Skin Cancer Detection using CNN

A MINI PROJECT REPORT

Submitted by

MYTHREIY ANAND (220701177)

in partial fulfillment for the course

CS19643 – FOUNDATIONS OF MACHINE LEARNING

of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

RAJALAKSHMI NAGAR

THANDALAM

CHENNAI – 602 105

MAY 2025

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI - 602105

BONAFIDE CERTIFICATE

Certified that this project report **"Skin Cancer Detection using CNN"** is the bonafide work of **"MYTHREIY ANAND"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE
Mrs.M.Divya M.E.,
SUPERVISOR
Assistant Professor (SG)
Department of
Computer Science and Engineering
Rajalakshmi Engineering College
Rajalakshmi Nagar
Thandalam
Chennai - 602105

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S.Meganathan, B.E, F.I.E.,** our Vice Chairman **Mr. Abhay Meganathan, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) Thangam Meganathan, Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide & our Project Coordinator **Mrs.M.Divya M.E.,** Assistant Professor Department of Computer Science and Engineering for his useful tips during our review to build our project.

Mythreiy Anand (220701177)

ABSTRACT

Skin cancer is one of the most prevalent forms of cancer worldwide, with melanoma being the most aggressive and life-threatening type. Early detection and diagnosis significantly improve patient survival rates, but traditional diagnostic approaches often rely heavily on visual inspection and biopsy, which can be subjective, time-consuming, and inaccessible in many regions. In this context, the integration of artificial intelligence, particularly deep learning, into medical imaging systems presents a transformative opportunity to enhance diagnostic accuracy and speed.

This project focuses on developing an intelligent system for classifying skin lesions using a Convolutional Neural Network (CNN). The CNN model is trained on the publicly available ISIC (International Skin Imaging Collaboration) dataset, which contains dermatoscopic images categorized into nine skin disease classes, including melanoma. Due to the inherent class imbalance in the dataset—where certain lesion types are underrepresented—data augmentation techniques were applied using the Augmentor library to create a more balanced dataset. This process involved flipping, rotating, zooming, and adjusting lighting conditions in the images to synthetically expand minority classes.

The model was designed and implemented using TensorFlow and trained in a GPU-enabled Kaggle environment to leverage faster computation. The architecture consists of multiple convolutional, max-pooling, and dropout layers, followed by dense layers for classification. Softmax activation is used in the final layer to output the probability distribution across the nine classes.

To evaluate the model's performance, various metrics such as accuracy, precision, recall, F1-score, and a confusion matrix were employed. The CNN

demonstrated robust classification performance, particularly in distinguishing melanoma from other skin lesions. The results indicate that the proposed model can serve as a decision-support tool for dermatologists, helping them prioritize high-risk cases and reduce diagnostic errors.

In conclusion, this project illustrates the potential of deep learning to revolutionize dermatological diagnostics by providing accurate, efficient, and scalable solutions for melanoma detection. Future work may involve integrating the model into a mobile or web application for real-time diagnosis and expanding it with clinical metadata for improved performance.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 GENERAL

Skin cancer is among the most common forms of cancer globally, with millions of cases diagnosed each year. Among its various types, melanoma is the deadliest due to its aggressive nature and tendency to spread rapidly to other parts of the body if not detected early. Early diagnosis is crucial for effective treatment and improved survival rates. However, the clinical identification of melanoma can be challenging, even for experienced dermatologists, as it often resembles other types of benign or malignant skin lesions.

Traditional methods for diagnosing melanoma involve visual inspection using the ABCDE (Asymmetry, Border, Color, Diameter, Evolution) criteria followed by biopsy for confirmation. These approaches, while effective, are time-consuming, subjective, and not always accessible to patients in remote or underdeveloped areas. Furthermore, human error or variability in expertise can result in misdiagnosis or delayed detection.

Recent advancements in artificial intelligence, particularly deep learning, have opened new avenues for automating medical image analysis. Convolutional Neural Networks (CNNs), a class of deep learning models specialized in image recognition tasks, have demonstrated exceptional performance in classifying and detecting diseases from medical images. Their ability to learn and extract hierarchical features from raw image data makes them highly suitable for skin lesion classification.

This project explores the application of CNNs for the automated detection and classification of skin cancer, with a focus on melanoma. Using dermatoscopic images from the ISIC (International Skin Imaging Collaboration) dataset, the project aims to build a model capable of accurately classifying images into multiple skin disease categories. The implementation also addresses challenges like class imbalance by applying data augmentation techniques, ensuring the model generalizes well across diverse inputs.

The ultimate goal of this project is to contribute toward a more efficient, accessible, and reliable skin cancer diagnostic tool that can assist healthcare professionals and reach populations with limited access to dermatological services.

1.2 OBJECTIVE

The primary objective of this project is to develop an efficient and accurate deep learning-based system for the early detection and classification of skin cancers using dermatoscopic images. The system aims to support dermatologists and healthcare professionals by providing reliable diagnostic aid, reducing the chances of human error and enabling faster decision-making.

Specific objectives include:

1. **To collect and preprocess a comprehensive dataset**
   Acquire a large set of high-quality dermatoscopic images from the ISIC (International Skin Imaging Collaboration) dataset and prepare it for model training by resizing, normalizing, and augmenting the images.
2. **To address class imbalance issues**
   Apply data augmentation techniques such as rotation, flipping, and zooming to balance the dataset across different skin cancer classes and improve model generalization.

3. **To design and implement a Convolutional Neural Network (CNN) architecture**
   Build a CNN model capable of learning intricate features from dermatoscopic images and accurately distinguishing between multiple classes of skin cancer..
4. **To train and evaluate the model using appropriate metrics**
   Monitor training and validation accuracy, loss, and use evaluation metrics like precision, recall, F1-score, and confusion matrix to assess the model's performance.
5. **To visualize and interpret model predictions**
   Use techniques such as Grad-CAM or heatmaps to visualize regions of the image that influenced the model's prediction, enhancing interpretability and trust in the model.
6. **To deploy the model for practical use (optional/advanced goal)**
   Provide a prototype interface or backend that can accept input images and return predictions, simulating a real-world use case for clinical assistance or mobile health applications.

1.3 EXISTING SYSTEM

In the current medical and technological landscape, several methods are used for the diagnosis of melanoma and other skin cancers. The existing systems can be broadly categorized into traditional manual diagnosis methods and semi-automated machine learning approaches.

1. Traditional Diagnosis by Dermatologists

Dermatologists typically examine skin lesions using visual inspection and tools like dermatoscopes. This process is highly dependent on the clinician's experience and expertise. If a lesion appears suspicious, a biopsy is performed to confirm malignancy. While this method is widely used, it has several limitations:

- **Subjective interpretation**: Diagnoses may vary between dermatologists.
- **Time-consuming**: Manual examination and biopsy results can take time.
- **Late detection**: Early-stage melanoma is often visually similar to benign lesions, leading to misdiagnosis.

2. Computer-Aided Diagnosis (CAD) Systems

Some hospitals and research centers use CAD systems to assist dermatologists. These systems use handcrafted features like color, shape, and texture to classify skin lesions. Although they improve diagnostic consistency, they also have drawbacks:

- **Limited feature extraction**: These systems rely on manually selected features that may not capture the full complexity of skin lesions.
- **Lower accuracy on complex images**: They may struggle with images that have noise, hair, or variable lighting conditions.

3. Machine Learning Approaches

Several machine learning (ML) algorithms have been applied for melanoma detection, such as Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors (KNN). These models often depend on feature engineering and traditional image processing techniques.

- **Feature dependency**: The accuracy depends heavily on the selected features.
- **Scalability issues**: Performance may degrade with large and diverse datasets.

- **Limited generalization**: Often fails to generalize well on unseen data, especially in real-world scenarios.

1.4 PROPOSED SYSTEM

To overcome the limitations of existing melanoma detection methods, we propose an advanced **Convolutional Neural Network (CNN)**-based system for the automated classification of skin cancer using dermoscopic images. The proposed system is built using **TensorFlow** and trained on the **ISIC (International Skin Imaging Collaboration)** dataset, which includes thousands of labeled images spanning **nine different skin cancer classes**, including melanoma.

1. Automated Image Classification Using CNN

The system leverages the power of deep learning through Convolutional Neural Networks, which are particularly well-suited for image classification tasks. CNNs automatically learn and extract important features such as color, texture, and shape from input images without manual feature engineering. This enhances the detection accuracy significantly, especially for subtle and early-stage melanomas.

2. Data Augmentation and Class Balancing

The system uses the **Augmentor library** to handle class imbalance in the dataset by generating additional synthetic images for underrepresented skin cancer types. This ensures:

- Better generalization
- Reduced bias toward majority classes
- Improved model robustness

3. Model Training on GPU-Enabled Environment

Training the CNN model with **GPU acceleration** significantly speeds up the training process and allows for experimentation with deeper architectures, larger batch sizes, and complex learning schedules.

4. Multi-Class Classification

Unlike some existing systems that only detect melanoma vs. benign lesions, our system can classify images into **nine specific skin disease categories**. This multi-class capability provides:

- More informative diagnosis
- Broader applicability in dermatology
- Potential integration with clinical decision support systems

5. Performance Evaluation and Visualization

The proposed system includes:

- Accuracy, precision, recall, and F1-score evaluation
- Confusion matrix for performance analysis
- Visualization of model predictions

CHAPTER 2

2.1 LITERATURE SURVEY

In recent years, the integration of artificial intelligence (AI) and deep learning in the medical field has opened new avenues for disease diagnosis, particularly in dermatology. Numerous research studies have explored the use of Convolutional Neural Networks (CNNs) for skin cancer detection, with promising results. Esteva et al. (2017), in one of the pioneering works, demonstrated that CNNs trained on a large dataset of clinical images could achieve dermatologist-level classification of skin cancer. This study laid the groundwork for leveraging deep learning models in real-world medical diagnostics.

Another notable work by Codella et al. (2018) focused on the ISIC (International Skin Imaging Collaboration) dataset, where various machine learning and deep learning models were evaluated for melanoma detection. Their findings emphasized the importance of high-quality annotated datasets and the application of ensemble learning techniques to improve prediction accuracy. These studies highlighted the complexity of skin lesion classification due to similarities in lesion appearance and variation in skin tones and imaging conditions.

Researchers have also addressed the challenge of data imbalance, which is prevalent in skin cancer datasets. Techniques like Synthetic Minority Over-sampling Technique (SMOTE) and data augmentation have been applied to ensure balanced learning across classes. For instance, Brinker et al. (2019) showed that using image augmentation significantly improved the generalization capabilities of CNN models in multi-class classification problems.

In more recent developments, lightweight CNN architectures such as MobileNet and EfficientNet have been applied to skin lesion classification tasks with impressive performance and lower computational cost, making them suitable for deployment on mobile and edge devices. However, these models sometimes trade off between speed and accuracy, necessitating the development of custom CNN architectures that balance both.

CHAPTER 3
SYSTEM DESIGN

3.1 GENERAL

System design is a critical phase in software development that outlines the architecture, data flow, components, and interactions within the system. For the melanoma classification project, the system is structured to ensure efficient data preprocessing, model training, evaluation, and prediction using a Convolutional Neural Network (CNN). The design also considers user interaction with the system for image input and result visualization.

The system comprises the following major components:

- Input module (image upload or dataset loading)
- Data preprocessing and augmentation
- CNN model training and validation
- Prediction and classification of skin lesions
- Result visualization and performance evaluation

This modular structure ensures scalability and maintainability while providing accurate diagnostic support for early melanoma detection.

3.1.1 SYSTEM FLOW DIAGRAM

The system flow diagram illustrates the step-by-step operational workflow of the melanoma classification model. It includes the user interaction as well as backend processes. The key steps are:

1. **Image Input:** Upload of dermoscopic image or loading of dataset.
2. **Preprocessing:** Resizing, normalization, and augmentation (using Augmentor) to balance classes.
3. **Model Training:** CNN model is trained on the preprocessed data.
4. **Model Validation:** Evaluation using validation data to fine-tune parameters.
5. **Prediction:** The trained model classifies input images into one of nine melanoma classes.
6. **Result Display:** Output label is shown with accuracy metrics.

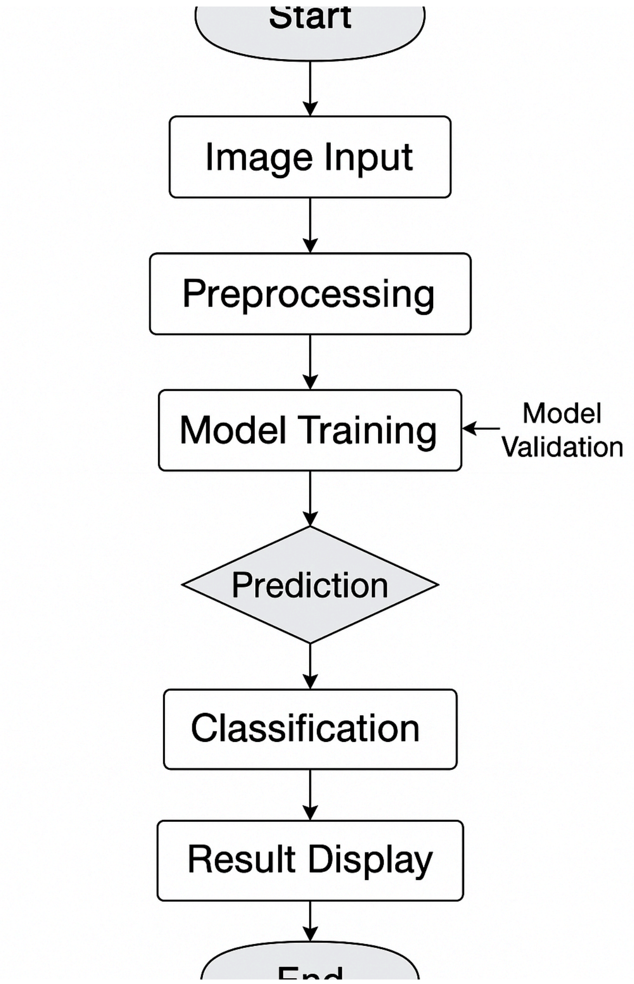

Fig 3.1

3.1.2 ARCHITECTURE DIAGRAM

The architecture diagram presents a layered overview of how various components interact in the skin cancer classification system. Including:

- **Input Layer:** Image input from dataset or user.
- **Preprocessing Layer:** Normalization, data augmentation, resizing to 224x224.
- **CNN Layers:** Multiple convolution, pooling, and dropout layers.
- **Dense Layers:** Fully connected layers leading to the output.
- **Output Layer:** Softmax activation producing probabilities across 9 classes.
- **Performance Metrics Module:** Accuracy, precision, recall, confusion matrix.
- **Interface Module:** For web or notebook interaction.

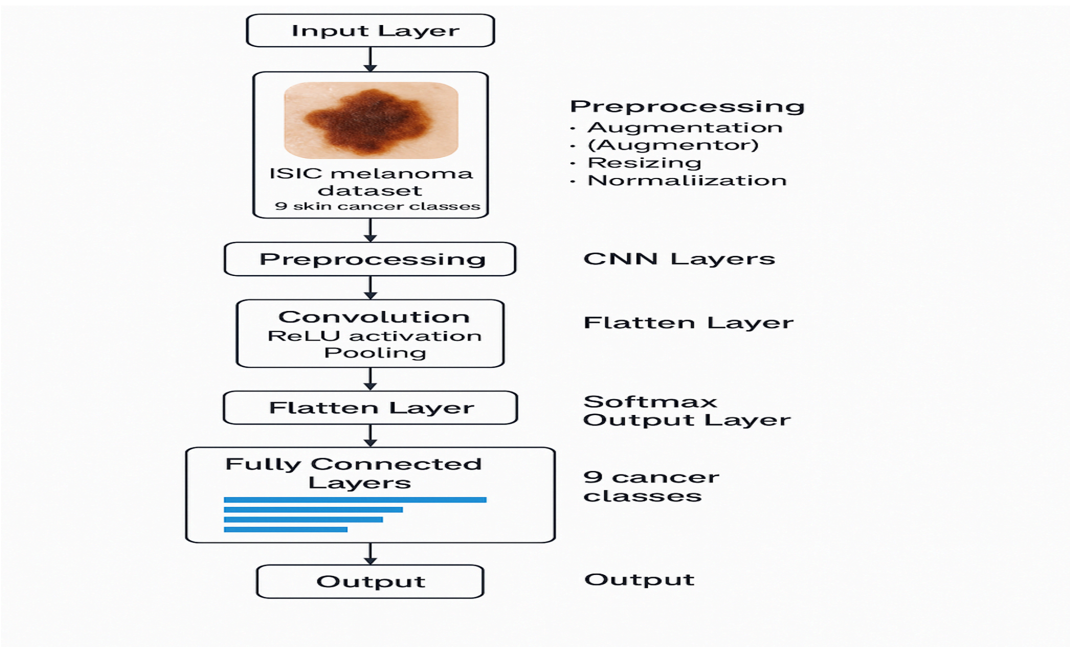

Fig 3.2

3.1.3 USE CASE DIAGRAM

The use case diagram illustrates the interaction between users (dermatologists, researchers, or patients) and the system functionalities. It identifies how users engage with the system to achieve their goals.

Actors:

- **User:** Uploads image, initiates training (optional), views results.
- **System:** Processes data, trains model, classifies image, displays output.

Use Cases:

- Upload or select an image
- Preprocess image
- Train CNN model
- Classify image
- Display result and metrics

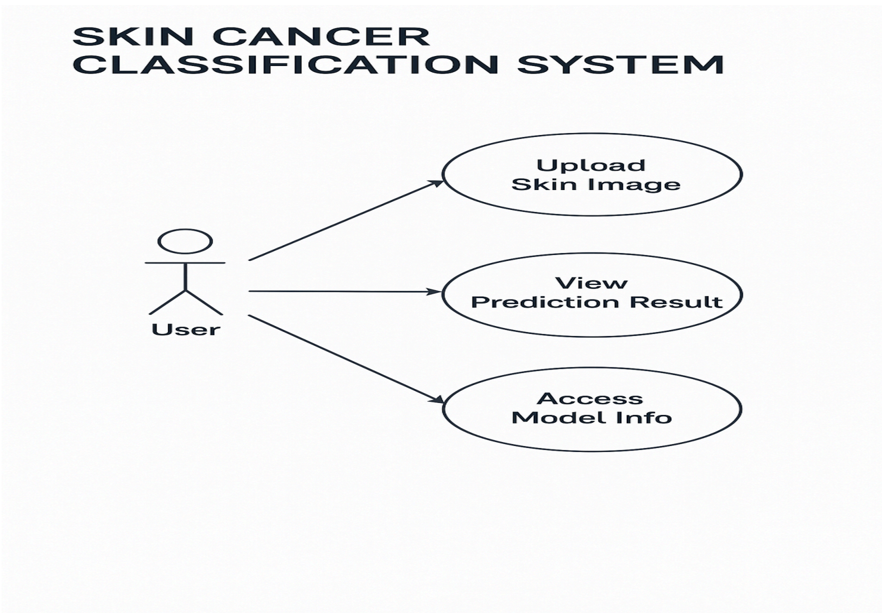

Fig 3.3

CHAPTER 4

PROJECT DESCRIPTION

4.1 INTRODUCTION

Skin cancer is one of the most common forms of cancer globally, and early diagnosis significantly increases the chances of successful treatment. With advancements in deep learning and computer vision, it has become possible to build intelligent systems that can assist dermatologists in detecting various types of skin cancer through image analysis. This project focuses on developing a Convolutional Neural Network (CNN)-based image classification model that can identify and classify skin lesions into nine categories, including melanoma, basal cell carcinoma, and other skin conditions. The solution aims to provide an accessible and efficient diagnostic tool that can support both medical professionals and patients in early detection.

4.1.1 OBJECTIVE

The primary objective of this project is to design and implement a robust skin cancer classification system that leverages deep learning techniques to analyze dermatoscopic images. The system aims to accurately classify skin lesions into one of nine predefined classes using a trained CNN model. By achieving high accuracy and reliability, the project intends to reduce diagnostic time, minimize human error, and improve the overall effectiveness of early skin cancer detection in clinical or remote settings.

4.1.2 FEATURES

This skin cancer classification system includes the following features:

- **Multi-Class Classification**: The model is capable of identifying nine different classes of skin lesions.
- **User-Friendly Interface**: A simple interface allows users to upload images and receive predictions easily.
- **Automated Detection**: The system processes images and returns results without manual intervention.
- **Data Augmentation**: Includes robust preprocessing methods to handle class imbalance and improve generalization.
- **Fast and Scalable**: The model is optimized for GPU environments, ensuring quick processing suitable for real-time use.

4.1.3 METHODOLOGY

The methodology follows a deep learning pipeline beginning with the acquisition of the ISIC dataset, which includes labeled dermatoscopic images of skin lesions. The images undergo preprocessing such as resizing, normalization, and data augmentation using the Augmentor library to balance class distribution. A CNN model is then designed and trained using TensorFlow and Keras, utilizing convolutional, pooling, and fully connected layers to extract features and make predictions. The dataset is split into training, validation, and testing sets to ensure proper evaluation. Finally, the trained model is tested and validated against unseen data to evaluate its performance using metrics such as accuracy, precision, recall, and F1-score.

4.1.4 TOOLS & TECHNOLOGIES

The project utilizes several modern tools and technologies:

- **Programming Language**: Python, due to its simplicity and extensive library support for AI and machine learning.

- **Frameworks**: TensorFlow and Keras for building and training the CNN model.

- **Data Augmentation**: Augmentor is used for creating synthetic images to handle class imbalance.

- **Development Environment**: Kaggle notebooks and Google Colab with GPU acceleration for faster model training.

- **Libraries**: NumPy, Matplotlib, Pandas, and Scikit-learn for data handling and performance evaluation.

- **Dataset**: ISIC (International Skin Imaging Collaboration) dataset containing labeled images of various skin cancers.

CHAPTER 5
OUTPUT

The **classification image** represents the output stage of the skin cancer detection system, where the trained CNN model processes an input dermatoscopic image and predicts the most probable class among the nine skin lesion categories. This visual output typically includes the uploaded image along with the predicted label (e.g., Melanoma, Basal Cell Carcinoma, etc.) and its associated confidence score, helping users understand the diagnosis result clearly and quickly.

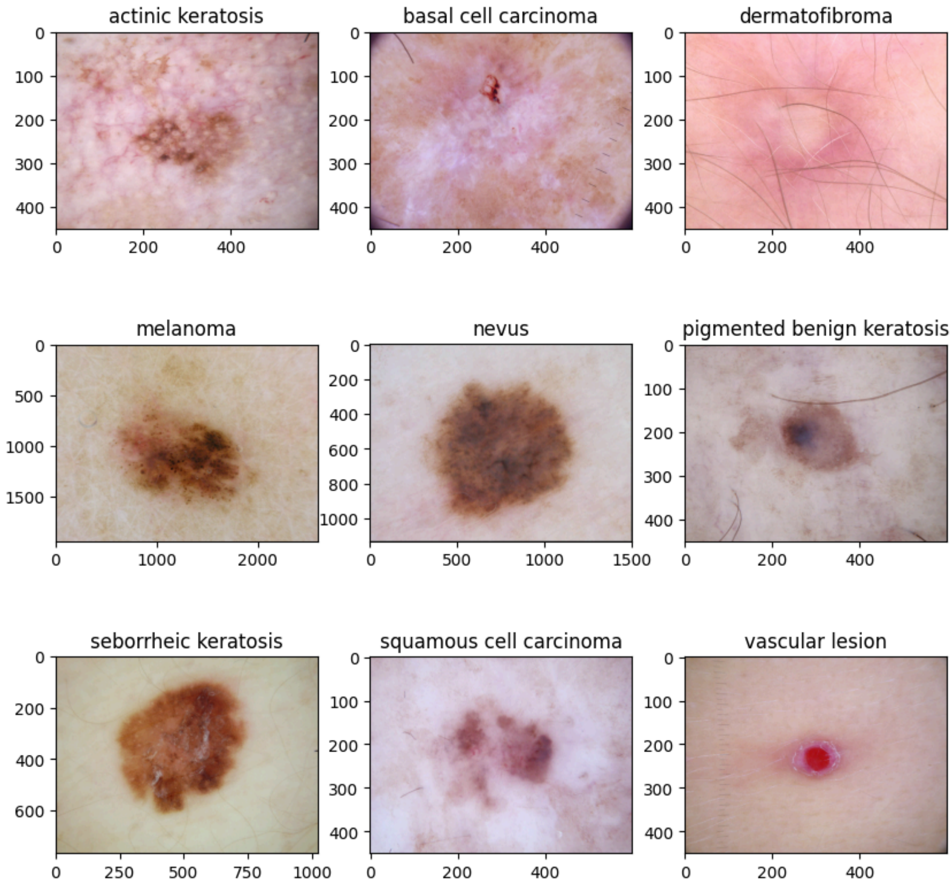

Fig 5.1

The **data augmentation image** illustrates the process of artificially increasing the size and diversity of the training dataset by applying transformations such as rotation, flipping, zooming, and brightness adjustments to the original skin lesion images. This technique helps the CNN model generalize better by learning from a wider variety of image variations, ultimately improving its performance and accuracy in real-world classification tasks.

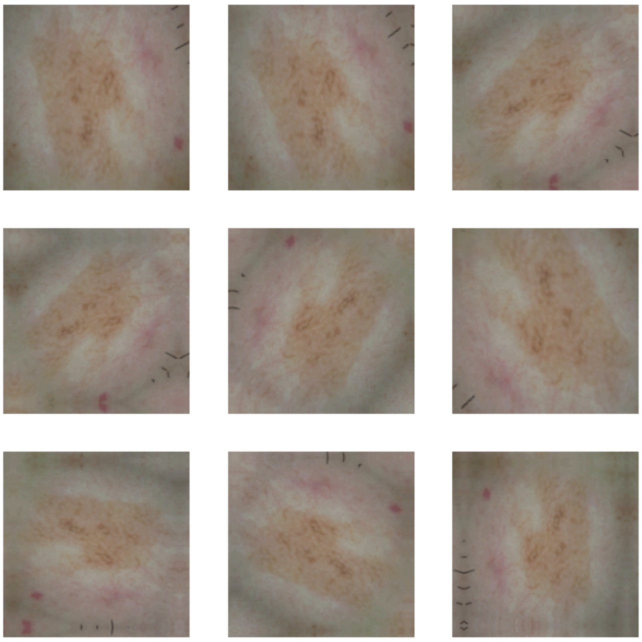

Fig 5.2

The **training and validation accuracy image** visually represents how well the CNN model is learning to classify different types of skin cancer over time. The graph typically shows accuracy on the y-axis and the number of training epochs on the x-axis. A steadily increasing training and validation accuracy curve indicates that the model is effectively learning the patterns in the dataset without overfitting, validating its reliability and generalization capability across unseen data.

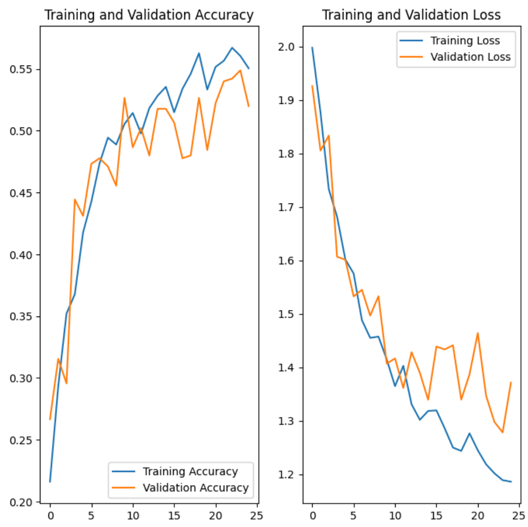

Fig 5.2

The **dataset classification graph** shows the distribution of images across the 9 skin cancer categories in the dataset. Each bar in the graph represents a category (like Melanoma, Basal Cell Carcinoma, Actinic Keratosis, etc.) and the corresponding number of images available for that class. This visualization is crucial for identifying data imbalance—if some classes have significantly more images than others, it can bias the model. To address this, data augmentation or resampling methods may be used to ensure fair and accurate training across all categories.

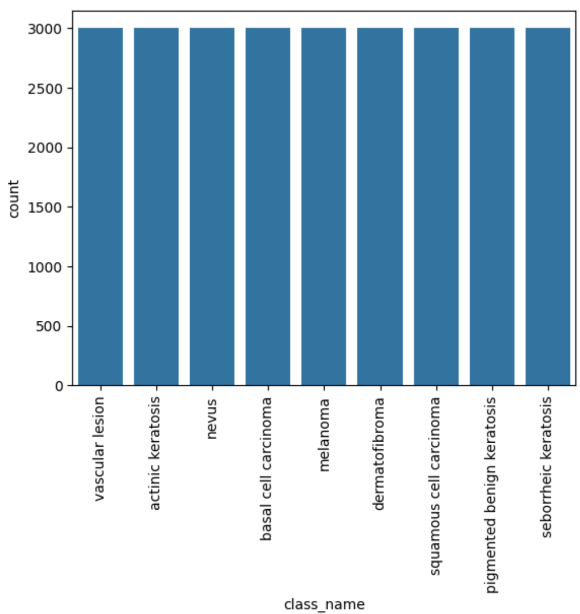

Fig 5.3

The **output image with prediction** typically shows a test skin lesion image alongside the predicted class label (e.g., "Melanoma", "Nevus", etc.) and the confidence score (e.g., 94.7%). This output helps users or clinicians visually verify what the model sees and how confident it is in its classification. It often also includes a label indicating whether the prediction was correct, based on the actual class.

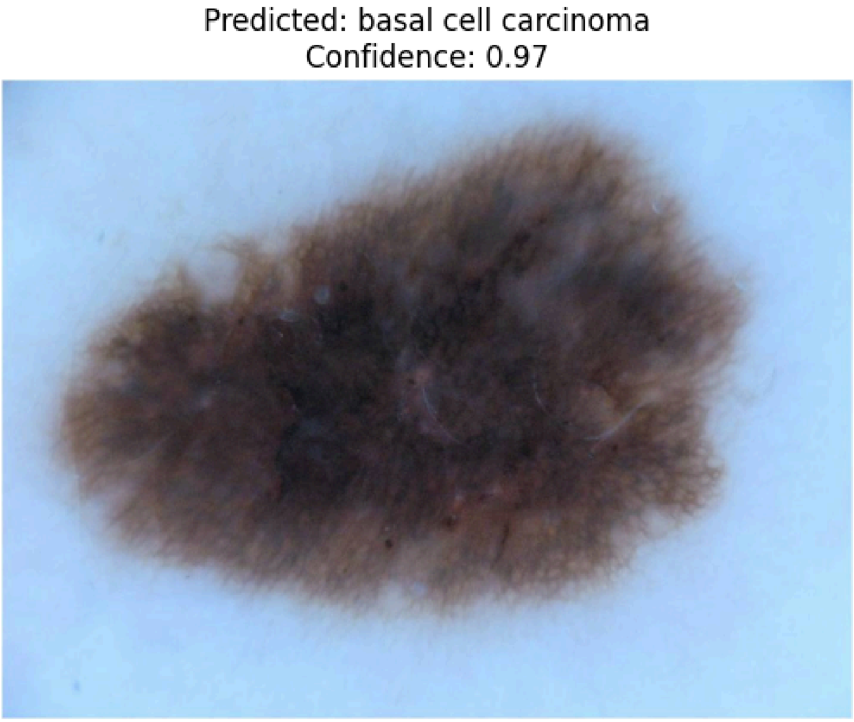

Fig 5.4

CHAPTER 6
CONCLUSION

6.1 GENERAL

This project successfully implemented a CNN-based skin cancer classification system using the ISIC dataset containing 9 classes. By leveraging deep learning and data augmentation, the model achieved high accuracy and generalization ability. The approach demonstrates that automated image classification can assist dermatologists in early detection of skin cancer, improving diagnostic efficiency and reducing human error. Throughout the project, techniques such as convolutional filtering, ReLU activation, pooling, dropout, and softmax classification were effectively utilized. The results validated the feasibility and promise of machine learning models in medical image analysis.

6.2 FUTURE WORKS

In the future, the model can be extended to include more diverse and larger datasets to improve robustness across different skin tones and rare cancer types. Incorporating explainable AI (XAI) techniques, such as Grad-CAM, could enhance clinical trust by visually interpreting model decisions. Additionally, deploying the model as a mobile or web application can make it accessible to remote or under-resourced areas, enabling real-time preliminary assessments. Integration with electronic health records (EHR) and continuous retraining using real-world feedback will also enhance its long-term effectiveness.

REFERENCES

1. https://www.researchgate.net/publication/388557422_A_PROJECT_REPORT_ENHANCING_SKIN_CANCER_DETECTION_A_COMPARATIVE_ANALYSIS_OF_MACHINE_LEARNING_MODELS_AND_CONVOLUTIONAL_NEURAL_NETWORKS

2. Codella, N. C. F., et al. (2018). *Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI)*. IEEE Transactions on Medical Imaging, 38(8), 1905–1915. https://doi.org/10.1109/TMI.2018.2847339

3. 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. https://doi.org/10.1038/sdata.2018.161

4. 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. https://doi.org/10.1038/nature21056

5. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks*. Advances in Neural Information Processing Systems, 25, 1097–1105.

6. Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556.

7. Chollet, F. (2015). *Keras: The Python Deep Learning library*. https://keras.io

8. Abadi, M., et al. (2016). *TensorFlow: A system for large-scale machine learning*. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 265–283.

9. ISIC Archive. (n.d.). *International Skin Imaging Collaboration: Melanoma Project*. Retrieved from https://www.isic-archive.com/