Mini Project Report

on

SKIN CANCER DETECTION

Submitted by

T.Jagadeesh -22bcs130
T.Dhanraj -22bcs132
Shiva teja A.K -22bcs117
P. Snehith -22bcs081

Under the guidance

of

Dr. Sunil Kumar PV

Designation

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD

09/04/2024

Certificate

This is to certify that the project, entitled **Skin Cancer Detection**, is a bonafide record of the Mini Project coursework presented by the students whose names are given below during 2024-2025 in partial fulfilment of the requirements of the degree of Bachelor of Technology in Computer Science and Engineering.

Roll No	Names of Students
22bcs130	T Jagadeesh
22bcs132	T Dhanraj
22bcs132	Shiva teja A.K
22bcs132	P Snehith

Dr. Sunil kumar Pv
(Project Supervisor)

Contents

List of Figures ii

1 Introduction 1

2 Related Work 2

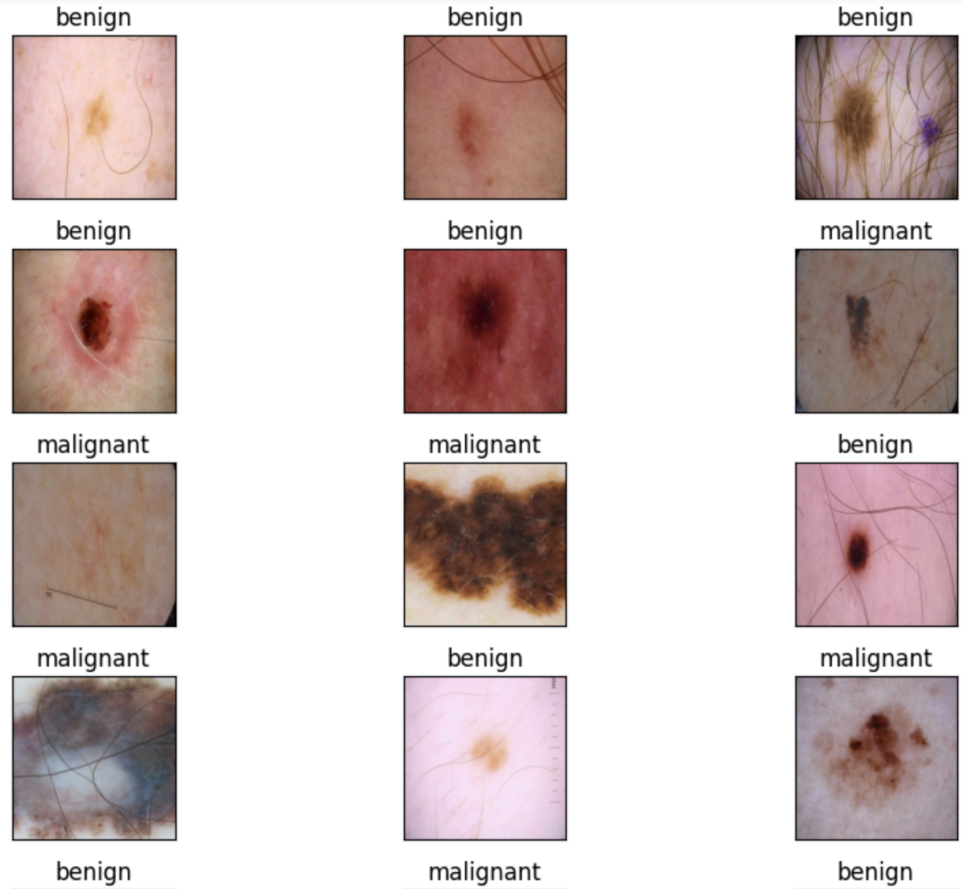
3 Data and Methods (Project Description) 3

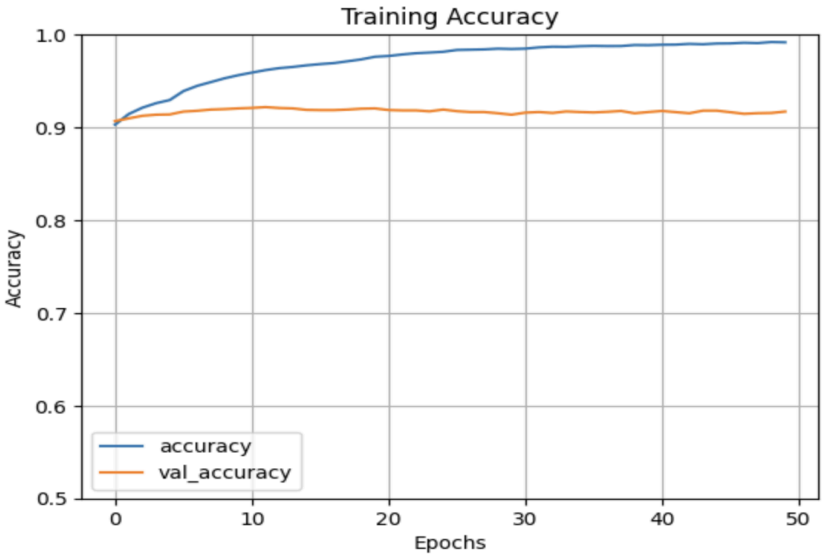
4 Results and Discussions 4

5 Conclusion 5

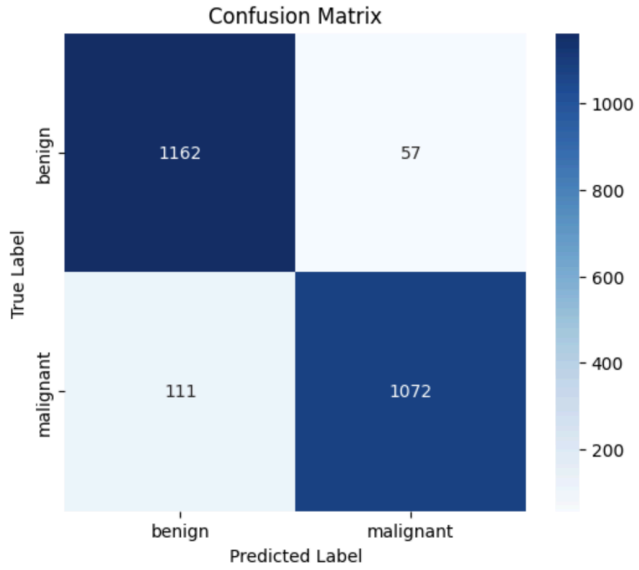
References 6

List of Figures

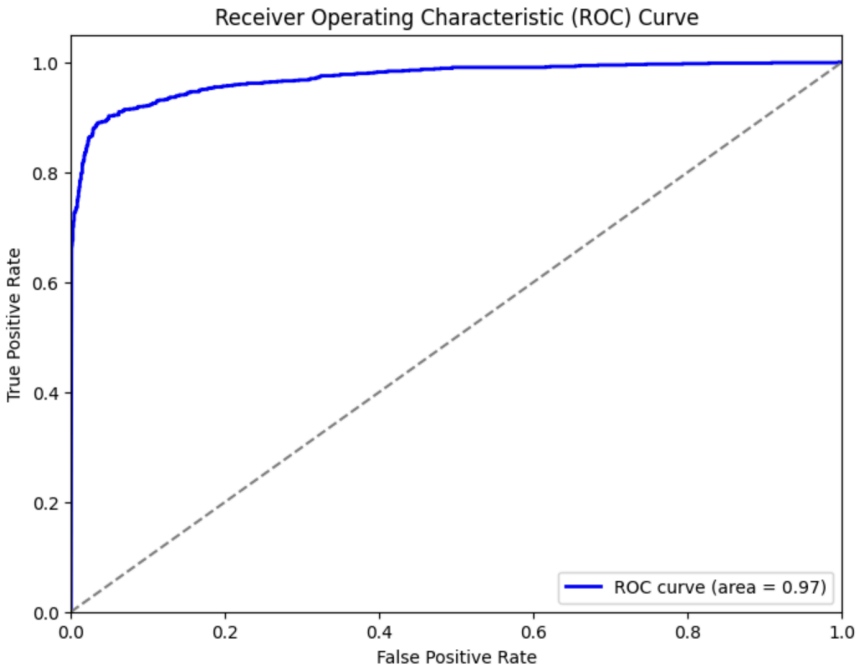




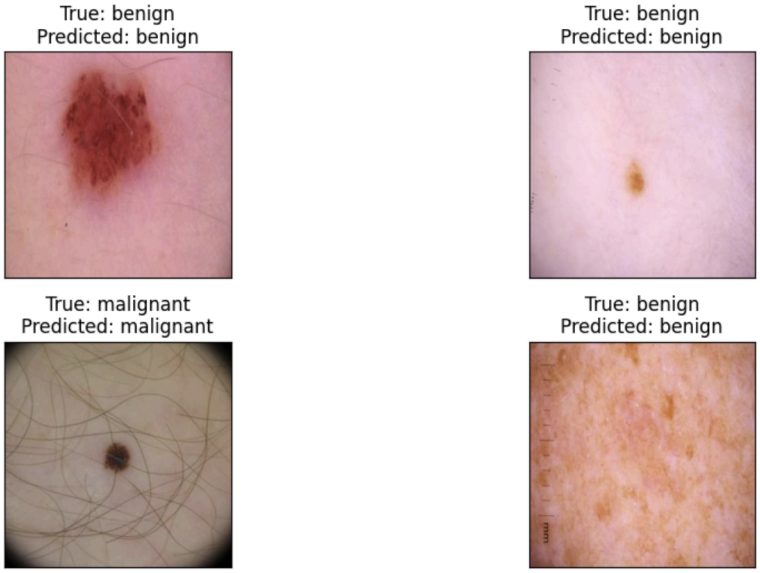
List of Figures



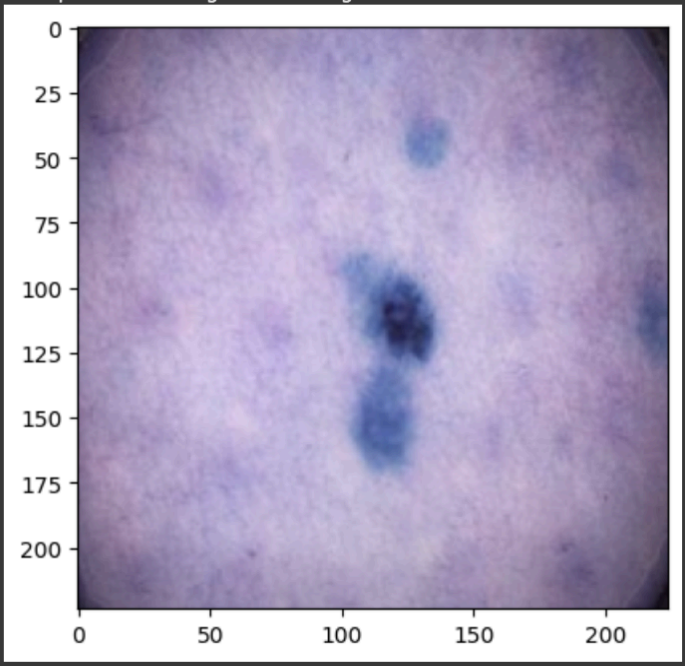
	precision	recall	f1-score	support
benign	0.92	0.95	0.93	1264
malignant	0.94	0.91	0.92	1138
accuracy			0.93	2402
macro avg	0.93	0.93	0.93	2402
weighted avg	0.93	0.93	0.93	2402




```
fig, axes = plt.subplots(nrows=3, ncols=2, figsize=(12, 8),
                         subplot_kw={'xticks': [], 'yticks': []})

for i, ax in enumerate(axes.flat):
    ax.imshow(plt.imread(test.Filepath.iloc[i]))
    ax.set_title(f"True: {test.Label.iloc[i]}\nPredicted: {pred[i]}")
plt.tight_layout()
plt.show()
```



```
1/1 ━━━━━━━━━━━━━━━━━━━━ 6s 6s/step
[[18 81]]
<matplotlib.image.AxesImage at 0x7c6417b3d750>
```




```
[ ] p=list((result*100).astype('int'))
    pp=list(p[0])
    print(pp)
```

```
[np.int64(18), np.int64(81)]
```

```
[ ] print("Largest element is:", max(pp))
```

```
Largest element is: 81
```

```
[ ] index = pp.index(max(pp))
```

```
[ ] name_class=['benign', 'Malignant']
```

```
[ ] name_class[index]
```

```
'Malignant'
```

1.INTRODUCTION

Skin cancer is one of the most prevalent and life-threatening diseases worldwide, with millions of new cases diagnosed each year. It primarily includes three major types: **melanoma**, **basal cell carcinoma (BCC)**, and **squamous cell carcinoma (SCC)**. Among these, melanoma is the most dangerous due to its high tendency to spread rapidly to other parts of the body. However, if detected early, melanoma has a **survival rate of up to 99%**, making early diagnosis critical for effective treatment and improved patient outcomes.

Traditional diagnostic methods typically involve **visual examination** by a dermatologist followed by **invasive biopsy procedures** to confirm malignancy. While these methods are effective, they are highly dependent on specialist expertise, are time-consuming, and may not be accessible in remote or under-resourced regions. As a result, there is a growing demand for automated, accurate, and accessible tools that can assist in early detection and classification of skin lesions.

With the advancement of **Artificial Intelligence (AI)**, particularly **Machine Learning (ML)** and **Deep Learning (DL)**, significant progress has been made in automating diagnostic processes in medical imaging. Among various deep learning techniques, **Convolutional Neural Networks (CNNs)** have proven exceptionally effective in analyzing and classifying complex image data, including medical images such as dermoscopic scans of skin lesions.

This project focuses on leveraging **transfer learning** using **EfficientNetB0**, a state-of-the-art CNN architecture developed by Google AI, known for achieving high accuracy with fewer parameters and lower computational cost compared to traditional models like VGG16 or ResNet. The model is trained on a large, curated dataset containing **10,000 dermoscopic images** of skin lesions, labeled as either benign or malignant.

The system is evaluated using standard performance metrics such as **accuracy**, **F1-score**, **recall**, and a **confusion matrix** to assess classification reliability. In summary, this project proposes a reliable, AI-assisted diagnostic solution for early detection of skin cancer using deep learning and transfer learning techniques. The ultimate goal is to bridge the gap between timely diagnosis and accessible healthcare, particularly in areas where expert dermatologists may not be readily available.

2. RELATED WORK

Traditional methods of diagnosing skin cancer involve manual examination by dermatologists, leading to variability in accuracy, delayed diagnosis, and the necessity of biopsies. There is a pressing need for an automated, AI-based solution that can analyze skin lesion images efficiently and accurately, reducing misdiagnoses and enabling early treatment.

Skin cancer detection has undergone significant transformation over the past decade with the integration of machine learning and deep learning techniques, particularly Convolutional Neural Networks (CNNs). This section explores previous research efforts and advancements in the field of automated skin lesion classification, highlighting key models, datasets, and methodologies used in recent years.

Early attempts at automated skin cancer detection were based on conventional image processing techniques, relying on handcrafted features such as ***color, texture, asymmetry, border irregularity, and diameter***—known collectively as the **ABCD** rule. These traditional methods, while useful, were often

limited by their dependence on domain expertise and susceptibility to variations in lighting, resolution, and lesion type.

The emergence of **machine learning algorithms** such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Random Forests marked a step forward. These methods allowed for basic lesion classification but still relied heavily on manually extracted features, which restricted their adaptability and scalability across different datasets.

The real breakthrough came with the advent of deep learning, especially **CNN**s, which significantly outperformed traditional approaches in image-based tasks. CNNs automatically learn hierarchical features from raw images, eliminating the need for manual feature extraction. Esteva et al. (2017) demonstrated that a CNN model trained on over 129,000 images could achieve dermatologist-level accuracy in classifying skin cancer, using a fine-tuned InceptionV3 architecture. Their research marked a pivotal moment in AI-assisted dermatology.

In subsequent years, several models including **VGG16, ResNet50, InceptionResNet**, and **DenseNet** have been applied to skin cancer classification tasks. Codella et al. (2018), in the ISIC Skin Lesion Analysis Challenge, showed that ensemble models combining deep networks with clinical metadata achieved improved diagnostic accuracy, reinforcing the role of CNNs in medical imaging.

The introduction of **Transfer learning**—where pretrained models on large datasets like ImageNet are fine-tuned for skin cancer detection—further improved model efficiency and performance, especially on smaller medical datasets. Recent research has shown that architectures like **EfficientNet** provide a better accuracy-parameter tradeoff, making them suitable for mobile deployment and resource-constrained environments.

Standard datasets such as **HAM10000**, **PH2**, and the **ISIC Archive** have become benchmarks for training and evaluating skin cancer detection models. These datasets offer a wide variety of labeled skin lesion images and have facilitated consistent comparisons across research efforts.

Despite these advancements, challenges remain. Many models are computationally expensive, lack interpretability, and are not always optimized for real-time or mobile applications. Additionally, class imbalance and intra-class variability in datasets continue to impact performance.

This project builds upon existing work by employing **EfficientNetB0**, a lightweight and scalable CNN architecture, to classify skin lesions as benign or malignant. Using a well-labeled dataset of 10,000 images, our approach aims to strike a balance between high accuracy and computational efficiency, offering potential for real-world deployment in dermatology tools.

3.DATA & METHODS

PROJECT DESCRIPTION

AIM :

To detect whether the skin lesion is Benign or Malignant.

OBJECTIVES :

Train and evaluate CNN model on melanoma skin lesions dataset to improve accuracy using transfer learning.

SCOPE :

Model predicts the lesion as Benign or Malignant.It aims to assist dermatologists by providing an diagnostic model.

METHODOLOGY:

- **Data Collection & Preprocessing:**
  - The dataset consists of 10,000 labeled skin lesion images.
  - Images are resized and enhanced for better feature extraction.
- **Model Development:**
  - Transfer learning is applied using "EfficientNetB0" or ResNet50.
  - The CNN model extracts important image features to classify lesions.
- **Training & Evaluation:**
  - The dataset is split into training, validation, and testing sets.
  - Performance metrics such as accuracy, F1-score, recall, and confusion matrix are used to evaluate the model

DATASET DESCRIPTION:

Dataset Name : Melanoma skin cancer Dataset.

Source : https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images

(This Dataset is licensed by ISIC.And similar to HAM10000 after folder classifications).

EXPECTED OUTCOMES :

1. Image/skin lesion is labelled("predicted") as Benign or Malignant.
2. Performance metrics demonstrating improved accuracy over traditional methods.

METHODOLOGY:

Skin cancer detection involves several stages, including data preprocessing, model development using transfer learning, training, and evaluation. The primary goal is to classify skin lesion images into two categories: **Benign** and **Malignant**. The workflow is designed to ensure high diagnostic accuracy while maintaining computational efficiency.

1. Data Collection and Preprocessing

The dataset used in this project consists of 10,000 dermoscopic images of skin lesions, categorized into benign and malignant classes. The images were sourced from a publicly available dataset that closely resembles the HAM10000 and ISIC archives in structure and quality.

- **Image Resizing:** All images were resized to 224x224 pixels to match the input requirements of EfficientNetB0.
- **Label Encoding:** Class labels (benign or malignant) were encoded for binary classification.
- **Data Augmentation:** To improve generalization and address class imbalance, various augmentation techniques were applied, including horizontal/vertical flipping, rotation, zoom, and brightness adjustment.
- **Normalization:** Pixel values were normalized to a range of [0, 1] to speed up model convergence.

2. Model Architecture: EfficientNetB0 with Transfer Learning

EfficientNetB0, a lightweight yet powerful convolutional neural network developed by Google, was chosen for this task due to its optimal balance between accuracy and efficiency. It uses a compound scaling method to scale depth, width, and resolution in a balanced way.

- **Transfer Learning Approach:**
  The base EfficientNetB0 model, pretrained on ImageNet, was imported with its convolutional layers frozen initially. A custom classification head was added, consisting of:
  - Global Average Pooling layer
  - Dense layer with ReLU activation
  - Final Dense layer with a Softmax activation for Probabilities as output.

- **Unfreezing Layers:** After initial training, the top layers of EfficientNetB0 were unfrozen to fine-tune the model using our skin lesion dataset, further improving performance.

EfficientNetB0: (Pretrained Model)

EfficientNetB0 is a convolutional neural network (CNN) model developed by Google AI, and it serves as the baseline architecture in the EfficientNet family. It is designed to deliver state-of-the-art accuracy while being significantly more efficient in terms of parameters and computation. EfficientNetB0 is trained on the **ImageNet** dataset, which contains over 1.2 million images across 1,000 classes, allowing it to learn powerful and generalized image features.

1. Compound Scaling Method

EfficientNet introduces a unique **compound scaling** technique that uniformly scales all dimensions of the network—**depth (number of layers)**, **width (number of channels)**, and **input resolution (image size)**—using a set of fixed scaling coefficients. This method allows the network to grow in a balanced manner, resulting in better performance with fewer resources compared to traditional CNNs.

2. Architecture Overview

EfficientNetB0 uses several advanced design techniques:

- **Mobile Inverted Bottleneck Convolution (MBConv)**: These layers are optimized for mobile devices and include Depth Wise separable convolutions to reduce the number of parameters and computations.

- **Squeeze-and-Excitation (SE) blocks**: These improve the network's sensitivity to important features by recalibrating channel-wise feature responses.

- **Swish activation function**: A smooth, non-monotonic activation function that outperforms ReLU in many tasks.

3. Model Training

- **Loss Function:** Binary Cross-Entropy Loss was used, as the task involves binary classification.
- **Optimizer:** Adam optimizer was used with a learning rate scheduler to adaptively update learning rate.
- **Epochs and Batch Size:** The model was trained for 50 epochs with a batch size of 32.
- **Validation Split:** 80% of the dataset was used for training and 20% for validation.

4. Evaluation Metrics

To evaluate the performance of the model, the following metrics were used:

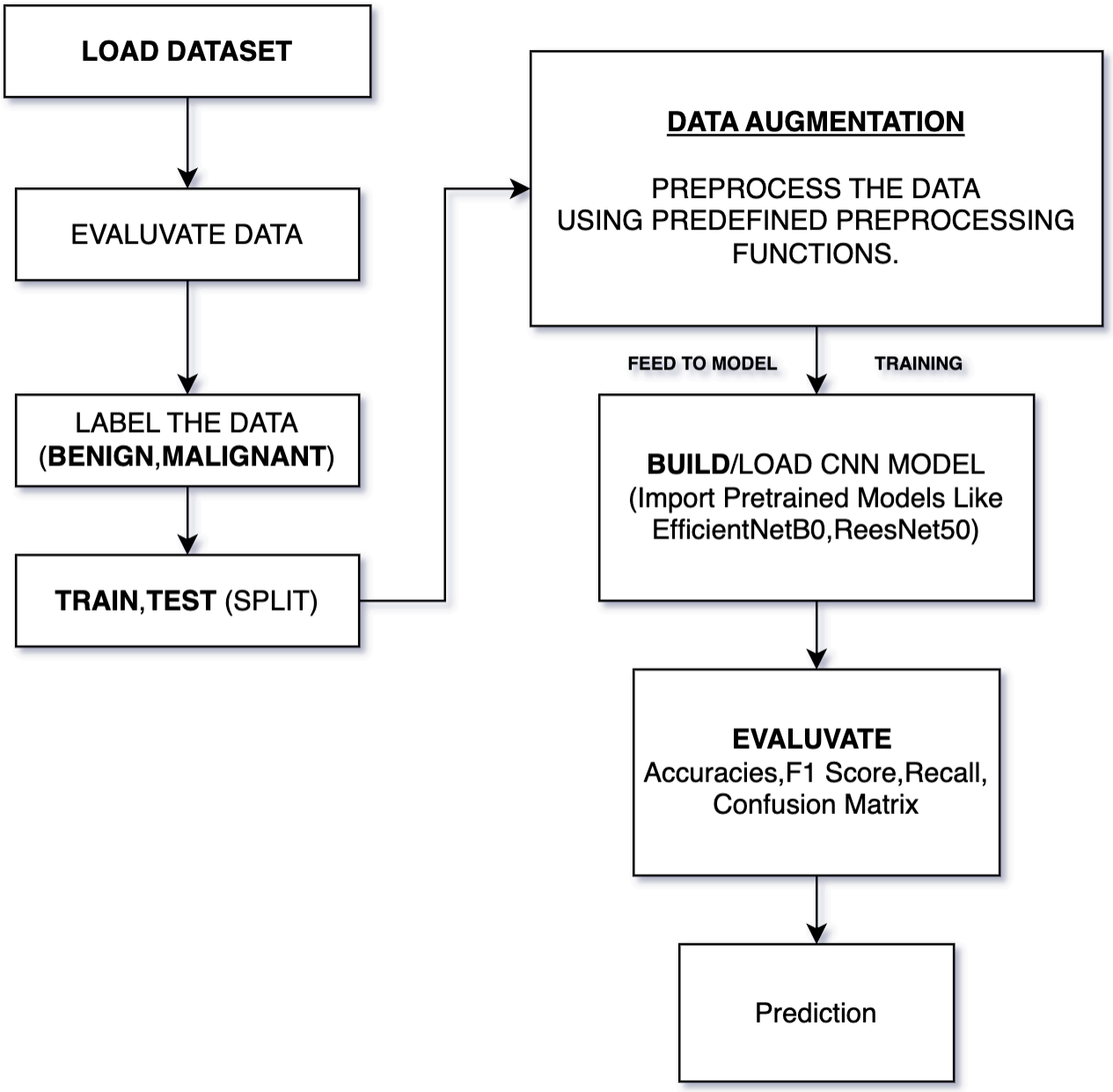
- **Accuracy**
- **Precision, Recall, F1-score**
- **Confusion Matrix**
- **ROC-AUC Curve**

These metrics help in understanding the model's ability to correctly classify malignant cases (reducing false negatives) which is critical in cancer diagnosis.

5. Prediction

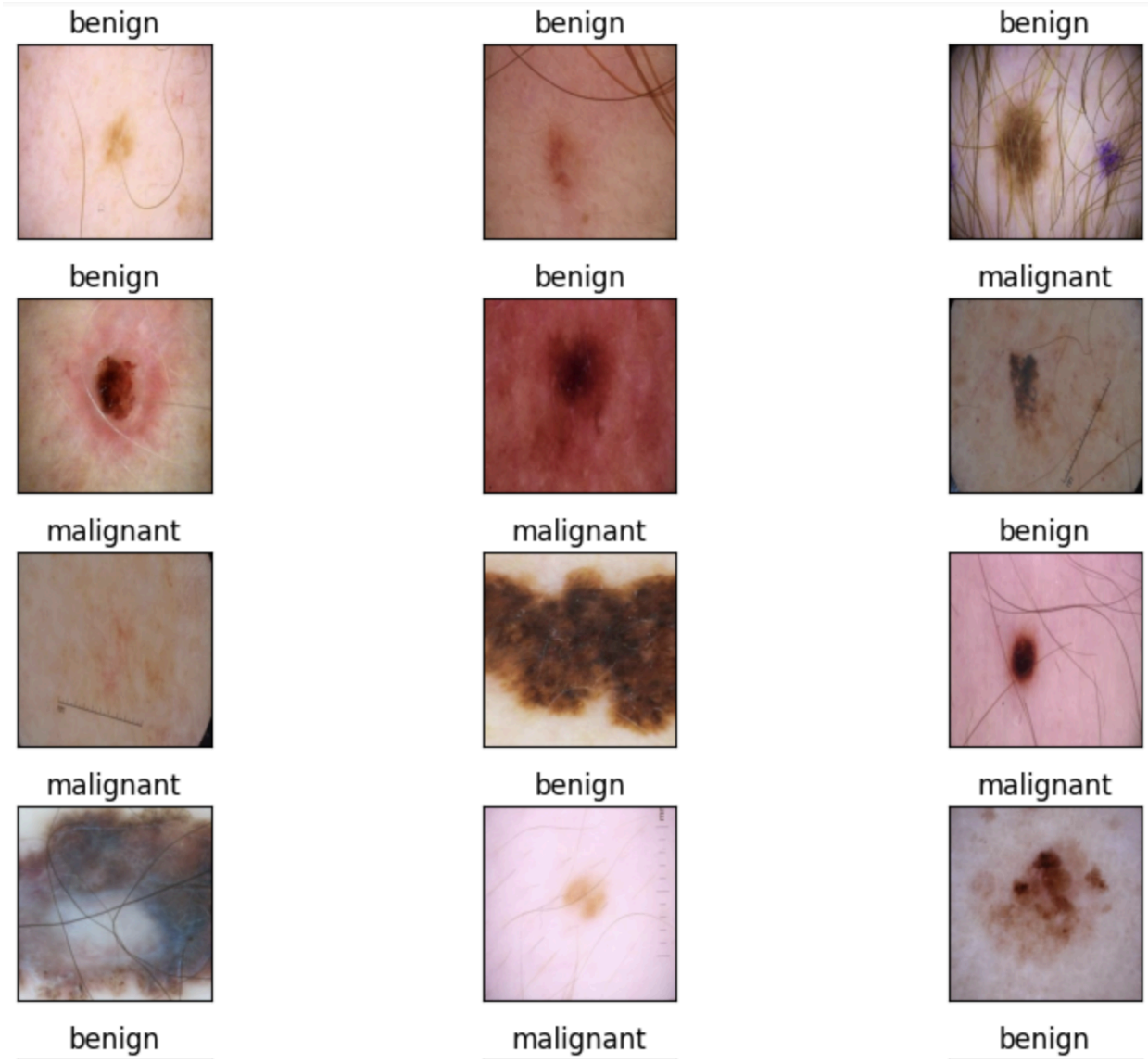
Once trained, the model was used to make predictions on unseen test data.

FLOW OF PROJECT :



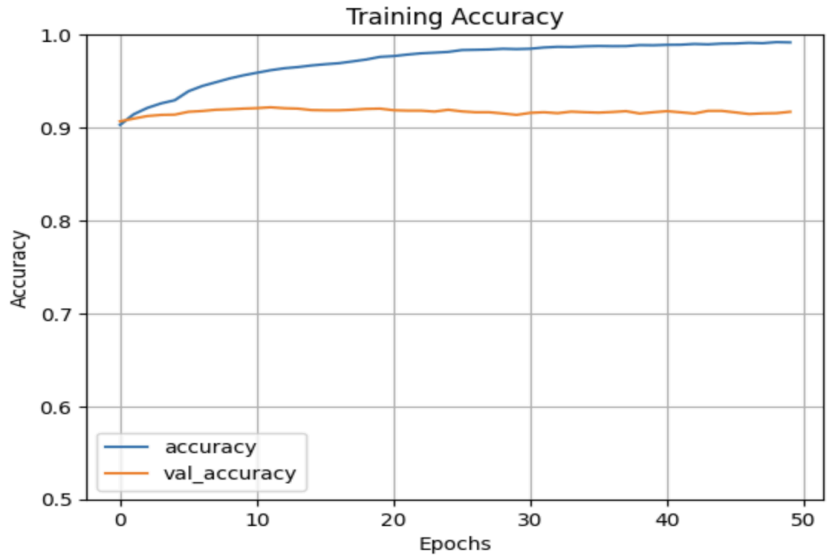
RESULTS & EVALUATION:

- Sample images in dataset after aligning the labels.

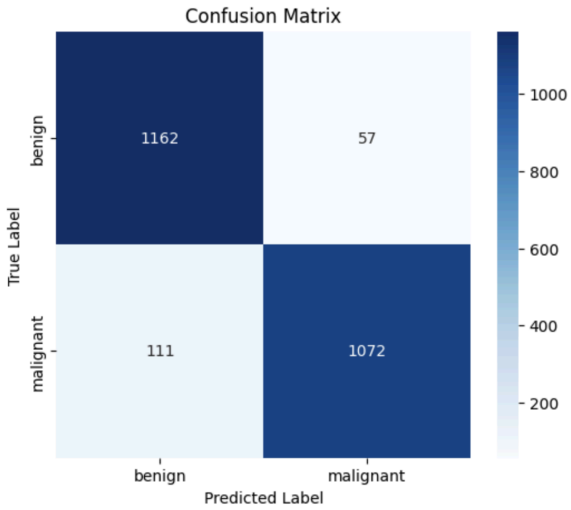


TRAINING:

- Epochs=50
- Training Accuracy = 99%
- Training Loss =0.0275
- Validation Accuracy = 93%
- Validation Loss =0.3808



- ***Confusion matrix***

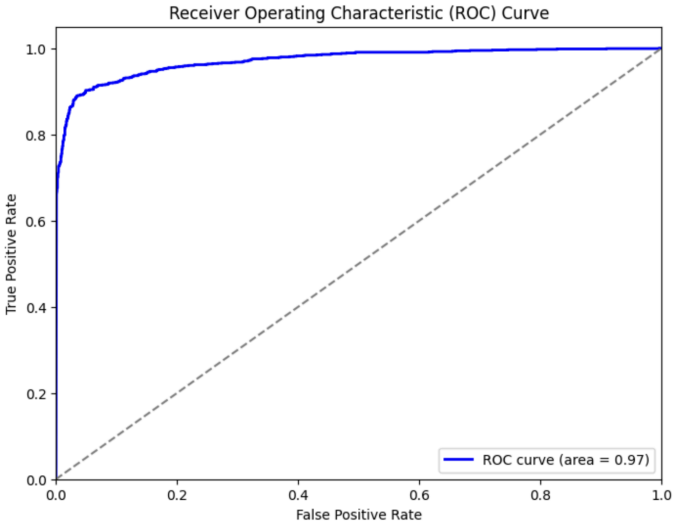


- ***precision,recall,f1-scores.***

```
y_test = list(test.Label)
print(classification_report(y_test, pred))
```

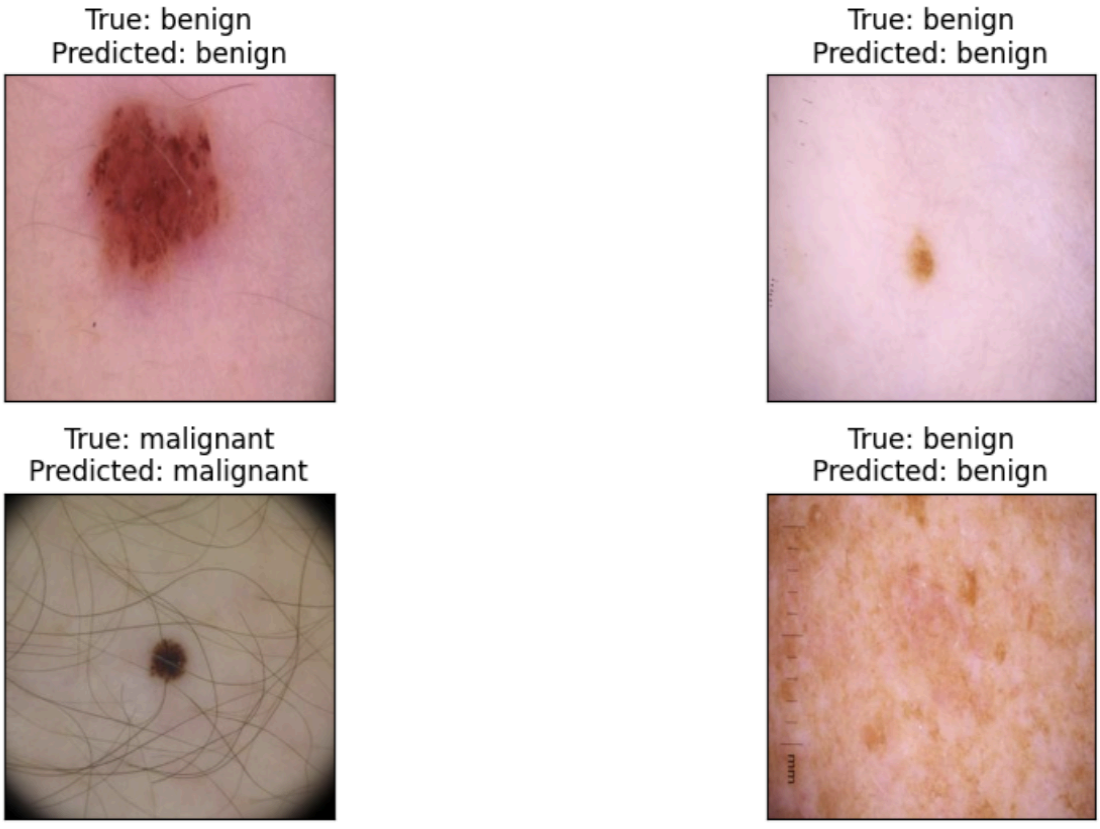
	precision	recall	f1-score	support
benign	0.92	0.95	0.93	1264
malignant	0.94	0.91	0.92	1138
accuracy			0.93	2402
macro avg	0.93	0.93	0.93	2402
weighted avg	0.93	0.93	0.93	2402

- ***Roc- Auc Curve***



- ***Some predictions.***

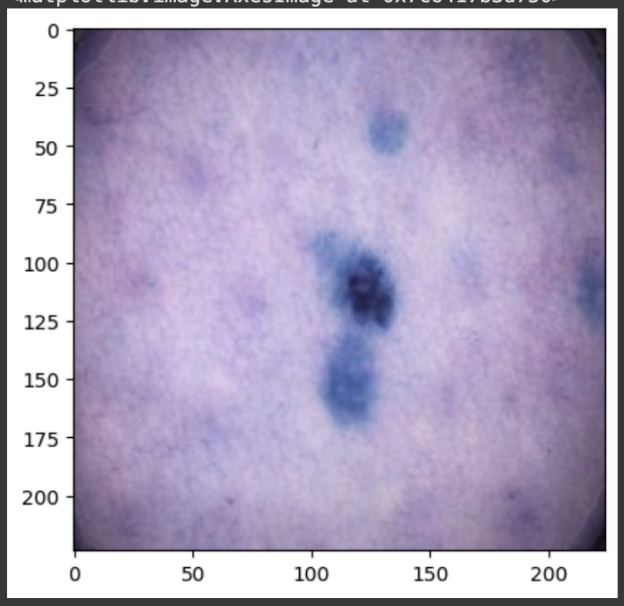
```
fig, axes = plt.subplots(nrows=3, ncols=2, figsize=(12, 8),
                         subplot_kw={'xticks': [], 'yticks': []})

for i, ax in enumerate(axes.flat):
    ax.imshow(plt.imread(test.Filepath.iloc[i]))
    ax.set_title(f"True: {test.Label.iloc[i]}\nPredicted: {pred[i]}")
plt.tight_layout()
plt.show()
```



PREDICTION:

- Given : /content/skin_cancer_dataset/melanoma_cancer_dataset/test/**malignant**/melanoma_10604.jpg
- Predicted = “Malignant”

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 6s 6s/step
[[18 81]]
<matplotlib.image.AxesImage at 0x7c6417b3d750>
```




```
[ ] p=list((result*100).astype('int'))
    pp=list(p[0])
    print(pp)
```

```
[np.int64(18), np.int64(81)]
```

```
[ ] print("Largest element is:", max(pp))
```

```
Largest element is: 81
```

```
[ ] index = pp.index(max(pp))
```

```
[ ] name_class=['benign', 'Malignant']
```

```
[ ] name_class[index]
```

```
'Malignant'
```

The high accuracy and precision scores indicate that EfficientNetB0, when fine-tuned on dermoscopic images, is highly capable of classifying skin lesions. The use of **data augmentation**, **transfer learning**, and **fine-tuning** were critical in achieving these results. The effectiveness of EfficientNetB0 for early skin cancer detection. With high accuracy and low misclassification rates, the system can serve as a reliable second-opinion tool for dermatologists and aid in early intervention strategies.

CONCLUSION

This project presents a deep learning-based approach for early skin cancer detection using Convolutional Neural Networks (CNNs), with a focus on improving diagnostic accuracy and reducing reliance on invasive procedures. By leveraging transfer learning with the EfficientNetB0 architecture, the model effectively classifies skin lesions into benign and malignant categories, achieving high training and validation performance. The use of a large and diverse dataset, combined with data augmentation and advanced evaluation metrics, ensures robust model generalization and reliability.

The results demonstrate the model's potential as a clinical decision support tool, aiding dermatologists in early and accurate diagnosis of skin cancer. With a training accuracy of 99% and validation accuracy of 93%, the proposed system shows promise in real-world deployment, especially in areas lacking access to specialized healthcare. The integration of lightweight and efficient deep learning models also opens possibilities for mobile and edge-based applications, making skin cancer screening more accessible and affordable.

Future work may involve expanding the model to multi-class classification for different types of skin cancer, incorporating patient metadata for personalized diagnosis, and enhancing interpretability using explainable AI techniques. Overall, this study contributes to the growing field of AI in healthcare, providing a reliable, non-invasive, and scalable solution for early skin cancer detection.

REFERENCE

1.A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images.
https://www.researchgate.net/publication/372139502_A_deep_neural_network_using_modified_EfficientNet_for_skin_cancer_detection_in_dermoscopic_images

2. a deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training timehttps://link.springer.com/article/10.1007/s00521-024-10225-y.

3 Skin Cancer Detection Using Deep Learning
_https://pmc.ncbi.nlm.nih.gov/articles/PMC10252190/

4. Skin cancer detection using deep machine learning techniques
https://www.sciencedirect.com/science/article/pii/S2666521224000589

5. Skin cancer detection and classification using machine learning
https://www.researchgate.net/publication/343673685_Skin_cancer_detection_and_classification_using_machine_learnin