires generating multiple noisy samples, increasing computation time. May still struggle with subtle variations in complex datasets.

2.5.4 Applications: Helps in refining visual explanations, making it easier for clinicians to interpret subtle features that may indicate malignancy.

2.6 Summary

2.6.1 InceptionV3 provides a balanced performance and is useful for tasks requiring reliable, general-purpose classification.
2.6.2 Xception excels in capturing complex features with its depthwise separable convolutions, making it effective in handling detailed image analysis tasks.
2.6.3 EfficientNet is highly efficient and performs well across different input sizes, suitable for scenarios with limited computational resources.
2.6.4 Grad-CAM offers region-based explanations, making it easy for clinicians to understand which areas of the image the model focused on.
2.6.5 SmoothGrad provides finer, pixel-level explanations, useful for highlighting subtle but critical features in the image. By combining high-performing deep learning models with XAI techniques, our project aims to deliver an accurate and interpretable diagnostic tool for skin cancer detection, enhancing the trustworthiness and clinical usability of AI systems in healthcare.

2.7 Methodology

2.7.1 Data Collection & Preprocessing

2.7.1.1 Algorithm: Use OpenCV for loading and resizing images. Convert images to NumPy arrays for processing.
2.7.1.2 Steps: Resize images to match the input sizes required by each model (128x128, 224x224, 299x299). Apply augmentation techniques to enhance training data.

2.7.2 Deep Learning Model Development

2.7.2.1 Algorithm: Develop and utilize EfficientNet, InceptionV3, and Xception architectures.
2.7.2.2 Steps:
EfficientNet: Optimized for smaller input sizes, balancing speed and accuracy.
InceptionV3: Suitable for handling larger input sizes, enabling detailed feature extraction.
Xception: Uses depthwise separable convolutions for improved performance with complex patterns.

2.7.3 Model Training & Testing

2.7.3.1 Algorithm: Adam optimizer with cross-entropy loss.

2.7.3.2 Steps: Split data into training and validation sets. Monitor accuracy, precision, recall, and F1-score during training. Apply early stopping or learning rate scheduling for optimization.

2.7.4 Prediction & Classification

2.7.4.1Steps: Use trained models to predict if the lesion is malignant or benign. Output the classification results.

2.7.5 Explainability using Grad-CAM and SmoothGrad

2.7.5.1 Algorithm: Generate visual explanations using Grad-CAM and SmoothGrad techniques.
2,7.5.2 Steps: 42 Grad-CAM: Produces heatmaps highlighting important regions in the images, showing where the model focuses.
SmoothGrad: Enhances Grad-CAM outputs by reducing noise, making the heatmaps clearer and more interpretable for clinicians.

2.7.6 Evaluation & Performance Metrics

2.7.6.1 Accuracy: Measures the proportion of correct predictions.
2.7.6.2 Precision & Recall: Evaluates the model's focus on malignant cases.
2.7.6.3 F1-score: Balances precision and recall for an overall effectiveness score.
2.7.6.4 Explainability Evaluation: Quality and clarity of Grad-CAM and SmoothGrad outputs assessed for clinical usefulness.

3. FINDINGS AND INTERPREATATION

3.1 Xception Model

3.1.1 Strengths: The Xception model demonstrates a well-balanced performance for classifying both benign and malignant skin lesions. It achieves high precision—0.87 for benign and 0.90 for malignant classifications—indicating that a large proportion of its positive predictions (either benign or malignant) are correct. Furthermore, its recall for benign lesions is 0.92, meaning it successfully identifies 92% of true benign cases, while for malignant lesions, recall is 0.84, capturing a substantial proportion of true malignant cases. These results yield high F1-scores of 0.89 for benign and 0.87 for malignant, leading to an overall accuracy of 0.88. This high accuracy makes the Xception model a strong candidate for clinical applications, as it maintains a robust balance across both sensitivity and specificity, minimizing the risk of false positives and false negatives.

3.1.2 Weaknesses: While Xception is overall well-balanced, it is computationally more intensive than some

of the other models, which may limit its deployment in resource constrained settings, such as mobile or remote diagnostic systems.

3.1.3 Reasoning: Xception's balanced performance is largely attributed to its unique architecture, which uses

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

depthwise separable convolutions instead of traditional convolutions. This approach allows the model to capture complex, high-level features more efficiently without a significant increase in computational cost. Depthwise separable convolutions separate spatial filtering from feature combination, enabling the model to analyze intricate patterns within the skin lesion images. This architectural advantage likely contributes to its effective handling of both benign and malignant cases, as it can capture detailed variations in lesion texture, shape, and color. Consequently, the Xception model offers high accuracy and generalizability, making it an optimal choice for balanced diagnostic tasks in dermatology.

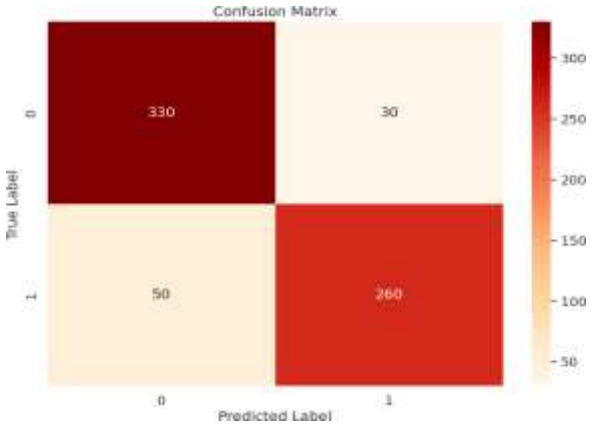


Fig 3.1 Xception model confusion matrix

3.2 EfficientNet (128x128 Input Size):

3.2.1 Strengths: EfficientNet with a 128x128 input size shows strong precision in identifying benign lesions,

achieving a precision score of 0.91. This indicates that when the model classifies a lesion as benign, it is accurate in 91% of cases, which is essential for reducing false positives in clinical settings. For malignant classifications, EfficientNet has a precision of 0.76 and a recall of 0.92, indicating that it is particularly adept at identifying true malignant cases, capturing 92% of actual malignant lesions. This high malignant recall is beneficial in a clinical setting, where missing a malignant case could lead to delayed treatment and worse patient outcomes.

3.2.2 Weaknesses: The model's recall for benign cases is somewhat lower at 0.75, indicating that it misses a quarter of true benign lesions, which may result in some benign cases being incorrectly flagged as potentially malignant. This trade-off could lead to unnecessary follow-ups or additional testing, placing a burden on both patients and healthcare systems. The low benign recall implies a risk of over-diagnosing benign cases, potentially causing anxiety for patients who are falsely classified as at risk.

3.2.3 Reasoning: EfficientNet achieves a balance between model depth, width, and resolution using a technique known as compound scaling, which optimizes performance without significantly increasing computational demands. However, with a smaller input size of 128x128, some fine-grained details in skin lesions may be lost, impacting the model's ability to capture subtle features in benign cases. The model's architecture is effective at recognizing malignant patterns, likely because malignant lesions often have distinct characteristics that are identifiable even in lower resolutions. Nevertheless, the lack of detailed input might explain its lower sensitivity for benign cases, as it may not pick up on the finer distinctions required to differentiate subtle benign features. Overall, EfficientNet at this input size performs well but has limitations in benign classification, making it more suitable for environments where malignant detection is prioritized.

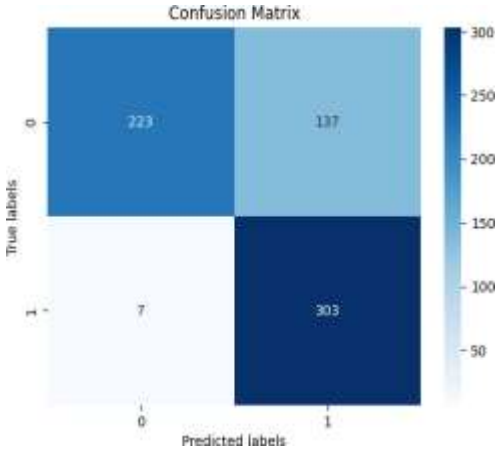


Fig 3.2 EfficientNet model confusion matrix

3.3 EfficientNetB0 (224x224 Input Size):

3.3.1 Strengths: When trained with an input size of 224x224, EfficientNetB0 excels in detecting malignant lesions, achieving an impressive recall of 0.98. This means that almost all malignant cases are correctly identified by the model, making it a highly reliable option for applications where detecting malignant lesions is the top priority. Additionally, it achieves a high precision of 0.97 for benign lesions, indicating that the model is effective in correctly identifying benign cases when it predicts them as such.

3.3.2 Weaknesses: The primary limitation of EfficientNetB0 at this input size is its recall for benign lesions, which is relatively low at 0.62. This means that 38% of benign cases are missed, leading to a considerable number of false positives. Such a shortcoming could contribute to patient anxiety, as benign lesions might be unnecessarily flagged as malignant, resulting in potentially avoidable biopsies or further investigations. This limitation may also impact resource allocation in

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

clinical settings, as additional follow ups for benign cases can strain healthcare systems.

3.3.3 Reasoning: EfficientNetB0's use of a larger input size (224x224) allows it to capture more detailed features in skin lesions, especially those indicative of malignancy, which explains its high malignant recall. The increase in input resolution enables the model to focus on subtle malignant characteristics, such as irregular borders, varied pigmentation, and asymmetry. However, this increased focus on malignant features may come at the cost of benign classification, as the model's attention to malignant patterns can overshadow the subtler details associated with benign lesions. This trade-off highlights EfficientNetB0's suitability for applications that prioritize malignant detection over benign accuracy, such as initial screenings where capturing all possible malignant cases is essential.

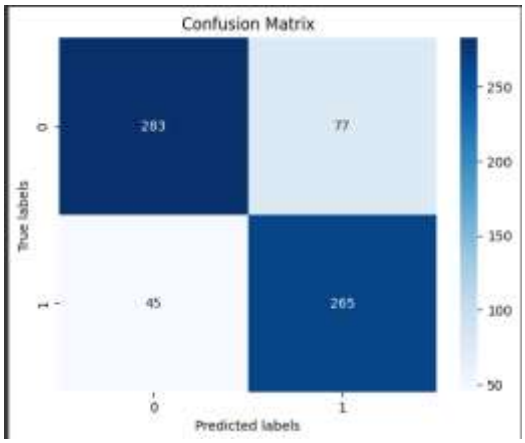


Fig 3.3 EfficientNetB0 model confusion matrix

3.4 InceptionV3:

3.4.1 Strengths: InceptionV3 demonstrates a balanced performance across both benign and malignant classifications, with an overall accuracy of 0.82. The model achieves a precision of 0.86 and recall of 0.79 for benign lesions, while for malignant lesions, it scores 0.77 for precision and 0.85 for recall. This balance ensures that InceptionV3 does not overemphasize either class, making it a dependable choice for general-purpose diagnostic tasks. By achieving relatively even scores in precision and recall for both classes, InceptionV3 offers a reliable model that can be deployed in situations requiring a balanced approach.

3.4.2 Weaknesses: Despite its steady performance, InceptionV3 does not lead in any specific metric, which might make it less appealing for tasks that require specialization, such as maximizing malignant recall or benign precision. While the model provides consistent results, it may not be the top choice for applications where extremely high accuracy in one class is critical. 23

3.4.3 Reasoning: The architecture of InceptionV3 is based on a multi-path approach, which processes images at multiple resolutions simultaneously. This design helps the model capture different spatial features and scale variations within images, allowing it to adapt to both benign and malignant cases. In dermatological contexts, where lesions can vary greatly in size and texture, this flexibility contributes to InceptionV3's balanced performance. However, the multi-path structure also leads to an averaging effect, as it does not prioritize any specific feature set over others. This design choice might explain why InceptionV3 achieves steady performance without excelling in any particular metric. It is, therefore, best suited for diagnostic workflows where a balanced, all-purpose model is needed rather than one that specializes in a single type of classification.

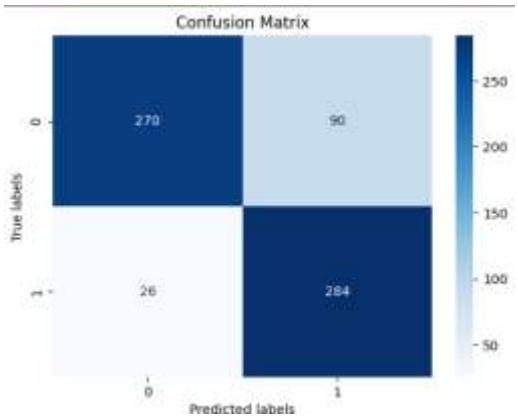


Fig 3.4 Inception model confusion matrix

4.ANALYSIS AND UNDERSTANDING:

4.1 Best Overall Model: Xception Model

The Xception model emerged as the most balanced among the models tested in this project. With strong precision and recall for both benign and malignant classifications, it achieves an overall accuracy of 0.88, making it highly reliable for clinical use. Its performance indicates that it minimizes both false positives and false negatives, which is crucial in a medical setting where diagnostic accuracy directly impacts patient outcomes. Xception's architecture, featuring depthwise separable convolutions, allows the model to capture complex features efficiently

without overburdening computational resources, thus offering a practical balance between accuracy and efficiency.This balanced capability makes Xception particularly suitable for clinical applications where high sensitivity (recall) and high specificity (precision) are needed to avoid both missed diagnoses and unnecessary follow-ups. Given its high F1-scores across both benign and malignant classes, Xception can be trusted to deliver consistent performance, aiding dermatologists in reliably

identifying both non-cancerous and cancerous skin lesions.

4.2 Best for Malignant Detection: EfficientNetB0

EfficientNetB0, with an input size of 224x224, proved to be the most effective model for detecting malignant lesions, achieving an exceptionally high recall of 0.98. This high recall rate indicates that the model captures nearly all malignant cases, which is particularly valuable in cancer detection where missing a malignant case could have serious consequences. For clinical applications prioritizing malignant detection, EfficientNetB0 stands out as it can serve as a robust early screening tool, catching nearly every potential cancerous case. However, its lower recall for benign lesions (0.62) means that some benign lesions might be misclassified as malignant, which could lead to over-diagnosis and unnecessary follow-ups. While this is a drawback, it can be acceptable in scenarios where the priority is to detect every possible malignant case, even at the cost of benign lesion misclassification. EfficientNetB0's ability to focus on malignant features is likely enhanced by its larger input size, allowing the model to detect minute patterns and irregularities associated with malignancy, such as asymmetric borders, uneven coloration, or rapid lesion growth.

4.3 Best for Balanced Performance: InceptionV3

The InceptionV3 model offers a balanced performance that does not excel in any single metric but achieves solid results across both benign and malignant classifications. With an overall accuracy of 0.82 and reasonably high precision and recall scores for both categories, InceptionV3 is suitable for general-purpose applications where a balanced diagnostic tool is required. Its ability to handle both benign and malignant cases without favoring one over the other makes it versatile for diverse clinical workflows. InceptionV3's architecture uses a multi-path approach, which processes images at various spatial resolutions, allowing the model to capture features across different scales. This makes it adaptable to variations in lesion size, shape, and texture, contributing to its balanced performance. Although it doesn't achieve the highest precision or recall for any single category, its consistent accuracy makes InceptionV3 a reliable model for clinical applications that require steady, all-around performance.

4.4 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.

4.4.1 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.

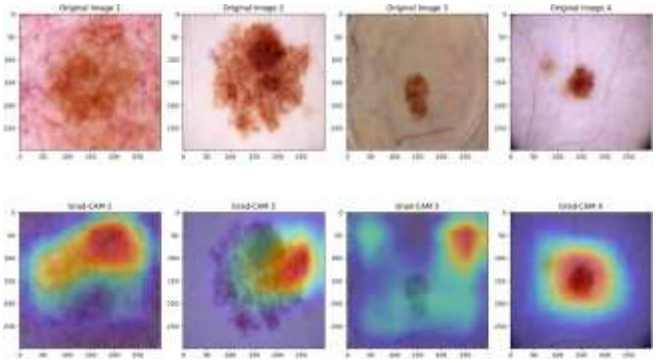


Fig 4.4.1 GRAD-CAM Analysis

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.

4.4.2 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

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

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.

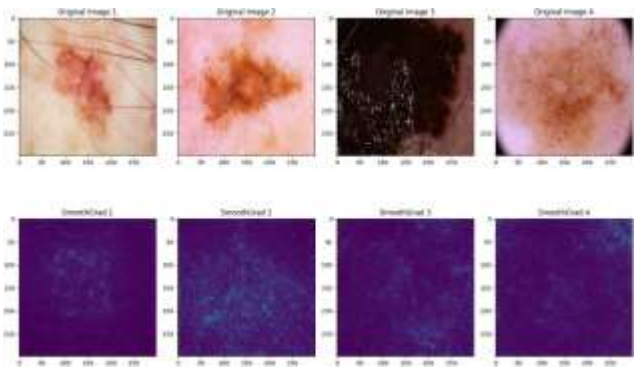


Fig 4.4.2 SmoothGrad Analysis

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.

4.4.3 Results of Applying XAI Techniques to the Models

By implementing Grad-CAM and SmoothGrad on the InceptionV3, Xception, and EfficientNet models, several benefits were observed:

Highlighting Critical Regions: Both Grad-CAM and SmoothGrad effectively identified significant areas within skin lesion images that influenced model decisions. These insights validated that the models were focusing on relevant features, such as lesion color changes, irregular shapes, and darker spots—key characteristics for dermatological assessment.

Enhancing Model Transparency: The application of Grad-CAM and SmoothGrad improved the transparency of each model, making it easier for healthcare professionals to trust and understand the AI's reasoning. Transparent models are more likely to be accepted in clinical settings, as they allow clinicians to interpret and validate the AI's focus and rationale.

Validating Model Behavior: Both techniques allowed for the validation of model behavior by ensuring that the AI focused on medically relevant parts of the lesions. For example, the heatmaps and saliency maps confirmed that the AI 30 models paid attention to critical aspects such as asymmetry, color variation, and border irregularities, which are commonly used criteria for identifying malignant skin lesions in dermatology.

5.VISUALIZATIONS

5.1 Grad-CAM Visualizations:

Gradient-weighted Class Activation Mapping (Grad-CAM) is an XAI technique used to create heatmaps that indicate the regions of an image that contributed most to the model's prediction. The primary goal of Grad-CAM is to highlight key areas that the model considers important when classifying an image as either malignant or benign.

Key Observations in Malignant Cases: In images classified as malignant, Grad-CAM heatmaps often highlighted irregular edges and darker spots within the lesion. These are crucial features used by dermatologists to identify potential skin cancer. Irregular Borders: Grad-CAM frequently showed high activation around uneven or asymmetrical borders of lesions. This is consistent with medical knowledge, as asymmetry is a common indicator of melanoma. Dark Patches: The technique also focused on dark or pigmented areas within the lesion. Malignant lesions often display uneven color distribution, with darker areas indicating potential malignancy due to irregular growth patterns.

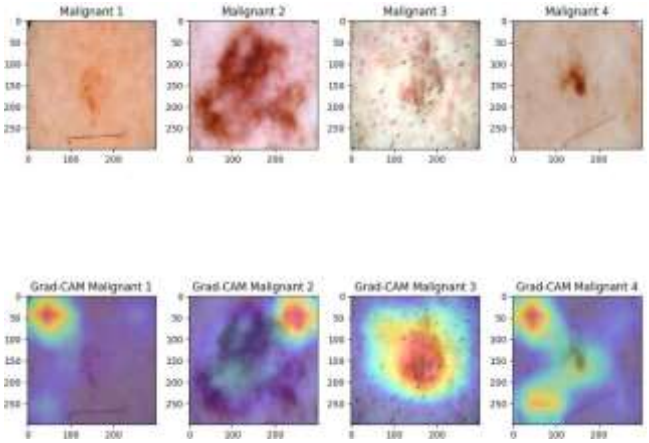
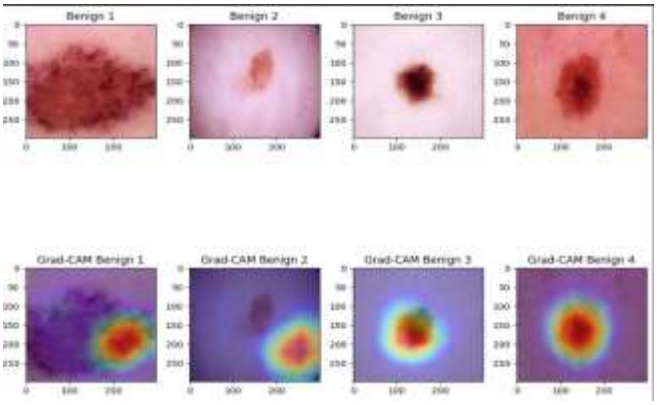


Fig 5.1 Malignant - GRAD-CAM Visualization



International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Fig 5.2 Benign - GRAD-CAM Visualization

5.2 SmoothGrad Visualizations

SmoothGrad is another XAI technique designed to improve gradient-based visual explanations by averaging over multiple noisy versions of the input image. This method reduces noise in the visualization, resulting in sharper, more focused explanations that highlight the most influential features at a finer level.

Key Observations in Malignant Cases: In malignant cases, SmoothGrad visualizations often showed emphasis on fine-grained features such as subtle texture changes and minor color variations within the lesion. Texture Patterns: Malignant lesions tend to have irregular textures, which can be indicative of uneven cell growth. SmoothGrad effectively highlights these textural differences by focusing on areas with high gradient influence. Color Variations: Small variations in color, such as the presence of tiny dark spots or uneven shading, are critical for distinguishing between benign and malignant lesions. SmoothGrad's pixel-level sensitivity helps bring out these subtle details, making it a valuable tool for identifying early signs of malignancy.

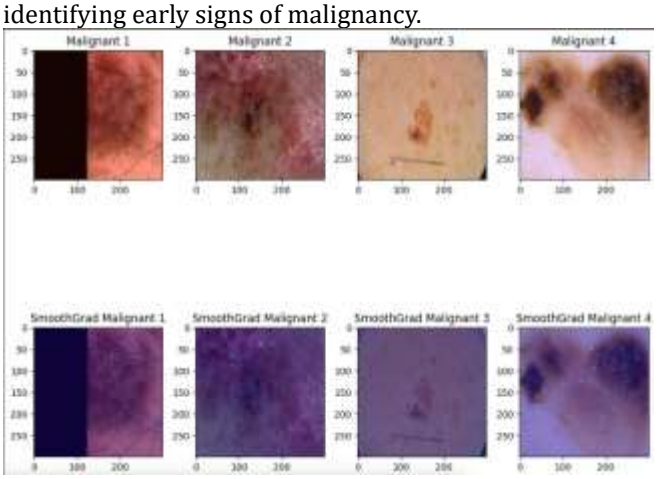


Fig 5.3 Malignant – SmoothGrad Visualization

Fig 5.4 Benign – SmoothGrad Visualization

6.MATHEMATICAL EQUATIONS:

6.1 Grad-CAM:

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

y^c output score for the target class c.
A_{ij}^k represents k-th feature map
$\frac{\partial y^c}{\partial A_{ij}^k}$ is the partial derivative of output y^c w.r.t the activation A_{ij}^k
Z is a normalization factor (number of elements in A_{ij}^k).

Intuition and Explanation:

Gradients: The partial derivatives measure how much the class score changes with respect to the activations at each position in the feature map. Larger values indicate that the model relies more heavily on that particular region of the feature map to make its prediction for class c.
Weighted Sum: The double sum is used to aggregate the gradients over all spatial positions in the feature map. This aggregation ensures that regions of the image corresponding to larger gradients have a greater influence on the final class activation map.

Normalization: The factor Z normalizes the computed weights, ensuring that the overall importance across feature maps is scaled properly. This normalization ensures the sum of the weights doesn't grow disproportionately with the size of the feature map.

6.2 SmoothGrad:

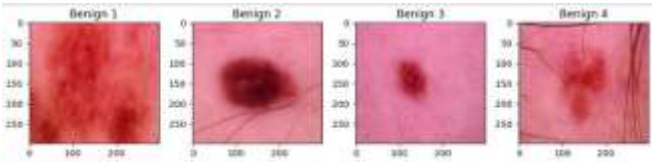
$$\text{SmoothGrad}(x) = \frac{1}{N}\sum_{i=1}^{n} \frac{\partial y^c}{\partial (x+N(0,\sigma^2)}$$

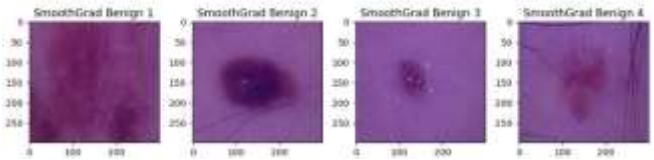
x is the input image.

- N is the number of noisy samples generated by adding Gaussian noise $N(0,\sigma^2$ (to x.
- $\frac{\partial y^c}{\partial x}$ represents the gradient of the class score ∂y^c with respect to the input image.
- σ is the standard deviation of the Gaussian noise applied to each sample.

Intuition and Explanation:

Noise Addition: By adding Gaussian noise to the input image, we create multiple noisy versions of the image.





This helps to capture the generalizable patterns in the

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

gradients, filtering out any high-frequency noise or artifacts that might distort the gradient signal.
Gradient Averaging: By averaging the gradients over multiple noisy samples, we get a more stable and reliable estimate of the gradients. This reduces the sensitivity of the gradient to small changes in the input image, helping to focus on the most important features and reducing the impact of random noise.
Standard Deviation σ: The standard deviation σ\sigmaσ controls the amount of noise added to the input image. A higher σ\sigmaσ results in more noise, making the gradients smoother but potentially less sensitive to fine details. A lower σ makes the gradients more precise but might not smooth out noise as effectively.

7.SYSTEM DESIGN:

This system diagram illustrates the end-to-end process for developing a deep learning model. It begins with collecting raw data, which is then preprocessed to ensure quality and consistency. Once the data is ready, a deep learning model is designed, where choices about the model's structure and functions are made. The next step is training the model on the data, fine-tuning parameters, and validating its performance to make sure it learns effectively. After training, the model is ready to make

predictions or classify new inputs. To make the model's decisions more understandable, explainability techniques are used to visualize and interpret its outputs. Finally, the model's performance is evaluated using metrics like accuracy and precision to determine how well it performs in real-world scenarios. This cycle helps refine the model and ensure it meets the desired goals.

Fig: 7.1 system design

8.EVALUATION METRICS COMPARISON

8.1 Metrics Table

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
EfficientNetB0 (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

Table : 10.1 : Metrics table

8.2 Analysis

8.2.1 Xception Model:
With its high precision and recall, Xception achieves strong overall accuracy, making it ideal for clinical scenarios that demand reliable, balanced predictions.
8.2.2 EfficientNet:
The 128x128 version is computationally lighter, offering good malignant recall. However, its performance on benign lesions is slightly lower, which may result in missed benign cases.
8.2.3 InceptionV3: This model provides reliable performance across metrics, making it versatile but not necessarily the top choice for highly specific classification needs.

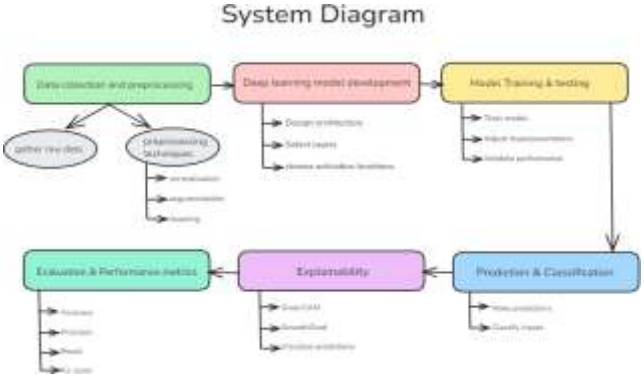
8.3 Importance of Evaluation Metrics in Medical AI

In a medical context, precision, recall, and F1-score are essential metrics that define the reliability and safety of AI systems:

8.3.1 Precision: Indicates the proportion of positive identifications (e.g., malignant cases) that were actually correct. High precision reduces false positives, which can prevent unnecessary patient anxiety and further testing.

8**.3.2 Recall:** Measures how well the model identifies true positive cases (e.g., malignant lesions correctly detected). High recall is vital to minimize false negatives, which could result in missed diagnoses.

8.3.3 F1-Score: Balances precision and recall, providing a single metric for understanding overall model performance. This is especially useful in clinical settings



where both false positives and false negatives have serious implications.

9.RESULTS AND DISCUSSION

9.1 Dataset Overview

In this project, we used a subset of the ISIC (International Skin Imaging Collaboration) Archive, a widely recognized dataset in the field of dermatology. The ISIC Archive is known for its extensive collection of dermoscopic images, aimed at improving skin cancer diagnosis using AI. For our study, we focused specifically on images labeled as either malignant (cancerous) or benign (non-cancerous). This binary classification is crucial for distinguishing potentially dangerous skin lesions from harmless ones, which is a key step in early skin cancer detection.

9.2 Focus on Binary Classification

The main goal of our project was to develop an AI model that can accurately classify skin lesions into two categories: malignant and benign. This binary classification task is particularly important in clinical settings because early and accurate identification of malignant lesions can lead to timely treatment and better patient outcomes. Benign lesions, on the other hand, are non-cancerous and usually do not require aggressive treatment. Misclassifying a malignant lesion as benign could delay necessary medical intervention, making this task a critical application of AI in healthcare.

9.3 Data Preparation

Given that our dataset only contains images with labels (malignant or benign), our preprocessing focused on standardizing these images for training the deep learning models.

Key steps included:

Image Resizing: Since different deep learning models require specific input sizes, we resized the images accordingly. For example: EfficientNet used input sizes of

128x128 pixels. EfficientNetB0 used 224x224 pixels. InceptionV3 and Xception used 299x299 pixels. This resizing ensures that all input images are compatible with the model architecture, allowing for consistent feature extraction.

Data Augmentation: To increase the variety of training data and improve the model's ability to generalize, we applied augmentation techniques such as rotation,

flipping, and zooming. These transformations help the model learn to recognize skin lesions from different perspectives and under various conditions, making it more robust when tested on new, unseen images.

9.4 Handling Class Imbalance

One of the challenges we encountered was the imbalance in our dataset, as benign cases were more prevalent than malignant ones. In medical datasets, this is common because most skin lesions are non-cancerous. However, this imbalance can lead the model to become biased, favoring benign predictions over malignant ones, which could reduce the sensitivity for detecting actual cancer cases. To address this, we implemented techniques like:

Oversampling: We artificially increased the number of malignant cases by duplicating existing malignant images, helping the model learn better from this minority class.

Class Weight Adjustment: During training, we adjusted the class weights to place more emphasis on malignant cases. This penalizes the model more when it makes errors in classifying malignant lesions, thereby improving its sensitivity to these critical cases.

10.CONCLUSION

The ultimate aim of this project was to create an AI-powered diagnostic tool that could assist dermatologists in quickly and accurately identifying malignant skin lesions. By providing a reliable second opinion, our model can help reduce the time taken for diagnosis and minimize the risk of missing a potentially cancerous lesion. This system can be especially beneficial in busy clinical environments where quick decision-making is essential. By focusing on this binary classification task, our project aims to contribute towards making AI tools an integral part of dermatological practice, providing enhanced diagnostic support and helping to improve patient outcomes through early detection.

REFERENCES

[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.

[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] 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.

[9] 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.