Empirical analysis of CNN & Transfer Learning Models with focus on Explainable AI techniques for interpretable Skin Cancer Classification

Ruchi Parmar - 501034872

Abstract

This report presents an empirical analysis of deep learning and transfer learning models for skin cancer classification, with a focus on the applicability of Explainable AI (XAI) techniques to improve interpretability. The study compares traditional CNN models with advanced transfer learning models such as EfficientNet_B0, EfficientNet_v2_B0, ResNet50, VGG16, and Vision Transformers, all pre-trained on extensive image datasets. It highlighted the challenges posed by data similarity in applying XAI methods effectively, as evidenced by overlapping class data points. The Vision Transformer model exhibited superior performance and enhanced explainability using attention mapping, achieving nearly 89% accuracy. The findings emphasize the need for continued enhancement of XAI methods to better support medical professionals in diagnosing skin cancer.

Keywords: Skin Cancer Classification, Deep Learning, Transfer Learning, Explainable AI, Vision Transformers, Model Interpretability.

1. Introduction

Skin cancer is a common type of cancer that usually forms on skin exposed to the sun, but it can also develop on areas not usually exposed to sunlight[1]. There are two main types: melanoma and non-melanoma. According to Skin Cancer Foundation Statistics skin cancer is the most diagnosed cancer, with one in every three cancers diagnosed is a skin cancer. A 10% decrease in ozone levels could lead to additional 300,000 non-melanoma and 4,500 melanoma skin cancer cases[16]. Early detection of skin cancer is crucial, as the survival rate is nearly 97% when caught early.

However, diagnosing skin cancer accurately can be challenging, time consuming and depends on the skill of the doctor. Even the best dermatologists have less than 80% accuracy in diagnosing skin cancer correctly[10]. Detecting skin cancer early is very tricky since its not always easy to see with naked eye. Not all dermatologists have same level of experience, and hence can't spot the early signs. Computer aided systems or models can really help dermatologists

to make diagnosis in early stages. To make the system efficient and easier for everyone, there is real need for precise automatic classification methods. It can benefit not only just doctors but also help patients and healthcare system[17].

Also, skin cancer diagnosis from images can be tricky because different types of skin lesions have lots of similarity between them, which makes it easy to misclassification[17]. Traditional Machine Learning algorithms rely on handpicked features and usually it only performs better on images similar to the data they are trained on. On the other hand Deep Learning methods like CNN are good with learning features from images and have become more accurate over time[12]. Issue with these models is, while working with large data or high-resolution images with large number of pixels, it results in high computational cost and time consumption. Transfer Learning methods takes this process a step further. It is becoming a popular approach while working with large images and text data, as it uses models that are trained on large amount on data(pre-trained models) and reuses it for training other tasks - like training set of data for skin cancer classification. This pre-trained models are trained on rich datasets and hence we can use its learnings in our model and save computational cost, time and gain more accuracy[9].

Although, these Deep Learning and Transfer Learning methods gives very accurate results, they are like black-box. Due to lake of transparency, their decision process is not easily explainable to humans[11]. In the healthcare systems mainly, it is very important to know the reasoning behind the decision to trust and rely on computer aided systems. Explainable AI (XAI) is an ongoing research area that is a remedy to this issue. It is like a peek into the thought process of these models. There are various XAI methods like GRAD-CAM, LIME, Attention mapping, Prototype based etc. which directs us to the part of images which were used to make decision or important behind decision[5] as shown example in figure 1. This helps humans understand reasoning before taking final decisions.

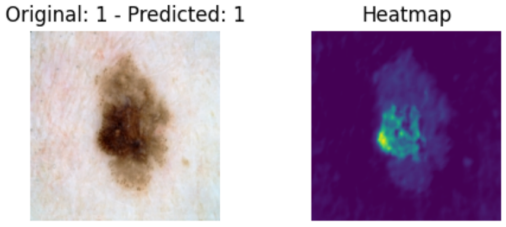


Figure 1: Interpretability of model's decision using GRAD-CAM

In this project, I set up a deep learning model using Convolutional Neural Networks (CNNs) to start with. This served as a baseline to compare other advanced models. Next I implemented and compared different transfer learning models like EfficientNetB0, EfficientNet_v2_B0, VGG, ResNet50, and Vision Transformers, to see which one could perform better while maintaining less False positives and False negatives. On top of that, I explored how to make these models' decisions transparent by using Explainable AI (XAI) methods such as GRAD-CAM, LIME, and Attention mapping. These methods help to understand why the models make their choices by highlighting the important parts of the images. I also analyzed similarity into the dataset by projecting images to lower dimension and how it affects XAI results.

The rest of the report is organized as follows. Section 2 summarizes the Literature review on recent studies on deep learning for skin cancer classification and XAI methods. Section 3 provides a brief overview the model architectures and XAI methods. Section 4 gives idea about experimental setup of the implementation. In Section 5, the results of model performances and illustrations that show the interpretability of the model predictions are provided. The paper concludes with a discussion on key findings, study limitations, and future research directions in Section 6.

2. Literature Review

Research of utilizing Deep Learning techniques for cancer classification has always been interesting area of research and it has made significant advances over the years. Exploring how these models make decisions has become a vibrant field of research, with numerous methods and algorithms emerging recently. In this section, I have analyzed existing studies performed in relative field.

The paper "Skin Cancer Classification With Deep Learning: A Systematic Review" by Wu et al. offers an extensive overview of deep learning applications in skin cancer classification, highlighting key challenges such as data imbalance, domain adaptation, and model robustness. It reviews various convolutional neural network (CNN) models that have shown promising results, often matching the diagnostic ability of dermatologists. The authors discuss frontier challenges in the field, proposing future directions focused on enhancing model efficiency and adaptability through structured, lightweight, and multimodal approaches. The review emphasizes the need for innovative techniques like pruning and knowledge distillation to improve clinical applicability and address the constraints of real-world clinical settings[17].

The paper "Skin Cancer Detection Using Deep Learning—A Review" by Naqvi et al. also presents an extensive review of deep learning methodologies applied to skin cancer classification. The review compares the performance of various deep learning models like CNN, VGG, ResNet etc. in accurately diagnosing skin cancers through image segmentation and classification. The paper emphasizes the potential of deep learning to surpass traditional visual inspections by dermatologists, which typically have lower accuracy rates. It also highlights the significance of computer-aided detection systems in improving diagnostic accuracy, which is crucial for early and effective treatment of skin cancers. Overall, this review synthesizes recent advancements in the field and suggests directions for future research to overcome existing challenges such as data imbalance and the need for more robust models[12].

The paper "Analysis of Skin Lesion Images with Deep Learning" by Josef Steppan and Sten Hanke evaluates advanced deep learning techniques for classifying skin lesion images, using models adapted from pre-trained ImageNet datasets for the ISIC-2019 Challenge. They enhance model performance and address class imbalances through real-time data augmentation and innovative balancing techniques. Their findings demonstrate substantial improvements in predictive accuracy for skin lesion classification, contributing to the field by making their

refined models. This study highlights the effectiveness of model fine-tuning and augmentation in improving the diagnostic capabilities of deep learning systems in dermatology[14].

Study titles "Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images " develops a model-driven architecture in the cloud for classifying dermal cell images and detecting skin cancer. It highlights the use of deep learning algorithms to improve accuracy and speed in diagnosis, achieving a high area under the curve (AUC) of 99.77%. This approach simplifies the creation of deep learning models, making it accessible for researchers without programming expertise, which is critical for rapid deployment in clinical settings[10].

The paper titled "Explainable Artificial Intelligence in Skin Cancer Recognition: A Systematic Review" by Hauser et al. provides a comprehensive analysis of the application of Explainable Artificial Intelligence (XAI) in the development of deep neural networks (DNNs) for skin cancer detection. The review identifies that while XAI is frequently integrated during the model development phase, there is a significant gap in systematic evaluations of its effectiveness, particularly in clinical settings. Most studies reviewed apply existing XAI techniques to assess the decision-making processes of their models, with a few proposing new or improved XAI methods. Despite the prevalent use of XAI, only three studies rigorously evaluate the impact of XAI on the performance and confidence of clinicians using these systems, highlighting a critical area for future research. The paper concludes that more robust studies are needed to truly understand the utility and impact of XAI in enhancing diagnostic accuracy and clinician trust in AI-supported dermatological assessments[5].

"Explainable Deep Learning Methods in Medical Image Classification: A Survey" by CRISTIANO PATRÍCIO et al. focuses on the advancements in explainable artificial intelligence (XAI), particularly in medical imaging. It details various XAI methods that make the outputs of deep learning models more interpretable for clinicians. The paper underscores the importance of explainability in medical diagnostics, where understanding the decision-making process of AI models can lead to greater trust and integration into clinical practice[13].

"Using ProtoPNet for Interpretable Alzheimer's Disease Classification" Mohammadjafari et al. uses the ProtoPNet architecture combined with other pretrained models to make Alzheimer's disease classifications more interpretable on MRI scans. It discusses the importance of interpretability in clinical applications, noting that while ProtoPNet slightly reduces model performance, it significantly increases trust and understanding in model predictions by showing which features lead to its conclusions. This concept is relevant for enhancing the transparency of skin cancer detection model[11]

3. Methodology

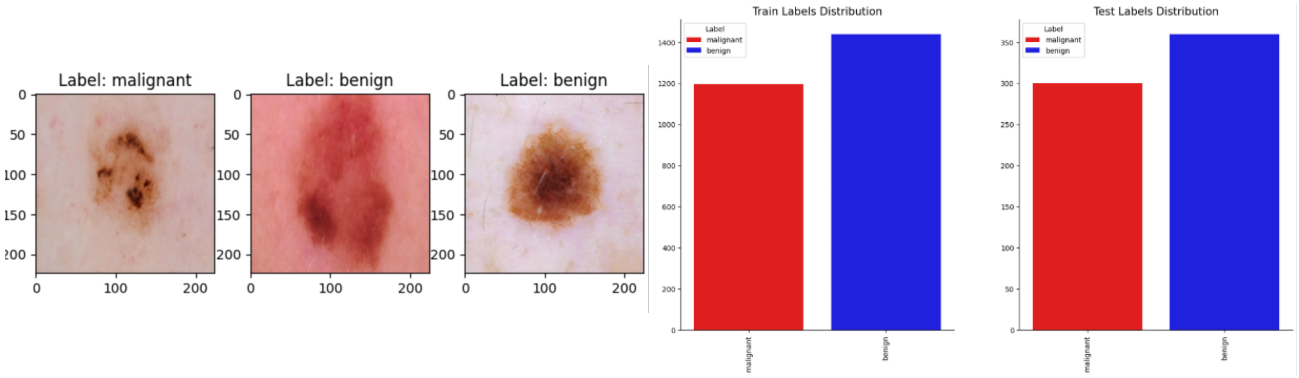
In this section information on dataset used, pre-processing steps is provided along with deep learning and transfer learning model architectures. Further details on the various XAI methods that are available and used is provided.

3.1 Dataset

For this project, I have used Kaggle version[7] of skin cancer classification data from International Skin Imaging Collaboration(ISIC). It is a collaboration between academia and industry, aims to advance the use of digital skin imaging to help lower the death rate from melanoma.

3.1.1 Exploratory Data Analysis

This dataset comprises a total of 3,297 samples, which are divided into a training set of 2,637 samples and a test set of 660 samples by default. An initial visual assessment of the dataset can be seen in below figure (a), where few sample images are displayed, providing a glimpse into the variety of skin lesions included in the study. Furthermore, as illustrated in the bar graph in below figure (b), the dataset exhibits a well-balanced distribution of data across both classes - "Benign" and "Malignant" in both the training and test sets. This balanced dataset is crucial for training unbiased and effective machine learning models for skin cancer classification.



(a) Sample images from dataset

(b) Data distribution between both classes

3.1.2 Pre-processing the data

To prepare the data for skin cancer classification using various models I have performed few pre-processing steps. Paths to image locations and their corresponding labels are organised within a data frame for streamlined access and manipulation. Training data is split into a training set and a validation set with an 80-20 ratio to ensure both robust learning and adequate model evaluation.

To fit the data in deep learning models, image data is converted into tensors using TensorFlow's tf.data.Dataset. This transformation is crucial for efficient batch processing during model training. Additionally, I implemented a data augmentation layer in the training pipeline, which introduces random flips and random zooms to the images. This augmentation not only helps in preventing the model from over-fitting but also ensures that it learns to recognize skin cancer from varied angles and scales.

All images were resized to 224 x 224 pixels with three color channels (RGB) and normalized by scaling the pixel values to a range of 0 to 1 by dividing them by 255. This normalization standardizes the input data and helps in accelerating the convergence of the model during training. Finally, the data was batched into sizes of 32 for efficient training, allowing for gradient updates that balance speed and memory usage effectively.

3.2 Deep & Transfer Learning Methods

I have utilised various open source deep learning Models with transfer learning along with baseline CNN model for this classification task. These models are - EfficientNet_B0, EfficientNet_v2_b0, ResNet50, VGG16 and Vision Transformers. These models are pretrained on large and rich image data. In this section I have summarised these models and their architecture. For the classification task I have used all these baseline models with transfer learning by adding dropout and two fully connected layers on top.

- **CNN:** Convolutional Neural Networks are foundational to deep learning for image classification tasks. A typical CNN architecture includes layers such as convolutional layers, pooling layers, and fully connected layers. These layers are designed to automatically learn spatial hierarchies of features or patterns, from low-level edges to high-level features, from images and perform tasks such as classification[2]. I have set CNN as a baseline model for understanding the basic functionalities to be expected in advanced transfer learning networks. Architecture of CNN model for this classification task can be seen in figure 3.

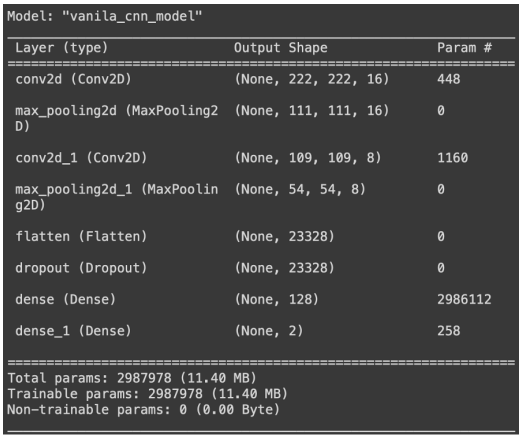


Figure 3: CNN architecture

- **EfficientNet_B0 & EfficientNet_V2_B0:** EfficientNet_B0 is part of the EfficientNet family, which scales up CNNs in a more structured way. EfficientNets use a compound coefficient to uniformly scale the depth, width, and resolution of the network. EfficientNet_B0, specifically, is optimized for accuracy and efficiency, and serves as the baseline for scaling up to other EfficientNet models (B1-B7).

  EfficientNet_V2 improves upon the original EfficientNet models by incorporating training-time optimizations and model simplifications. EfficientNet_V2_B0, the baseline model

for this series, is designed for even faster training speeds and greater efficiency. It introduces improvements such as reduced parameter count and optimized training processes[15].Figure 4 is architecture of EfficientNet_B0 architecture.

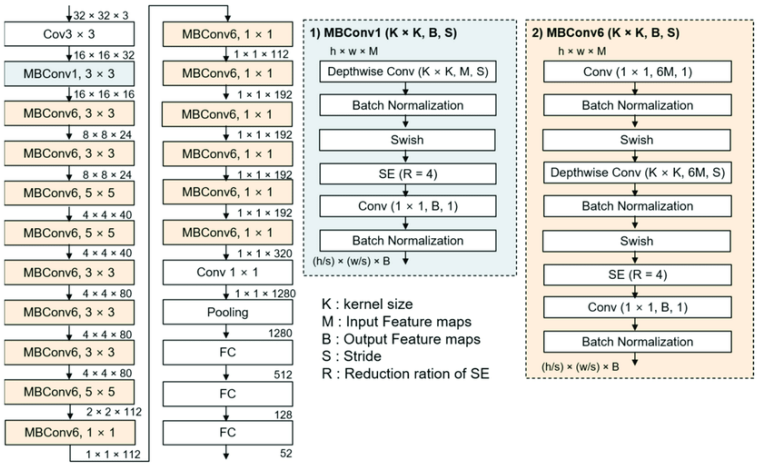


Figure 4: EfficientNet_B0 architecture

- **ResNet50:** ResNet50 is part of the Residual Networks family, which introduces residual learning to facilitate the training of much deeper networks. In ResNet50, layers learn residual functions with reference to the layer inputs, allowing it to have a much deeper architecture without the vanishing gradient problem. This model consists of 50 layers and is widely used due to its impressive performance across various datasets[6].

- **VGG16:** The VGG16 model is designed so that the number of filters doubles within each block, while the size of the feature maps halves. This structure helps the model progress from recognizing simple features early on to identifying complex shapes deeper in the network. The model consistently uses the 'ReLU' activation function in all layers except the output layer. However, a notable drawback of VGG16 is its large number of parameters—138 million—which can lead to lengthy training times depending on the dataset size and hardware capabilities[11].

- **Vision Transformers:** The Vision Transformer (ViT) model was introduced in the paper "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," at The International Conference on Learning Representations (ICLR) in 2021. ViTs are utilized for various image recognition tasks including object detection, image segmentation, action recognition, and image classification[3][4][8].

  Originating from transformer architectures in Natural Language Processing (NLP), which turn text into sequence tokens and generate text embeddings, ViTs adapt this concept to images(as seen in figure 5). An image is divided into patches similar to word tokens in NLP. These patches are processed by a transformer encoder to produce image embeddings. The transformer encoder includes three key components:

- **Layer Normalization:** Enhances computational efficiency by normalizing the patches and attention outputs.
- **Multi-head Attention:** Creates multiple attention heads for the patches, allowing the model to capture both local and global image dependencies.
- **Multi-Layer Perceptrons (MLP):** Processes the attention outputs through two dense layers using the Gaussian Error Linear Unit (GELU) as the activation function.

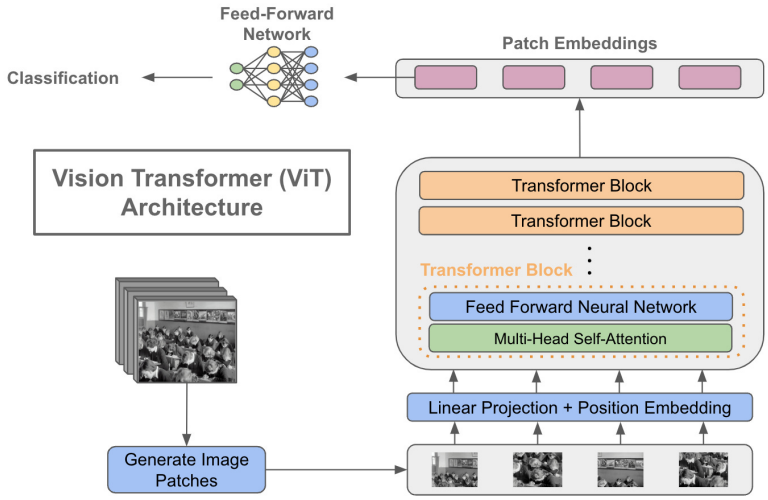


Figure 5: Basic Vision Transformer architecture

3.3 Explainable-AI Methods

This section covers brief overview on Explainable AI, various XAI methods categories based on their approach to explanation and 3 methods(Grad-CAM, LIME and Attention Mapping) that are implemented in this project.

Explainable AI (XAI) methods aim to make the decision-making processes of AI models more transparent and understandable. XAI is particularly crucial in fields like medical imaging, where understanding how decisions are made is essential for trust and implementation[13]. XAI methods vary widely, each offering different insights into how models make their decisions. These methods can broadly categorized based on their approach to explanation:

- **Visual explanations:** Visual explanations aim to make the internal workings of a model transparent by highlighting important features or regions within the input data that significantly influence the model's predictions. Techniques like **Saliency Maps, Grad-CAM, and Attention Mapping** visualize these aspects, often through heatmaps or overlay markings on the input images, making it easier to understand why a model makes certain decisions[13].
- **Textual explanations:** Textual explanations provide insights into a model's decision-making process through descriptive text. These can be **automated reports or natural language statements** that explain the reasoning behind a prediction[13].

- **Example-based:** Example-based explanations help users understand model behavior by referencing specific instances from the dataset. Methods like **Prototypes and Criticisms based reasoning** highlight representative examples and notable outliers that significantly influence the model's learning[13].

- **Concept-based:** Concept-based explanations seek to link model decisions to high-level human-understandable concepts. Techniques such as **Concept Activation Vectors (CAVs) and Testing with Concept Activation (TCAV)** analyze directions in the model's internal representation space that correlate with these concepts, helping to bridge the gap between abstract model computations and intuitive human reasoning[13].

In this project, I implemented several XAI techniques to enhance the interpretability of baseline - Vanilla CNN and Best performer - Vision Transformer models:

- **CNN + Grad-CAM:** Grad-CAM is used to create visual explanations for CNN models by highlighting the regions of the input image that are important for predictions. This is achieved by using the gradient information flowing into the final convolutional layer of the CNN. GradcamPlusPlus from tf_keras_vis.gradcam is used for this task. once GradCAM++ is received using the library, it can be plotted as heatmap and compared with original image.

- **CNN + LIME & Vision Transformer + LIME:** LIME generates local interpretable explanations for predictions made by models. It perturbs the input image and observes the effect on the output to identify regions that significantly influence the model's decision. This method is applied to both CNN and Vision Transformers in the project to provide insights into each model's decision-making process. For this task lime_image.LimeImage-Explainer is utilized.

- **Vision Transformer + Attention Mapping:** This technique utilizes the attention mechanisms inherent in Vision Transformers to highlight areas of the image that most influenced the model's prediction. This method is particularly useful for understanding and visualizing how Vision Transformers process and prioritize different parts of an image for classification tasks. Few example of attention mappings are shown in figure 6.

4. Experimental Setup

I trained all models on 80% of the training data, using the remaining 20% for validation purposes. Testing is carried out on a separate test dataset, where I assessed model performance using metrics such as accuracy, precision, recall, ROC values, and confusion matrices, with a particular focus on the ratios of false negatives (FN) and false positives (FP).

I compared the performance of a baseline CNN model against five transfer learning models. For the transfer learning models. Pre-trained models are used to extract features and added a

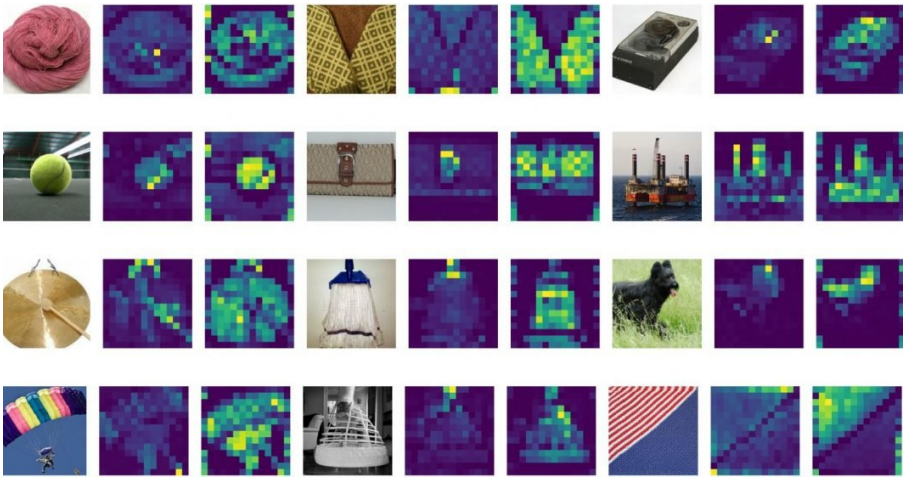


Figure 6: Attention Mapping examples

dropout layer and two dense layers(with 'relu' and 'sigmoid' activation respectively) to finalize the training process. Hyperparameter tuning was conducted across various parameters including the learning rate, dropout rate, and optimizer type and below parameters are finalised to achieve best performance(all models are trained on same parameters for better comparison) -

Table 1: Hyperparameters for the models

hyperparameter	value
learning rate	0.001
Optimizer	Adam
Dropout rate	0.2

All the models are trained for 50 epochs with an early stopping mechanism that monitored the validation loss, halting training if there were no improvements after five epochs.

For the explainable AI (XAI) analysis, five images from each class are selected randomly in the test data to apply three XAI methods previously discussed — Grad-CAM, LIME, and Attention Mapping with both the baseline CNN model and the best-performing model, the Vision Transformer. This allowed to examine how each model made its decisions and to validate the interpretability of the outputs.

5. Results

In this sections, results and findings on performances of each model is presented. It also includes interpretability of two models (CNN and Vision Transformer) using XAI methods. Further I have also discussed limitations of XAI with particular dataset.

5.1 Performance of Deep & Transfer Learning Models

Table 2 presents the performance metrics for all models, including accuracy, precision, recall, as well as the number of false negatives and false positives for each model. The values in the table are averages from multiple runs performed for each model. The baseline CNN model achieves an accuracy of nearly 83%, which is slightly better than that of dermatologists. However, this model has a high number of false positives, where non-cancerous cases are incorrectly predicted as cancerous. The EfficientNet_B0 model, using transfer learning, did not achieve high accuracy due to a large number of false negatives, indicating a bias towards classifying cases as benign. A similar issue is observed with the ResNet50 model. In contrast, the VGG16 model performs similarly to the CNN model but requires more time to train. The EfficientNet_v2_B0 model achieves about 86% accuracy but still struggles with a high number of false negatives.

Among all the models, the Vision Transformers achieved the best performance, with an accuracy of nearly 89%. It not only improved precision and recall but also maintained a balance between false positives and false negatives, effectively identifying true labels while minimizing incorrect classifications. Maintaining low number of False negatives while reducing False positives is extremely important as, patient with cancer should not be identified as non-cancerous.

Table 2: Comparison of model's performances

Metrics	CNN	EfficientNet_B0	ResNet50	VGG16	EfficientNet_v2_B0	ViT
accuracy	0.83	0.55	0.55	0.83	0.86	0.89
precision_benign	0.86	0.55	0.55	0.86	0.88	0.88
precision_malignant	0.80	0.00	1	0.81	0.83	0.89
recall_benign	0.82	1	1	0.83	0.86	0.91
recall_malignant	0.84	0.00	0.01	0.84	0.86	0.85
False Negative	49	300	296	48	41	44
False Positive	63	0	0	61	52	31

5.2 Interpreting predictions with Explainable-AI

In this section, I present the results from applying XAI methods to sample data. The GRAD-CAM method was applied to the CNN model, utilizing weights from the last convolutional layer to highlight important regions within the images. GRAD-CAM results are displayed in Figure 7, where blue indicates areas of low importance and yellow denotes the most significant regions. It is notable that for two images where the class was incorrectly predicted, the model failed to assign clear importance to specific areas due to unclear image.

For attention mapping with the CNN, it would be necessary to modify the model architecture to include an attention layer. During testing with the CNN_attention model, it is observed that while the model performance remained similar, the results from attention mapping closely

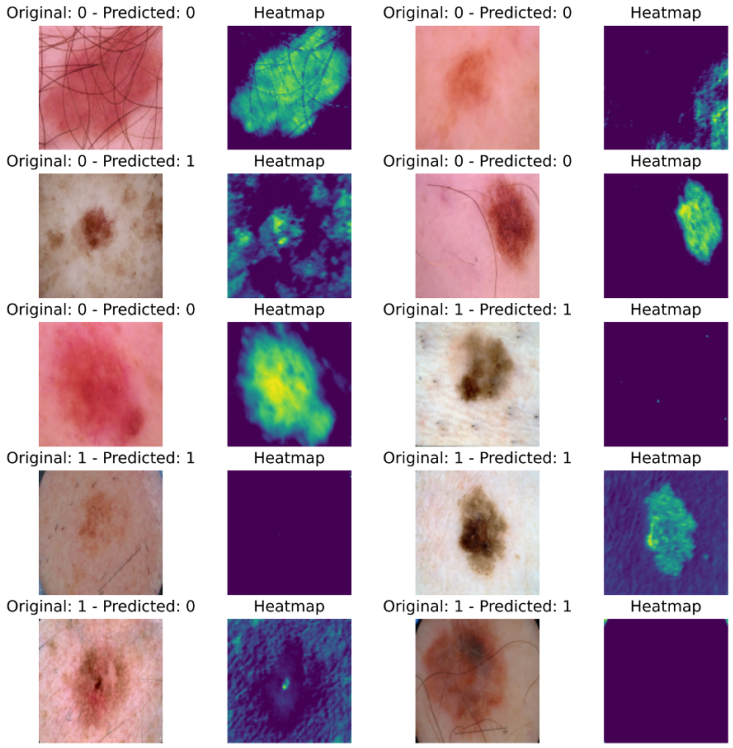


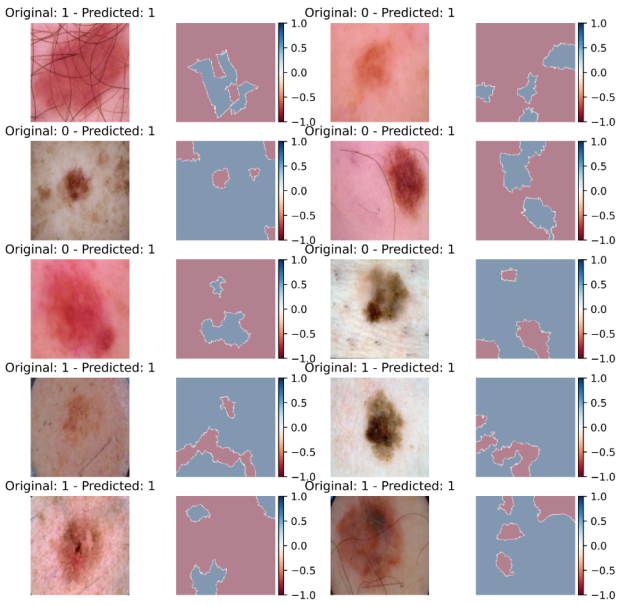
Figure 7: CNN GRAD_CAM results

resembled those from GRAD-CAM. Consequently, I decided to use only GRAD-CAM with the CNN.

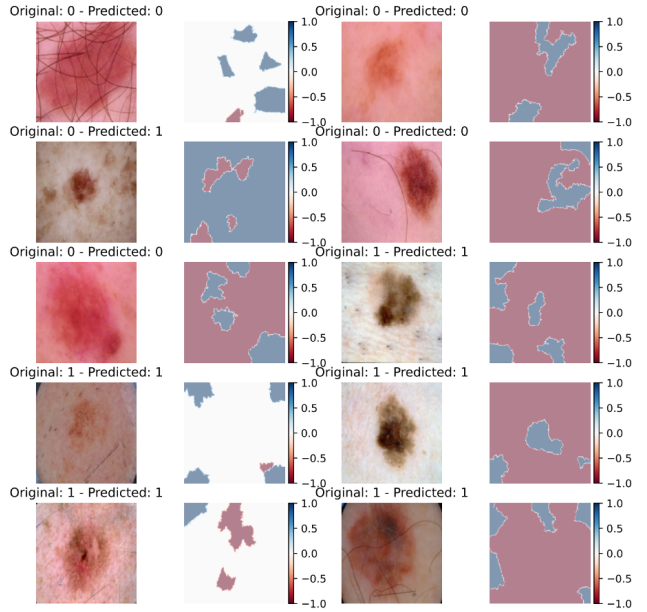
The LIME method was applied to both the CNN and Vision Transformer models using the LIME library. The outcomes are shown in Figure 8, where red indicates a negative impact and blue a positive impact on the model's decision-making. Compared to CNN, LIME provided clearer explanations for the Vision Transformer, though it was still not particularly effective for this dataset. This issue is further discussed in the next section 5.2.

Lastly, the Vision Transformer model processes image data in a manner similar to how transformers handle NLP data, focusing attention on individual pixels across 12 encoder blocks. Figures 9 and 10 illustrate how the model learns to focus its attention in each encoder block. The issue of similar attention patterns across different images is explored further in Section 5.2. It can be seen that with attention mapping we can get very clear importance mapping for each pixels of the image.

Using these XAI methods, healthcare professionals can verify whether the model is focusing on the correct regions of the images when making decisions. Specifically, the LIME method, when applied to a more suitable dataset, enables visualization of which regions positively or negatively impact the model's decision-making process. This capability is crucial for ensuring that the deep learning models are reliable and trustworthy, providing staff with actionable insights into the models' operational dynamics.

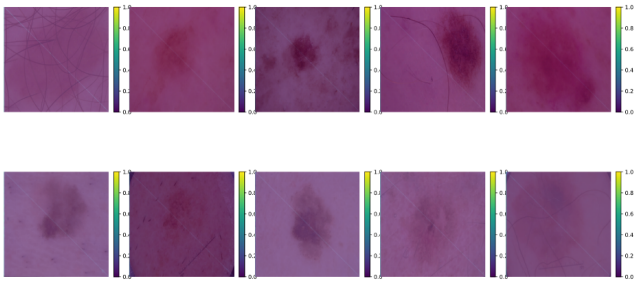


(a) LIME with CNN

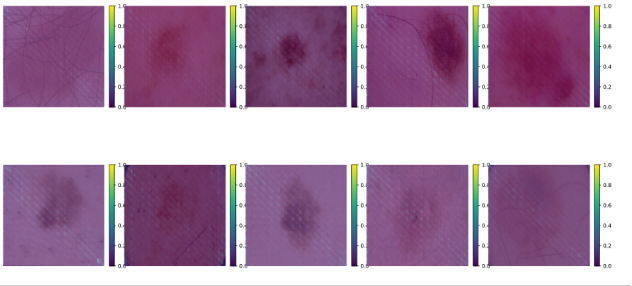


(b) LIME with ViT

Figure 8: LIME method



(a) Attentions after first encoder block



(b) Attention after five encoder blocks

Figure 9: Attentions

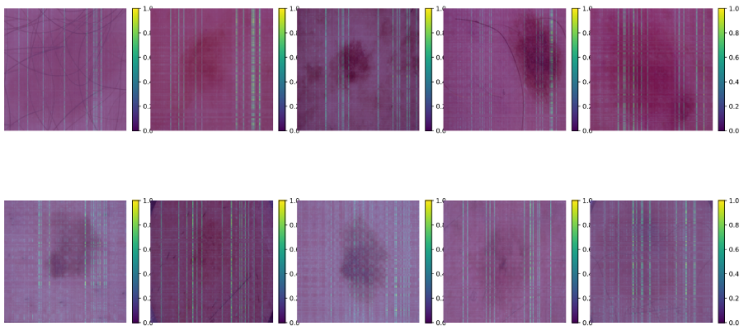


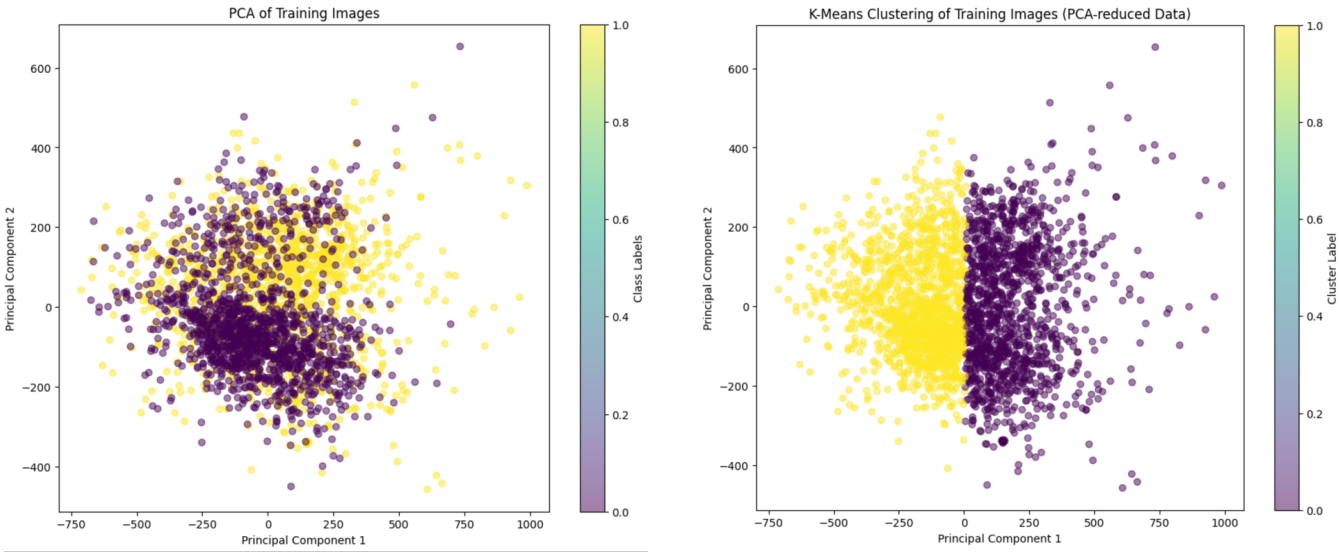
Figure 10: Attentions after final encoder blocks

5.3 Challenges faced in applying Explainable-AI

In exploration of XAI methods for this dataset, I encountered several challenges. As seen above, the LIME method struggled to assign significant weights across all images uniformly,

and the attention mechanisms in the Vision Transformer is remarkably similar patterns for different images. An attempt to implement a prototype-based method revealed another issue: all prototypes generated were very similar, with no distinguishable differences between them. Furthermore, mapping test images to these prototypes did not reveal any distinct variations, suggesting a lack of diversity in the features being captured.

These observations drew attention towards looking into the data's similarity. I applied Principal Component Analysis (PCA) to reduce the dimensionality of the image data and visualized it in two dimensions. As shown in the figure 11, the data points from both classes overlapped significantly, indicating a lack of clear separation. Subsequent clustering attempts using K-means also split the data down the middle, further indicating the absence of distinct groupings within the data.



(a) Training data projection in 2-dimension

(b) Image clustering on training data

Figure 11: Data projection and clustering

This overlap and the challenges in distinguishing between classes make interpreting model decisions with XAI particularly difficult for this dataset. While methods like GRAD-CAM successfully identified relevant regions influencing model decisions, more advanced methods like prototype-based approaches and attention mapping did not provide the expected clarity or transparency in illustrating how decisions were made.

6. Conclusion

This study explored the effectiveness of various deep learning and transfer learning models for skin cancer classification, emphasizing the role of Explainable AI (XAI) methods in enhancing model transparency. While Vision Transformers demonstrated significant predictive capabilities maintaining other performance metrics like False Negatives and False positives, other models

like CNN, VGG16, EfficientNet_v2_B0 predicts high number of False positives. On the other hand ResNet50 and EfficientNet_B0 doesn't perform better with this dataset. With Vision Transformer, 8 - 9% higher accuracy can be achieved than current accuracy of dermatologists with naked eye.

The study revealed challenges faced in interpreting model decisions due to data overlap and similarity. The application of XAI methods such as GRAD-CAM, LIME, and Attention Mapping provided insights into the decision-making processes. Although, GRAD_CAM can confirm the region of image used to take decision for doctors, advanced methods showed limitations when applied to datasets with high similarity. The best performance was observed in the Vision Transformer model, which balanced accuracy and interpretability better than others.

These Transfer Learning models along with XAI methods can further be tested with larger dataset to observe their performance on dataset without similarity between classes. Also, similar study can be conducted in other healthcare areas to help healthcare system trust computer aided models.

References

[1] Mayo Clinic. *Skin Care*. URL: https://www.mayoclinic.org/diseases-conditions/skin-cancer/.

[2] Datacamp. *An Introduction to Convolutional Neural Networks (CNNs)*. URL: https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns.

[3] Alexey Dosovitskiy et al. 'AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE'. In: (). URL: https://arxiv.org/pdf/2010.11929.pdf.

[4] Deep (Learning) Focus. *AWhat are Vision Transformers?* URL: https://cameronrwolfe.substack.com/p/vision-transformers.

[5] Katja Hauser et al. 'Explainable artificial intelligence in skin cancer recognition: A systematic review'. In: (). URL: https://www.sciencedirect.com/science/article/pii/S095980492200123X.

[6] Kaiming He et al. 'Deep Residual Learning for Image Recognition'. In: (). URL: https://arxiv.org/pdf/1512.03385.pdf.

[7] Kaggle. *Skin Cancer: Malignant vs. Benign*. URL: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign.

[8] Kaggle. *Transfer Learning | Skin Cancer Classification*. URL: https://www.kaggle.com/code/matthewjansen/transfer-learning-skin-cancer-classification.

[9] Machine Learning Mastery. *A Gentle Introduction to Transfer Learning for Deep Learning*. URL: https://machinelearningmastery.com/transfer-learning-for-deep-learning//.

[10] Sulaiman Al Riyaee Mohammad Ali Kadampur. *Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images*. URL: https://www.sciencedirect.com/science/article/pii/S2352914819302047#bib26.

[11] Sanaz Mohammadjafari et al. 'Using ProtoPNet for Interpretable Alzheimer's Disease Classification'. In: (). URL: https://assets.pubpub.org/gzl17h3e/21624570427985.pdf.

[12] Maryam Naqvi et al. 'Skin Cancer Detection Using Deep Learning—A Review'. In: (). URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10252190/.

[13] CRISTIANO PATRÍCIO et al. 'Explainable Deep Learning Methods in Medical Image Classification: A Survey'. In: (). URL: https://arxiv.org/pdf/2205.04766.pdf.

[14] Josef Steppan & Sten Hanke. 'Analysis of skin lesion images with deep learning'. In: (). URL: https://arxiv.org/pdf/2101.03814v1.pdf.

[15] Mingxing Tan & Quoc V Le. 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks'. In: (). URL: https://arxiv.org/pdf/1905.11946.pdf.

[16] WHO. *UV radiation and skin cancer*. URL: https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer/.

[17] Yinhao Wu et al. 'Skin Cancer Classification With Deep Learning: A Systematic Review'. In: (). URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9327733/.