

## **Melanoma detection through AI models**

**Aditya N. Jas**

*5/25/24 (date completed)*



## **Abstract**

Skin cancer is one of the most prominent forms of cancer, and early detection can significantly improve the recovery process for those affected. Through this project, I aimed to understand the ability to detect cancer from images of skin lesions using artificial intelligence (AI) models. I attempted to build software which would allow users to predict the possibility of cancerous lesions by processing images through these models. The approach involved the use of different AI algorithms which were trained on a dataset of thousands of individual images, and identifying which models yielded the best results. Just from graphing the data, it was shown that the malignant images were darker on average, with lower average color values for each pixel. The test accuracy of most models ranged between 80 - 90%, proving that the AI models were able to correctly predict malignant v/s non-malignant 80 - 90% of the time. Through my research, I discovered that AI models have the ability to correctly predict melanoma by processing images using a variety of algorithms through which the data is trained. Additionally, I learned of the numerous physical attributes of an image which could help people predict the malignancy of skin lesions before using an AI model for prediction.

## **1. Introduction**

In my research, I decided to explore whether or not skin cancer detection through trained AI models was a realistic possibility, contributing to early skin cancer detection. Due to the nature of skin cancer, and the necessity of its early detection, a form of AI software aiding melanoma detection could significantly limit the effects of cancerous cells across individuals. The use of AI in diagnosing several forms of cancer is often used by medical professionals, however the images often require high standards of technology which inhibit an efficient diagnosis. Currently, AI is used for diagnosing patterns in brain waves, brain activity, medical scans, and has other implementations which contribute to the diagnostic abilities of medical professionals. This problem would fall under the classification category, as the goal of the project is to correctly identify malignancy in images of the skin, with the intention of generating a proper diagnosis. The input data would be in the form of images, with the diagnosis in the form of a text label.

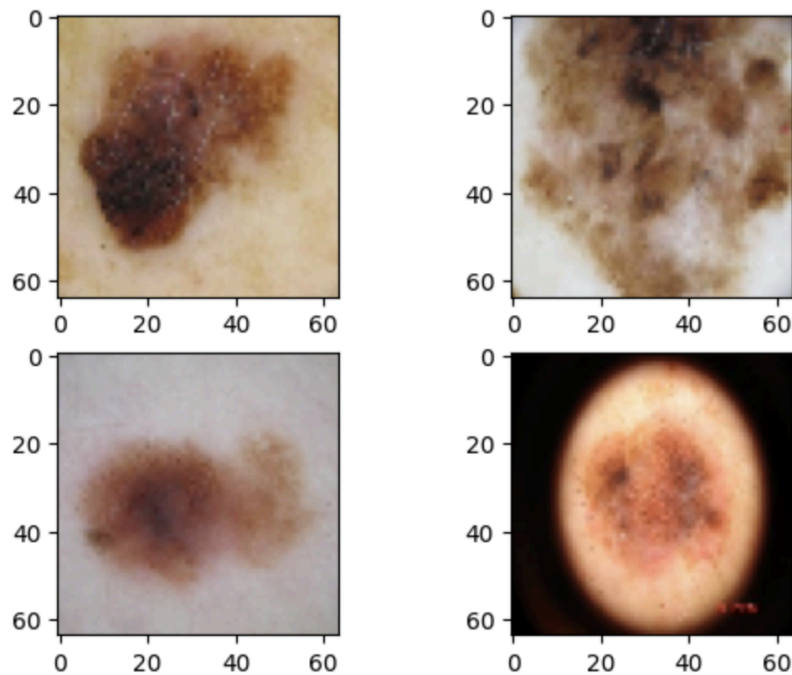
## **2. Background**

An approach from the Institute of Electrical and Electronic Engineers describes a method of classifying images of skin lesions through the use of processing filters, allowing for a clearer image of the lesion, and for the algorithm to easily identify melanoma in the images. In addition to the filters, the article describes the use of k-means clustering in order to provide a more accurate diagnosis. While this step may contribute to higher accuracy scores, the implementation of such processes ensures that the images of the lesions are of optimal quality, and can be processed as such. This differs from my approach, as one of the primary goals of this project was to create an algorithm which used images with minimal preprocessing. By focusing on

developing an algorithm that performs well even with raw, unprocessed images, the goal is to create a more robust and versatile solution that can be applied in real-world scenarios where high-quality, preprocessed images may not always be available. This approach aims to enhance the accessibility and practicality of the algorithm, making it useful in a wider range of applications and environments.

### 3. Dataset

The dataset used for this project was sourced from Kaggle and created by Data Scientist Muhammad Hasnain Javid. It consists of JPEG files with a resolution of 300x300 pixels. The dataset had 9605 images (5000 benign, 4605 malignant). It was a balanced dataset. Before processing the images, I created several visualizations to depict the general differences between the malignant and benign images. Through these visualizations, I was able to determine the average RGB pixel values for the malignant and benign images. Additionally, I noticed that the malignant images on average had lower average RGB values, and that they were generally darker. Before processing the images, I had to downsize them from 300x300 to 64x64 in order for the AI algorithms to run more efficiently. This change didn't create any significant change in the accuracy of the training models.



**Figure 1:** Examples of some images from the dataset

#### 4. Methodology / Models

To train and evaluate the models, I split the dataset into training and testing sets with an 80-20 split. The training set was used to train the models, while the testing set was reserved for evaluating their performance.

Since it was a classification problem, I used the four traditional machine learning algorithms for Classification: Logistic Regression, Random Forest, AdaBoost, and Support Vector Classifier (SVC), all implemented with the `scikit-learn` library.

Logistic Regression works by predicting the probability of a binary outcome using the logistic function. I used the training dataset and used the Logistic Regression function of `scikit-learn`.

AdaBoost combines multiple weak classifiers to create a strong classifier by adjusting the weights of incorrectly classified instances in subsequent iterations.

Support Vector Classifier (SVC) works by finding the optimal hyperplane that separates data points of different classes with the maximum margin.

Random Forest operates by constructing multiple decision trees during training and outputting the class that is the mode of the classes of the individual trees.

In my testing, Random Forest provided the best average scores. As a result, I further tuned it by changing the hyper-parameters.

Hyperparameter tuning was conducted using `GridSearchCV` to find the best parameters for the Random Forest model.

While these traditional algorithms yielded good results, I wanted to see if these could be improved further using CNNs.

Convolutional Neural Networks (CNNs) are designed to process structured grid data such as images. They use convolutional layers with filters to detect features, followed by pooling layers to reduce dimensionality, and fully connected layers for classification.

The models were evaluated using metrics such as accuracy, recall, precision, and the ROC-AUC score. Confusion matrices were generated to visualize performance in terms of true positives, false positives, true negatives, and false negatives. Additionally, ROC curves were plotted to better understand the trade-off between sensitivity and specificity.

## 5. Results and Discussion

The performance of the models was evaluated using accuracy, precision, recall, and F1-score. The models developed include Logistic Regression, AdaBoost, Support Vector Classifier (SVC), Random Forest, and CNN.

### ***Model Learning Procedure***

The dataset was split into training and testing sets with an 80-20 ratio. Features were extracted from the images, including mean RGB values, brightness differences, and estimated lesion sizes. These features were input into the models, which were then trained and evaluated.

1. Logistic Regression: Implemented using [scikit-learn](#), achieved an accuracy of 85.32%.
2. AdaBoost: Enhanced using grid search, resulting in a final accuracy of 88.65%.
3. Support Vector Classifier (SVC): Achieved an accuracy of 83.86% with a linear kernel.
4. Random Forest: Without tuning, it achieved 90.62% accuracy. After tuning, it slightly increased to 90.78%.
5. Convolutional Neural Network (CNN): Achieved 91% accuracy, demonstrating robust performance on image data.

### ***Hyperparameter Tuning***

Hyperparameter tuning was conducted to optimize the performance of the AdaBoost and Random Forest models. For AdaBoost, the grid search focused on the number of estimators and learning rate, resulting in the following optimal hyperparameters:

- [n\\_estimators](#): 100
- [learning\\_rate](#): 1.0

For the Random Forest model, tuning was performed on the number of estimators and maximum depth, yielding these optimal hyperparameters:

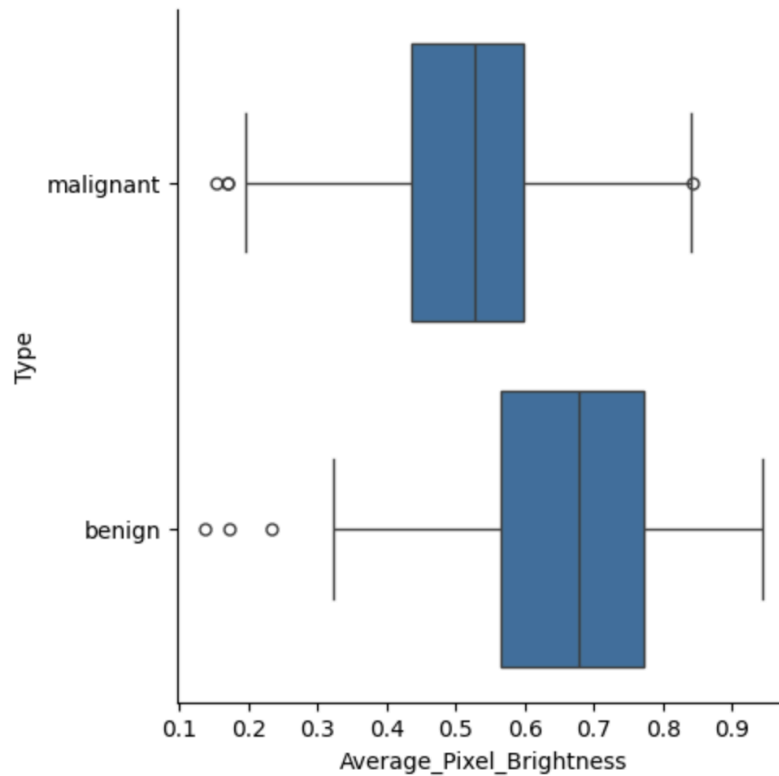
- [n\\_estimators](#): 150
- [max\\_depth](#): 9

Below are the summarized performance metrics for each model:

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression Model	85.32%	86%	85%	85%
Random Forest Model	90.62%	91%	91%	91%
ADABOOST Model with Hyperparameter Tuning	88.65%	89%	89%	89%
Support Vector Model	83.86%	84%	84%	84%
Random Forest with Hyperparameter Tuning	90.78%	91%	91%	91%

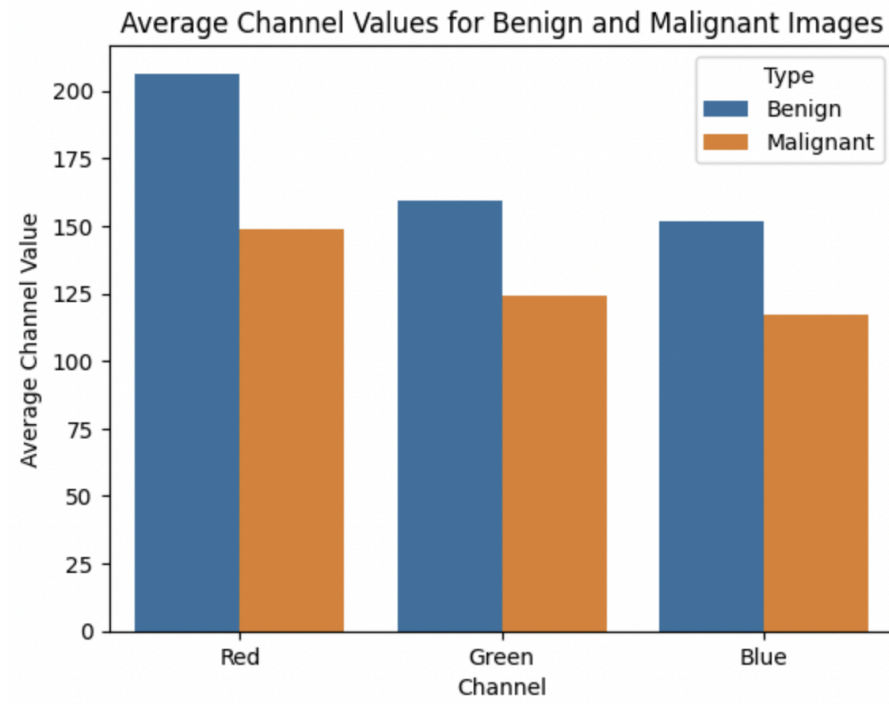
**Table 1:** Performance metrics for each model

## Visualizations and Figures



**Figure 2:** Box plot for pixel brightness for malignant and benign images

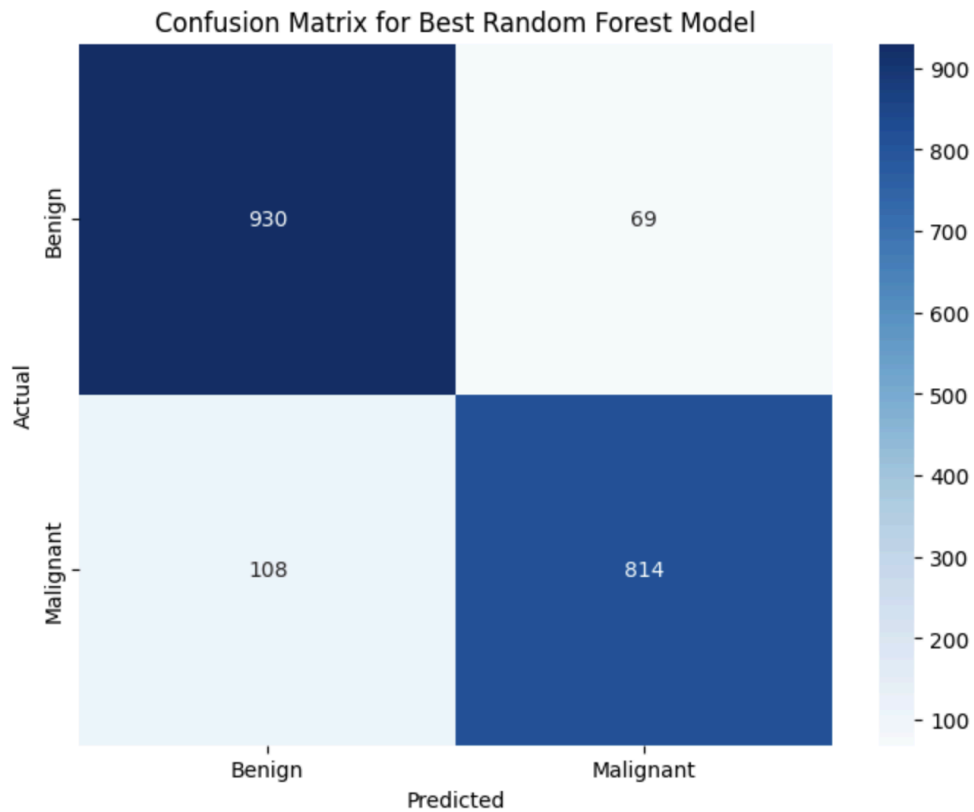
- Significance: By plotting the pixel values for malignant and benign images, we can see that on average, the malignant images are darker compared to benign ones.



**Figure 3:** RGB values of pixels for malignant and benign images

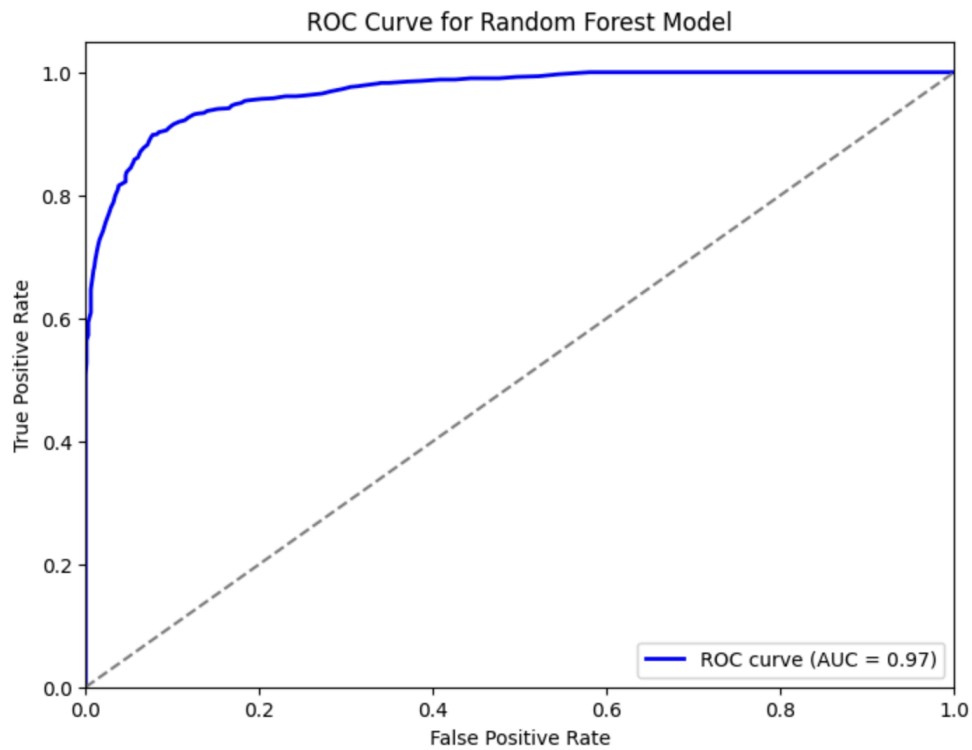
- Significance: By plotting the RGB values of pixels for malignant and benign images, we can observe that on average, the malignant images have lower values and are darker compared to benign ones.





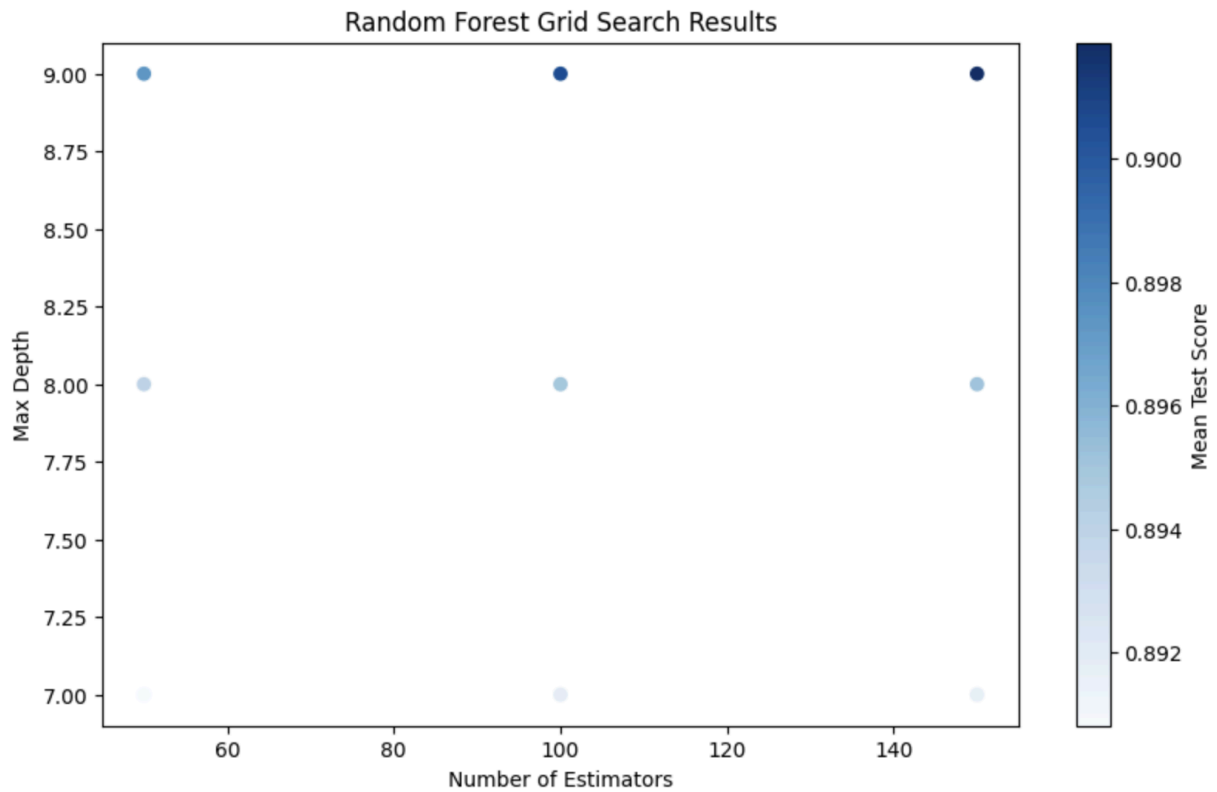
**Figure 4:** Confusion Matrix for the Hyperparameter tuned Random Forest Model

- Significance: The confusion matrix highlights the true positives, false positives, true negatives, and false negatives, indicating the model's performance in distinguishing between benign and malignant images.



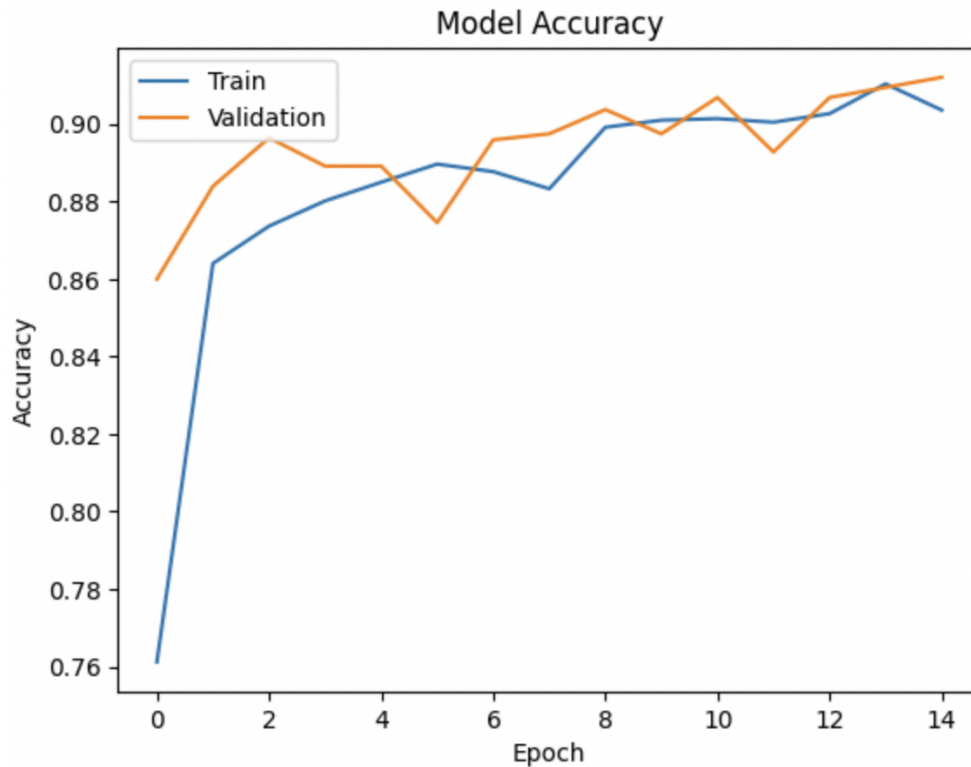
**Figure 5:** ROC Curve for the Hyperparameter tuned Random Forest Model

- Significance: The ROC curve illustrates the trade-off between sensitivity and specificity. The area under the curve (AUC) provides a single metric to evaluate the model's performance.



**Figure 6:** Scatter Plot of Grid Search Results for Hyperparameter tuned Random Forest Model

- Significance: This scatter plot shows the hyperparameters' impact on the model's mean test score, guiding the selection of optimal parameters.



**Figure 6:** Accuracy for the CNN Model

- Significance: This graph depicts the Convolutional Neural Network model's accuracy increasing exponentially as the number of epochs increases.

### ***Errors and Model Performance Discussion***

Despite the high accuracy of the Random Forest and CNN models, some errors were noted:

1. False Positives: Instances where benign images were misclassified as malignant, possibly due to noise or features in benign images that resemble malignant characteristics.
2. False Negatives: Malignant images misclassified as benign, which is critical to minimize. These errors could arise from subtle features in malignant lesions not being captured effectively by the models.

The SVC and AdaBoost models showed lower performance compared to the Random Forest and CNN models. This might be due to the high dimensionality and complexity of image data, where traditional models might struggle without extensive feature engineering.

## Conclusion

The research demonstrated that both traditional machine learning models and CNNs can effectively classify medical images, with CNNs and Random Forests showing superior performance. Hyperparameter tuning further optimized the models, enhancing their performance metrics. Visualizations such as confusion matrices and ROC curves provided insights into the models' performance, highlighting areas for improvement. Future work could focus on refining feature extraction methods and exploring advanced deep learning architectures to further improve classification accuracy.

## 6. Conclusion

By integrating traditional machine learning models and a CNN, the medical images were effectively classified into benign and malignant categories. The models were rigorously trained and evaluated, with hyperparameter tuning to optimize their performance. This comprehensive approach, leveraging both traditional algorithms and deep learning, provided robust solutions to the classification problem.

## 7. Acknowledgments

I would like to acknowledge my mentor, Alaisha Alexander, for guiding me on this project.

## 8. References

- [1] Das, Kinnor, et al. "Machine Learning and Its Application in Skin Cancer." *MDPI, Multidisciplinary Digital Publishing Institute*, 20 Dec. 2021, [www.mdpi.com/1660-4601/18/24/13409](http://www.mdpi.com/1660-4601/18/24/13409).
- [2] "Early Detection." *The Skin Cancer Foundation*, 27 Feb. 2023, [www.skincancer.org/early-detection/](http://www.skincancer.org/early-detection/).
- [3] Khan, Muhammad, et al. *IEEE Xplore*, [ieeexplore.ieee.org/abstract/document/8756036](http://ieeexplore.ieee.org/abstract/document/8756036). Accessed 22 June 2024.
- [4] Mehta, Palak, et al. "Diagnosis of Skin Cancer Using Machine Learning Techniques." *Microprocessors and Microsystems*, Elsevier, 29 Dec. 2020, [www.sciencedirect.com/science/article/abs/pii/S0141933120308723](http://www.sciencedirect.com/science/article/abs/pii/S0141933120308723).
- [5] Sharma, Pulkit. "Introduction to K-Means Clustering Algorithm." *Analytics Vidhya*, 4 June 2024, [www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/](http://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/).
- [6] "Skin Cancer." *Cleveland Clinic*, [my.clevelandclinic.org/health/diseases/15818-skin-cancer](http://my.clevelandclinic.org/health/diseases/15818-skin-cancer). Accessed 19 June 2024.

[7] Warren, Homa. “Researchers Create AI Model to Understand How Brain Activity Relates to Illness.” *Baylor College of Medicine*, [www.bcm.edu/news/researchers-create-ai-model-to-understand-how-brain-activity-relates-to-illness#:~:text=The%20researchers%20used%20generative%20AI,their%20illness%2C%20history%20or%20age](http://www.bcm.edu/news/researchers-create-ai-model-to-understand-how-brain-activity-relates-to-illness#:~:text=The%20researchers%20used%20generative%20AI,their%20illness%2C%20history%20or%20age). Accessed 19 June 2024.