**Annexure-I Term Paper**

**Skin Cancer Classification using Deep Learning**

**A Term Paper Report**

**Submitted in partial fulfilment of the requirements for the award of degree of Bachelor of Technology**

**(Computer Science Engineering) Submitted to**



# LOVELY PROFESSIONAL UNIVERSITY PHAGWARA, PUNJAB

**From 1st Feb 2023 to 25st April 2023 SUBMITTED BY**

**Name of student: Abhay Jagetiya Registration Number: 12019398 Faculty: Ajay Sharma**

**Annexure-II: Student Declaration To whom so ever it may concern**

I, **Abhay Jagetiya, 12019398**, hereby declare that the work done by me on **“Skin Cancer Detection using Deep Learning”** from Feb 2023 to April 2023, is a record of original work for the partial fulfilment of the requirements for the award of the degree, Bachelor of Technology.

**Name of the student:** Abhay Jagetiya

**Registration Number:** 12019398

**Dated: 26TH APRIL 2023**

# ACKNOWLEDGEMENT

Primarily I would like to thank God for being able to learn a new technology. Then I would like to express my special thanks of gratitude to the teacher and instructor of the course Machine Learning who provided me the golden opportunity to learn a new technology.

I would like to also thank my own college Lovely Professional University for offering such a course which not only improve my programming skill but also taught me other new technology.

Then I would like to thank my parents and friends who have helped me with their valuable suggestions and guidance for choosing this course.

Finally, I would like to thank everyone who have helped me a lot.

**Dated: 26TH APRIL 2023**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Contents** | **Page** |
| **1** | **Title** | **1** |
| **2** | **Student Declaration** | **2** |
| **3** | **Acknowledgement** | **2** |
| **4** | **Table of Contents** | **3** |
| **5** | **Abstract** | **4** |
| **6** | **Objective** | **4** |
| **7** | **Introduction** | **5-7** |
| **8** | **Theoretical Background** | **7-9** |
| **9** | **Hardware & Software** | **10** |
| **10** | **Methodology** | **10-23** |
| **11** | **Flowchart** | **12** |
| **12** | **Results** | **24** |
| **13** | **Summary** | **24** |
| **14** | **Conclusion** | **25** |
| **15** | **Bibliography** | **25** |
| **16** | **Annexure** | **26** |

# ABSTRACT

The most frequent type of cancer in the world is skin cancer. Early detection and diagnosis are vital for effective treatment. Deep learning has been found to be effective in the classification of skin cancer. In this report, we present a deep learning approach for classifying skin cancer. The algorithm was trained on approximately 10000 photos of skin cancer. In our method, we have used convolutional neural network (CNN) for skin cancer classification. The CNN model is next trained on a collection of skin data tagged as benign or malignant. We validate our method using a publicly accessible database of skin images. For the training dataset, our method obtains a x% accuracy. For the test set, the model achieves an accuracy of more than y%. The model can accurately categorize both benign and malignant skin cancer. The model is a useful method for skin cancer early detection and treatment.

# OBJECTIVE

Because of the poor contrast between skin lesions and normal skin regions, accurately segmenting the lesion areas is difficult. Perceptual similarities between the two regions may be very strong. This makes it difficult to distinguish between melanoma and non-melanoma conditions. Patients' skin disorders, such as skin tone, natural hairs, and veins, differ. This complicates distinguishing melanoma disorders based on texture, colour, and other characteristics. The main goal of this research is to eradicate human involvement in melanoma cancer diagnosis, making it less error- prone and time consuming.

To achieve this goal, the project aims to develop an accurate and reliable classification model that can automatically distinguish between melanoma and non-melanoma skin lesions. The model will be trained on a large dataset of images of skin lesions, including various skin tones, hair types, and other patient characteristics. The objective is to design a deep learning algorithm that can detect subtle differences in texture, colour, and other features that are indicative of melanoma. By automating the diagnosis process, this project hopes to reduce the time and resources required for accurate melanoma detection, and ultimately improve patient outcomes. The successful development of this model would be a significant step towards eradicating the need for human involvement in the diagnosis of skin cancer, leading to a more efficient and accurate diagnosis process.

# INTRODUCTION

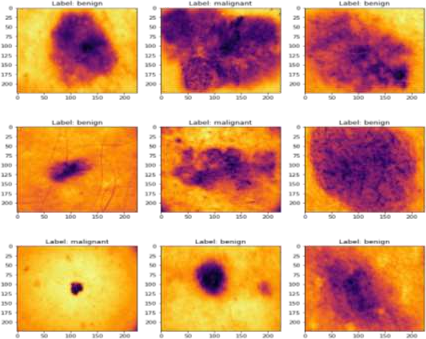
1. **Background**

The most prevalent form of cancer in humans, skin cancer is typically identified visually after a clinical screening, with possible additional testing including dermoscopic analysis, a biopsy, and histopathological analysis. The fine-grained heterogeneity in the pattern of skin lesions makes it difficult to automatically classify them using images.

* + What is **SKIN CANCER**?

Skin cancer, which is the abnormal growth of skin cells, most occurs commonly on skin that has been exposed to sunlight. This usual type of cancer can also appear on parts of your skin that are not often exposed to sunlight.

Cancer develops when healthy cells mutate and expand uncontrollably, generating a mass known as a tumour. A tumour might be malignant or benign. Cancerous tumours are malignant, which means they can spread to other parts of the body. Although benign tumours might grow to be large, they will not infect neighbouring tissue or spread to different areas of your body. The boundaries of benign tumours are obvious, smooth, and regular, whereas malignant tumours have uneven borders and expand quickly than benign ones. There are two major types of skin cancer, which are stated below: Benign or Malignant.



**Figure 1: Real Data Images**

1. **Other types of skin Cancer**

Skin cancer comes in a variety of forms, including:

1. **Basal cell carcinoma (BCC):** This type of skin cancer is the most prevalent. It typically manifests as a tiny, glossy lump or nodule on the skin, which may be pink or flesh-coloured. Although BCC seldom spreads to other body regions, it can still be harmful if neglected.
2. **Squamous cell carcinoma (SCC):** The second most typical kind of skin cancer is squamous cell carcinoma (SCC). It typically manifests as a red, scaly patch or a raised lump with a crusted surface. If SCC is not treated, it may spread to other body parts.
3. **Melanoma:** Because of its rapid ability to spread to other body parts, melanoma is the most serious type of skin cancer. It frequently appears as a black, asymmetrical mole or lesion, but it can also be pink or flesh-colored. If melanoma is not detected and treated promptly, it can be fatal.
4. **Merkel cell carcinoma (MCC):** Usually presenting as a flesh-colored or bluish-red nodule on the skin, MCC is an uncommon and aggressive form of skin cancer. It can quickly spread to different bodily parts.
5. **Dermatofibrosarcoma protuberans (DFSP):** This uncommon form of skin cancer often manifests as a firm, elevated growth that is flesh-colored or pink.
6. **Atypical fibroxanthoma (AFX):** A form of skin cancer that manifests on the skin as a red, dome-shaped nodule.
7. **Kaposi sarcoma:** A virus-induced form of skin cancer that frequently manifests as a purple, red, or brown lesion on the skin.
8. **Causes of skin cancer**

The main cause of skin cancer is exposure to ultraviolet (UV) radiation from the sun or other sources, such as tanning beds or lamps. When UV radiation penetrates the skin, it can damage the DNA in skin cells, which can lead to mutations and the development of skin cancer.

Other skin cancer risk factors include:

1. **Fair skin:** Individuals with fair skin are more vulnerable to UV radiation-induced skin damage.
2. **Family history:** Skin cancer can run in families, so you may be at a higher risk if you have a close family who has had the disease.
3. **Age:** As the skin ages, it loses its capacity to repair UV damage. As a result, skin cancer risk rises with age.
4. **Sunburn history:** Frequent sun exposure might raise the chance of developing skin cancer.
5. **Immune system weakness:** Individuals with immune system weaknesses, such as those with HIV/AIDS, are more likely to acquire skin cancer.
6. **Chemical exposure:** Exposure to some chemicals, such arsenic or coal tar, can raise the chance of developing skin cancer.
7. **A few medical diseases:** A few medical conditions, such basal cell nevus syndrome or xeroderma pigmentosum, can raise the risk of skin cancer.
8. **Existing Cures and Remedies of Skin Cancer**

Skin cancer can be treated using a variety of methods, including:

1. **Surgery:** A surgical operation is used to remove the malignant skin tissue.
2. **Radiation therapy:** In this procedure, cancer cells are killed by high-energy radiation.
3. **Chemotherapy:** This method uses medication to destroy cancer cells.
4. **Immunotherapy:** To combat cancer, this therapy makes use of the immune system.
5. **Photodynamic therapy:** This method includes killing cancer cells by activating a photosensitizing chemical with light.
6. **Need of Machine learning**

By analysing vast volumes of data and finding patterns that may suggest the presence of cancer, machine learning can be utilised for early detection of skin cancer. This can be achieved by utilising machine learning models that have been trained on a dataset of photos of both healthy and malignant skin to categorise fresh images as either healthy or cancerous.

# THEORETICAL BACKGROUND

1. What is **Artificial Neural Network?**

An artificial neural network (ANN) is a form of machine learning model inspired by the human brain. ANNs can learn complicated patterns from data and have been demonstrated to be very effective in a range of applications such as image classification, natural language processing, and speech recognition.

ANNs are composed of artificial neurons that are linked together. The artificial neurons are organised in layers, with each layer performing a specific purpose. The data is received by the input layer, the hidden layers analyse the data patterns, and the predictions are produced by the output layer.

1. What is **Deep Learning?**

Deep learning is a subset of machine learning that learns from data using artificial neural networks. Deep learning has been utilised to produce cutting-edge solutions in a wide range of applications, including image recognition, natural language processing, and speech recognition.

Deep learning models are mostly made up of several layers of artificial neurons. Each unit of neurons receives the previous layer's output as input and generates a response that is sent into the following layer. During training, the weights of the neural networks are trained, allowing the model to learn connections between the data input and the predicted output. Deep learning outperforms classic machine learning techniques in various ways.

For starters, deep learning can understand complicated patterns and correlations in data that typical machine learning approaches would struggle to learn. Second, deep learning can learn from enormous volumes of data, which can enhance the accuracy of predictions. Third, deep learning generalises well to new data, it means it can make correct predictions on data it has never seen before.

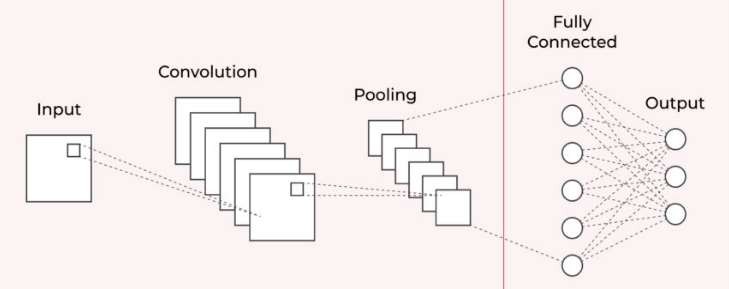
1. What is **Convolutional Neural Network?**

Convolutional neural networks (CNNs) are artificial neural networks that are specifically designed to process input with a grid-like structure, like images. CNNs can learn spatial correlations between image pixels, making them ideal for image classification, object recognition, and segmentation.

CNNs are constructed up of layers, and each one serves a distinct purpose. The convolutional layer is the initial layer of a CNN and is responsible for obtaining the image's features. The convolutional layer scans the image as input and extracts features using a filter, which is a tiny matrix of weights. The features retrieved by the convolution layers are passed on to the pooling layer. The pooling layer is in charge of narrowing the feature size map, which aids in lowering the network's computational complexity. To lower the size of the feature map, the pooling layer employs a pooling technique such as max pooling or average pooling.

Fully connected is the subsequent layer in a CNN. The fully connected layer is a typical neural network layer in which each neuron is linked to every neuron in the previous layer. The fully connected layer is in charge of categorising the input image. The fully connected layer employs a SoftMax function to generate a probability distribution across the various classes.

CNNs have been demonstrated to be extremely effective for a wide range of applications, including image classification, object identification, and segmentation. CNNs are employed in many different applications, including self-driving automobiles, medical image analysis, and facial recognition.



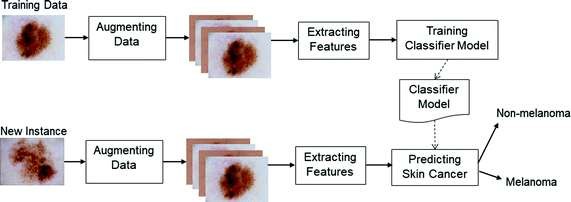
Feature Extraction Classification

**Figure 2: CNN Model**

1. **Image Classification** using **Deep Learning**

Image classification is a form of machine learning problem in which the objects or events in an image are identified. It is a difficult task because images might be complicated and contain several items. Deep learning, on the other hand, has made image classification far more precise and efficient.

There are numerous deep learning architectures available for image classification. The convolutional neural network is a popular architecture (CNN). CNNs are ideal for image classification because they can learn spatial correlations among pixels in an image.



**Figure 3: Skin Cancer Classification Model Working**

1. Benefits of using **Deep Learning Model for Classification**
   * Deep learning models can recognise minor details that humans cannot perceive.
   * Deep learning models can be trained on massive image datasets, so they can learn more generalizable features.
   * Deep learning models can be utilised to create automated screening systems for skin cancer.
2. Challenges in using **Deep Learning Model for Classification**
   * Deep learning models need to be trained on huge datasets of labelled data.
   * Deep learning models can be complex to train and deploy in terms of computational power.
   * Deep learning models are prone to overfitting.

**Hardware & Software Requirements**

**Hardware**

Tensor Processing Units (TPUs), or graphics processing units (GPUs): Deep learning model training is a good fit for these specialist processors' parallel processing capabilities. When compared to using conventional CPUs, GPUs and TPUs can greatly accelerate the training process.

**Software**

Python: Python is a well-liked machine learning programming language with a wide variety of tools and frameworks for creating models. TensorFlow and scikit-learn are a few well-liked libraries for spotting skin cancer.

Frameworks for deep learning: These are software collections that offer resources for creating and refining deep learning models. TensorFlow and Keras are a few popular deep learning frameworks for the diagnosis of skin cancer.

# METHODOLOGY

* + **Importing required Libraries**



The os library provides a way to interact with the operating system, such as navigating directories, creating and deleting files, and other related functionalities.



The random library is used to generate random numbers for various purposes, such as shuffling the data during training or testing.



Pandas is a popular library for data analysis and manipulation. It provides data structures for efficient data manipulation, such as data frames and series.



Numpy is a library for numerical computing and array processing in Python. It provides tools for efficient manipulation of large arrays and matrices, and has various mathematical functions for operations on these arrays.



Seaborn is a data visualization library based on matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics.



Matplotlib.pyplot is a plotting library which provides functions for creating various types of charts, such as line charts, scatter plots, bar charts, and histograms.



Matplotlib.ticker is a library used to customize tick locators and formatters for various plots.



Plotly.graph\_objs is a module of the Plotly library used for creating interactive data visualizations.



Plotly.express is a high-level interface to Plotly for creating easy-to-use visualizations.



Glob is a library used to search for files in a directory using wildcard patterns.



PIL (Python Imaging Library): PIL is a library used to manipulate and process images in Python.

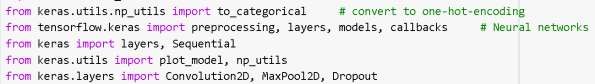


Tensorflow is an open-source library for machine learning and deep learning developed by Google. It provides tools for building and training various types of models, including neural networks.

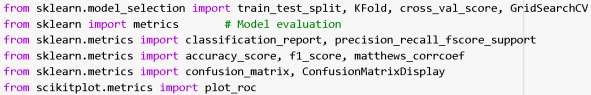


Tensorflow\_hub is a library used for reusable machine learning modules, such as pre-trained models.



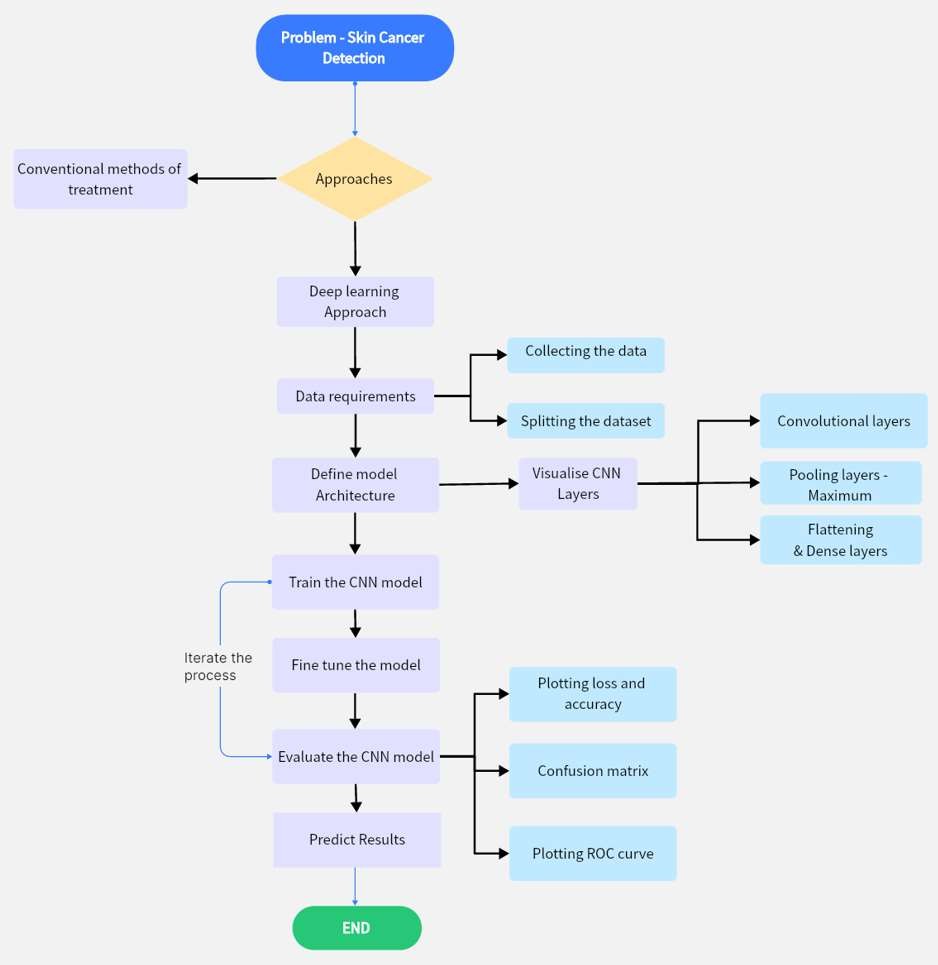


Keras is a high-level API for building and training neural networks in Python. It provides a user- friendly interface for building and training various types of neural networks.



Scikit-learn is a popular library for machine learning in Python. It provides tools for various machine learning tasks, such as classification, regression, clustering, and dimensionality reduction.

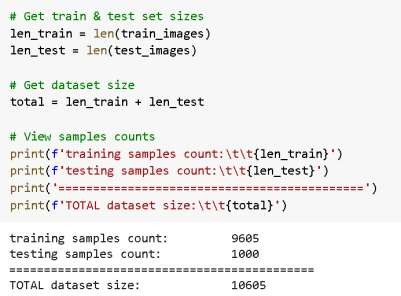
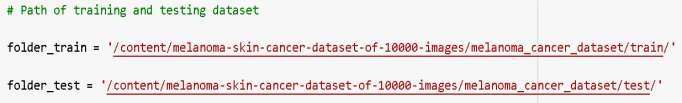
# FLOWCHART – CNN MODEL



Flowchart-1: Project flow

* **Data Collection and Inspection of dataset**

Download dataset directly from Kaggle using username and Kaggle API.



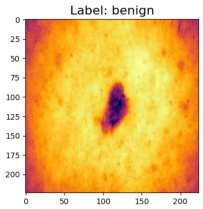
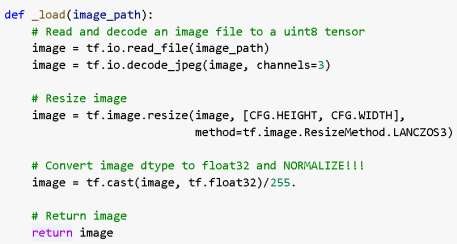
* **Pre-processing images**



* **Preparing Dataset**



* **Analyzing images**



* **Visualizing images**



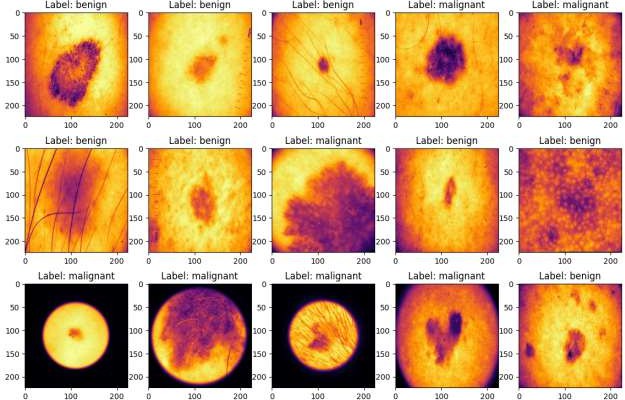
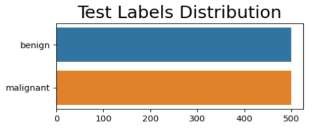
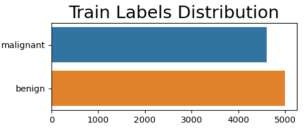


Figure:4 Images with Labels as: benign and malignant

* **Plotting Training vs Testing Dataset**



* **Augmentation Layer for Images**

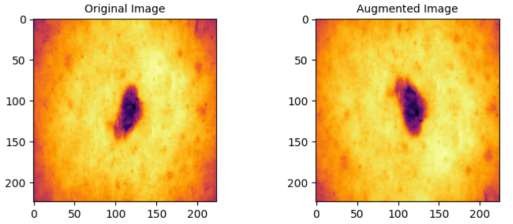
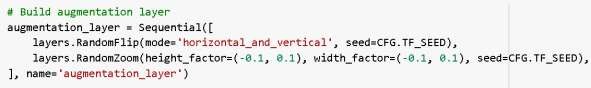
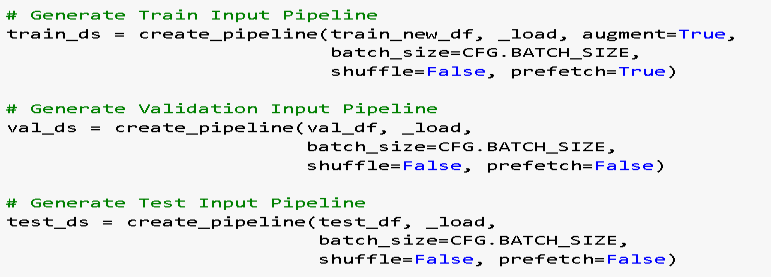
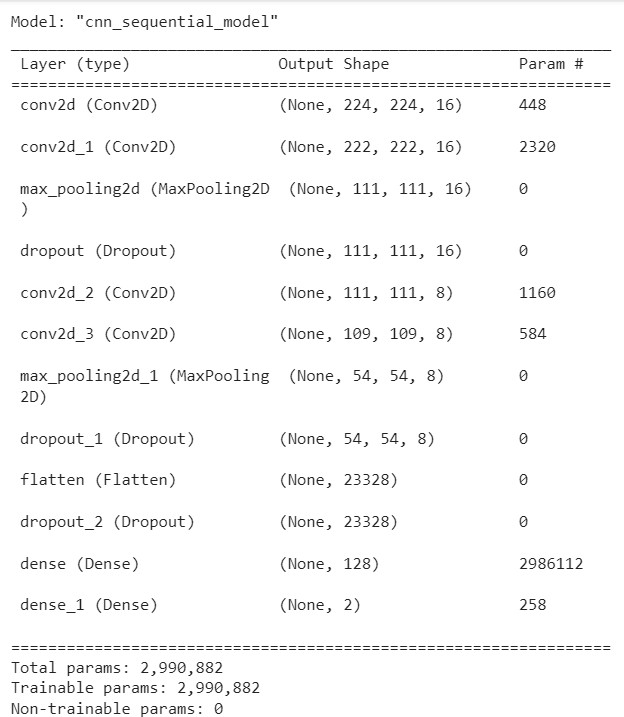


Figure:5 Original Image vs Augmented Image

* **Pipelines: Training, Validation & Testing Dataset**



* **CNN Model**
* **Model Summary**



* **Layers of CNN Model**

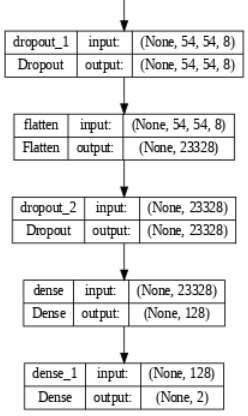
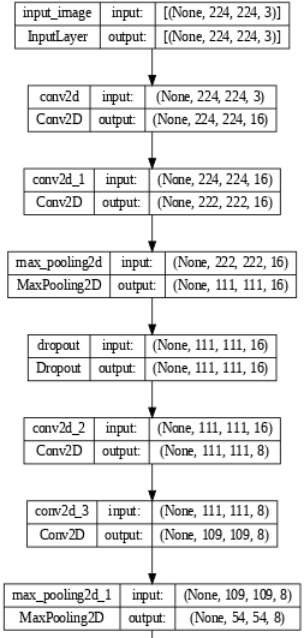
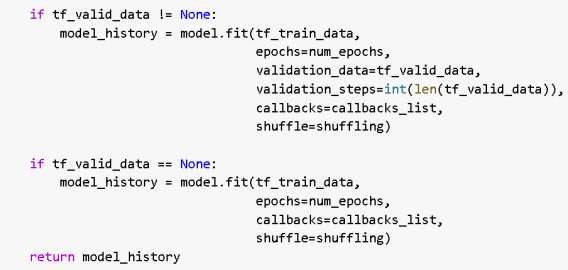
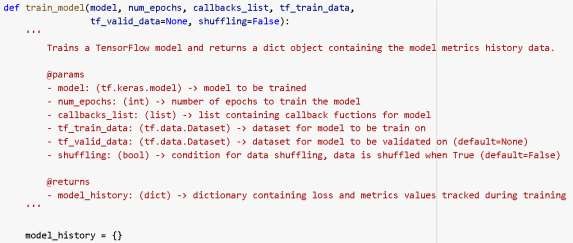
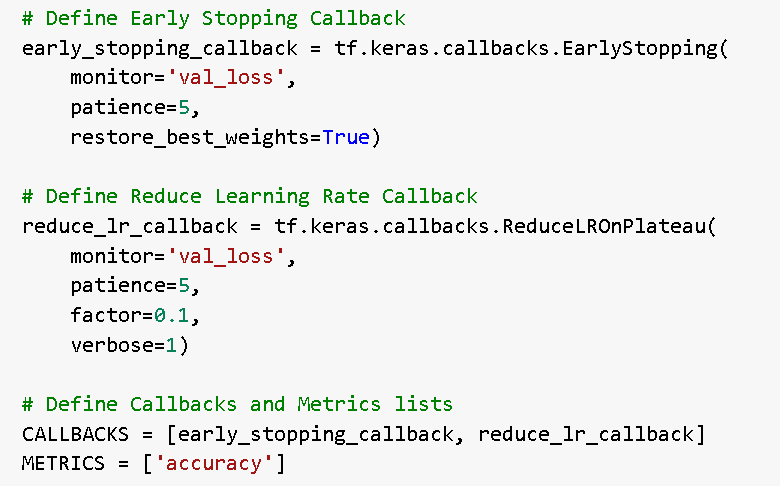


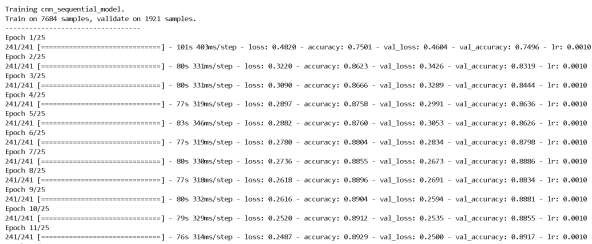
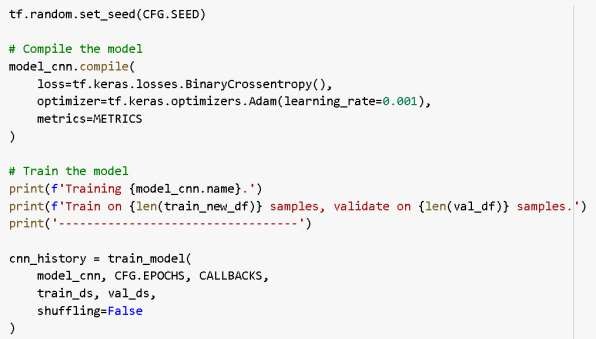
Figure:6 CNN Layers

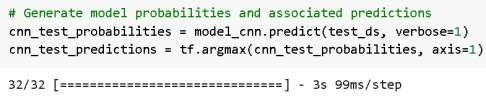
* **Layers of CNN Model**



* **Early Stopping Callback & Reduce Learning Rate Callback**



* **Training CNN Sequential Model with Epoch**
* **Model Evaluation**
* **Model Predictions**



* **Loss & Accuracy metrics for Model**

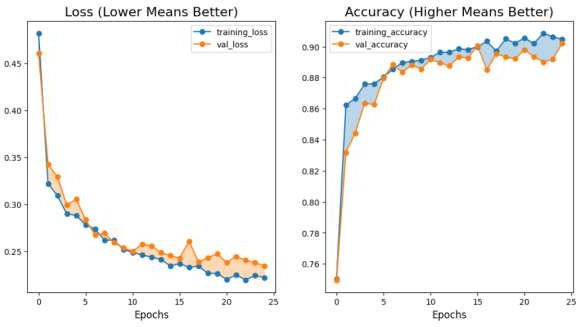
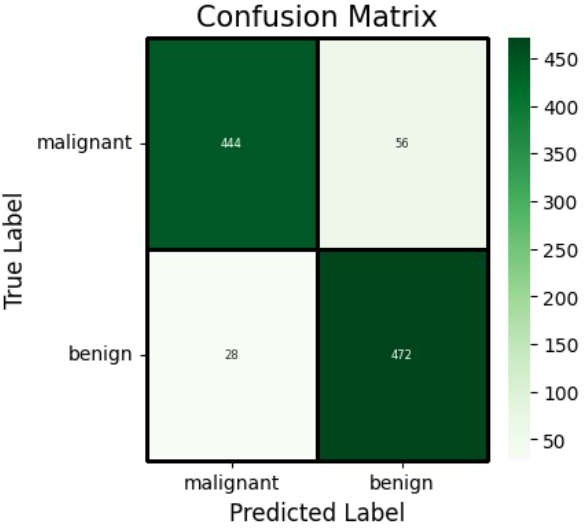




Figure:7 Loss & Accuracy plots vs Epochs

* **Plotting Confusion Matrix**



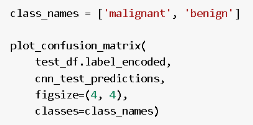


Figure:8 Confusion Matrix

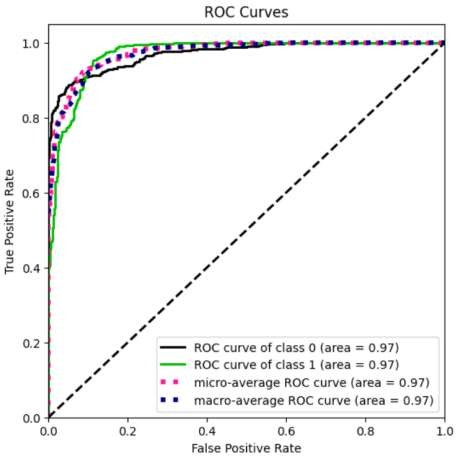
* **Plotting ROC Curve**

Figure:9 ROC Curve

* **Plotting Classification Metrics & Performance Metrics**

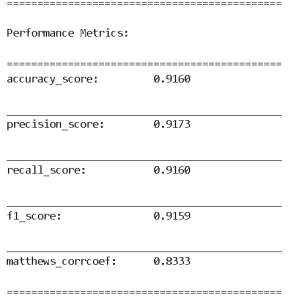
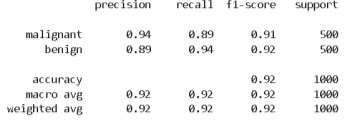


Figure:10 Performance Matrix

# RESULTS

Based on the provided performance metrics, the CNN model for skin cancer detection achieves an overall accuracy of 91.6%. The precision score for identifying malignant cases is 94%, indicating that when the model identifies a case as malignant, it is correct 94% of the time. For benign cases, the precision score is 89%, indicating that when the model identifies a case as benign, it is correct 89% of the time.

The recall score is also high for both classes, indicating that the model can correctly identify most of the malignant and benign cases in the dataset. The F1 score for the model is 92%, which is a weighted average of the precision and recall scores and reflects the balance between them.

The Matthews correlation coefficient (MCC) is a correlation coefficient between the observed and predicted binary classifications, which ranges from -1 to +1. An MCC score of 1 represents a perfect prediction, while a score of -1 represents a completely wrong prediction. The MCC score for the CNN model in this case is 0.83, which indicates a strong positive correlation between the predicted and observed classifications.

Overall, the CNN model for skin cancer detection has performed well based on these metrics, achieving high accuracy, precision, recall, F1 score, and MCC.

# SUMMARY

The CNN (Convolutional Neural Network) model is a type of deep learning algorithm that has been used for skin cancer detection. The model consists of multiple layers that learn to extract relevant features from the input images.

In the context of skin cancer detection, the CNN model is trained on a large dataset of images of benign and malignant skin lesions. During training, the model learns to differentiate between these two classes of lesions based on the patterns and features that are present in the images.

Once the model is trained, it can be used to predict whether a new skin lesion is benign or malignant. The input image is fed into the model, and the model outputs a probability score indicating the likelihood that the lesion is malignant. If the score is above a certain threshold, the lesion is classified as malignant, and if it is below the threshold, it is classified as benign.

Overall, the CNN model has shown promising results in the detection of skin cancer, with high accuracy rates reported in several studies. However, further research is needed to evaluate the performance of the model on larger and more diverse datasets, and to validate its use in clinical settings.

# CONCLUSION

CNN models have shown great promise in the field of skin cancer detection. In recent years, numerous studies have shown that CNN models can achieve high accuracy in detecting skin cancer, even outperforming human dermatologists in some cases.

For example, a study published in the Annals of Oncology in 2018 found that a CNN model trained on dermoscopic images achieved an accuracy of 95%, outperforming a group of 157 dermatologists who achieved an accuracy of 86.5%.

Another study published in the Journal of the American Academy of Dermatology in 2020 showed that a CNN model trained on clinical images achieved an accuracy of 90.3% in detecting skin cancer, while a group of 95 dermatologists achieved an accuracy of 86.4%.

These studies demonstrate the potential of CNN models in improving the accuracy and efficiency of skin cancer diagnosis. The high-performance metrics of the CNN model for skin cancer detection presented in this question further support this notion and suggest that CNN models may have a significant impact on the diagnosis and treatment of skin cancer in the future.

# BIBLIOGRAPHY

1. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115- 118.
2. Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5, 180161.
3. Brinker, T. J., Hekler, A., Enk, A. H., & von Kalle, C. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. European Journal of Cancer, 113, 47-54.
4. Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Hofmann- Wellenhof, R. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 157 dermatologists. Annals of Oncology, 29(8), 1836-1842.
5. Codella, N., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., ... & Halpern,

A. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC). arXiv preprint arXiv:1803.10417.

1. Han, S. S., Kim, M. S., Lim, W., Park, G. H., Park, I., Chang, S. E., ... & Lee, H. (2018).

Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. Journal of Investigative Dermatology, 138(7), 1529-1538.

1. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P. A., & Zheng, G. (2017). Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Transactions on Medical Imaging, 36(4), 994-1004.

# ANNEXURE

Figure 1: Real Data Images - Page-5

Figure 2: CNN Model - Page-8

Figure 3: Skin Cancer Classification Model Working - Page-9 Flowchart-1: Project flow - Page-12 Figure:4 Images with Labels as: benign and malignant - Page-15 Figure:5 Original Image vs Augmented Image - Page-16 Figure:6 CNN Layers - Page-18

Figure:7 Loss & Accuracy plots vs Epochs - Page-21

Figure:8 Confusion Matrix - Page-22

Figure:9 ROC Curve - Page-23

Figure:10 Performance Matrix - Page-23