GROUP ID: GE13

A PROJECT REPORT ON

ENHANCED MELANOMA SKIN CANCER DETECTION THROUGH ENSEMBLE OF THEPADE'S SBTC AND GLCM FEATURES USING MACHINE LEARNING

SUBMITTED TO THE PIMPRI CHINCHWAD COLLEGE OF ENGINEERING
AN AUTONOMOUS INSTITUTE, PUNE
IN THE FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF TECHNOLOGY
COMPUTER ENGINEERING (REGIONAL LANGUAGE)

SUBMITTED BY

DEEP NAVALE	121B1D015
PANKAJ KANAME	121B1D016
OM MAHAJAN	121B1D034
ROHAN PATIL	121B1D036

DEPARTMENT OF COMPUTER ENGINEERING (REGIONAL LANGUAGE)

PCET'S PIMPRI CHINCHWAD COLLEGE OF ENGINEERING

Sector No. 26, Pradhikaran, Nigdi, Pimpri-Chinchwad, PUNE 411044

2024-2025

CERTIFICATE

This is to certify that the project report entitles

"ENHANCED MELANOMA SKIN CANCER DETECTION THROUGH ENSEMBLE OF THEPADE'S SBTC AND GLCM FEATURES USING MACHINE LEARNING"

Submitted by

DEEP NAVALE	**121B1D015**
PANKAJ KANAME	**121B1D016**
OM MAHAJAN	**121B1D034**
ROHAN PATIL	**121B1D036**

is a bonafide student of this institute and the work has been carried out by him under the supervision of **Prof. Minal Shahakar** and it is approved for the partial fulfillment of the requirement of Pimpri Chinchwad College of Engineering an autonomous institute, for the award of the B. Tech. degree in Computer Engineering.

(Prof. Minal Shahakar)
Guide
Department of Computer Engineering
(Regional Language)

(Prof. Dr. Rachana Y. Patil)
Head,
Department of Computer Engineering
(Regional Language)

(Prof. Dr. G.N. Kulkarni)
Director,
Pimpri Chinchwad College of Engineering Pune – 411044

Place: Pune
Date:

ACKNOWLEDGEMENT

We express our sincere gratitude to our **Guide, Prof. Minal Shahakar** , for their constant encouragement and invaluable support throughout our project. Their insightful suggestions and guidance have played a crucial role in shaping the success of this work. We would also like to extend our heartfelt thanks to our **Project Coordinator, Prof. Rohini Y. Sarode**, for their unwavering assistance and direction from the initial stages of the project. Our sincere appreciation goes to **Prof. Dr. Rachana Y. Patili, Head of Computer Engineering (Regional Language) Department**, for their continuous support and motivation throughout our project journey.

We are deeply thankful to our **Director, Prof. Dr. G.N. Kulkarni**, for providing us with the necessary resources and a conducive environment to successfully complete our project. We also extend our gratitude to all faculty members, staff, and technicians of our college for their valuable assistance in making this project a success.

Furthermore, we sincerely appreciate the contributions of all web committees for enriching us with their vast knowledge. Lastly, we take this opportunity to express our deep appreciation to our family and friends for their constant encouragement and unwavering support during this crucial phase of our project.

DEEP NAVALE
PANKAJ KANAME
OM MAHAJAN
ROHAN PATIL

ABSTRACT

Melanoma is one of the most aggressive forms of skin cancer, often caused by prolonged sun exposure and genetic factors. Early detection is critical, as melanoma spreads rapidly and can become life-threatening if not treated promptly. Traditional diagnostic methods rely heavily on visual examination, which can be subjective and prone to errors. Therefore, there is a strong need for advanced and automated melanoma detection techniques to improve accuracy and reliability.

This study proposes an enhanced melanoma detection approach by combining Thepade's Sorted Block Truncation Coding (SBTC) and Gray Level Co-occurrence Matrix (GLCM) for feature extraction. The extracted features are classified using an ensemble of machine learning models, including Multilayer Perceptron and Random Forest. The HAM-10000 dataset was used for evaluation, and experimental results show that the fusion of Thepade's SBTC and GLCM achieved an accuracy of 88.46%, outperforming individual methods.

The proposed method demonstrates the potential for improving melanoma detection through machine learning, providing a more reliable alternative to conventional diagnostic techniques. Future work can focus on integrating deep learning models and expanding the dataset for further performance enhancement.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABBREVIATION	ILLUSTRATION
SBTC	Sorted Block Truncation Coding
GLCM	Gray Level Co-occurrence Matrix
MLP	Multilayer Perceptron
RF	Random Forest
HAM-10000	Human Against Machine with 10000 Images Dataset
CNN	Convolutional Neural Network
YOLO	You Only Look Once
SVM	Support Vector Machine
PCA	Principal Component Analysis
GANs	Generative Adversarial Networks
CSV	Comma-Separated Values
SQL	Structured Query Language
MySQL	Relational Database Management System

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Melanoma is one of the most aggressive and life-threatening forms of skin cancer, known for its rapid spread if not detected in its early stages. The increasing prevalence of melanoma, especially among young individuals, highlights the urgent need for more efficient and accurate detection methods. Traditional diagnosis relies heavily on visual inspection by dermatologists, which can be subjective and prone to errors. Even with dermatoscopic tools, diagnostic accuracy largely depends on the clinician's experience and training. Therefore, an automated and precise detection system using machine learning can significantly aid in early diagnosis, improving patient survival rates. Such systems can reduce diagnostic delays, assist in triaging suspicious lesions, and offer decision support in remote or resource-constrained settings.

This project proposes an advanced melanoma detection system that leverages Thepade's Sorted Block Truncation Coding (SBTC) and Gray Level Co-occurrence Matrix (GLCM) for feature extraction. SBTC extracts color-based features from Red, Green, and Blue (RGB) color planes, while GLCM extracts texture-based features such as contrast, correlation, energy, and homogeneity. These features are then used to classify skin lesions as benign or malignant using machine learning classifiers. The integration of both color and texture dimensions ensures a more comprehensive representation of lesion characteristics, improving model performance.

The study utilizes the HAM-10000 dataset, which contains a significant number of labeled skin lesion images. Prior to training, preprocessing techniques like artifact removal and normalization are applied to enhance image quality. Various machine learning classifiers, including Multilayer Perceptron, Random Forest, Naïve Bayes, and others, are tested for classification accuracy. Results indicate that while individual classifiers perform well, an ensemble approach combining Multilayer Perceptron and Random Forest achieves the highest accuracy of 88.46%. This highlights the potential of combining neural and tree-based models for robust classification.

The proposed system demonstrates that combining SBTC and GLCM feature extraction techniques enhances melanoma detection accuracy, outperforming traditional methods. This research provides a promising foundation for developing AI-driven melanoma detection systems, which can assist medical professionals in early diagnosis and treatment planning. It also opens avenues for integrating the model into mobile applications for accessible skin screening. Future improvements may include deep learning models and larger datasets to further refine the classification process. Additionally, real-time deployment in clinical settings can be explored to evaluate its performance in real-world scenarios.

1.1 OVERVIEW

This project enhances melanoma detection using Thepade's SBTC and GLCM for feature extraction. The HAM-10000 dataset is used to classify skin lesions as benign or malignant with machine learning. Various classifiers are tested, with Multilayer Perceptron (87.51%) and Random Forest (86.33%) showing high accuracy individually. Combining SBTC and GLCM with an ensemble of MLP and Random Forest achieves 88.46% accuracy, proving the effectiveness of the proposed method. This approach improves early melanoma detection, aiding in timely diagnosis and treatment.

1.2 MOTIVATION

Melanoma is one of the deadliest forms of skin cancer, known for its rapid spread if not detected early. Early diagnosis significantly increases the chances of successful treatment, but visual identification remains challenging due to variations in appearance. Traditional methods rely on dermatologists expertise, which can be subjective and time-consuming. This highlights the need for an automated, accurate, and efficient melanoma detection system that can assist in early diagnosis and reduce mortality rates.

Recent advancements in machine learning and image processing have shown great potential in medical diagnostics. By leveraging computational techniques, we can improve the accuracy and efficiency of melanoma detection. Feature extraction methods like Thepade's SBTC and GLCM provide valuable color and texture-based information, which, when combined with machine learning classifiers, can significantly enhance classification performance.

The motivation behind this project is to develop a robust melanoma detection system that integrates SBTC and GLCM features with machine learning to improve classification accuracy. By using the HAM-10000 dataset and evaluating multiple classifiers, this research aims to propose an optimized approach for early and reliable melanoma diagnosis. The ultimate goal is to contribute to better healthcare solutions, supporting dermatologists in making faster and more accurate decisions.

1.3 PROBLEM STATEMENT AND OBJECTIVES

PROBLEM STATEMENT

Melanoma is an aggressive form of skin cancer that spreads rapidly if not diagnosed at an early stage. Traditional detection methods rely heavily on dermatologists' visual examination, which can be subjective, time-consuming, and prone to errors. The similarity of melanoma to benign skin lesions further complicates accurate classification. Therefore, there is a need for an automated, efficient, and accurate melanoma detection system that leverages advanced machine learning techniques to enhance early diagnosis and improve treatment outcomes.

OBJECTIVES

The objectives of this project are:

1. To develop an automated melanoma detection system using Thepade's SBTC and GLCM for feature extraction.
2. To classify skin lesions as benign or malignant using multiple machine learning classifiers and evaluate their performance.
3. To improve classification accuracy by combining SBTC and GLCM features with an ensemble of Multilayer Perceptron and Random Forest classifiers.
4. To utilize the HAM-10000 dataset for training and validating the proposed system with real-world skin lesion images.
5. To enhance early melanoma detection and contribute to better healthcare solutions by reducing the dependency on manual diagnosis.

1.4 SCOPE OF THE WORK

This project focuses on developing an automated melanoma detection system by leveraging advanced machine learning techniques. The primary goal is to improve the accuracy of melanoma classification by combining Thepade's Sorted Block Truncation Coding (SBTC) and Gray Level Co-occurrence Matrix (GLCM) for feature extraction. The scope of this study includes image preprocessing, feature extraction, machine learning-based classification, and performance evaluation.

The system is trained and tested using the HAM-10000 dataset, which consists of real-world skin lesion images labeled as benign or malignant. Various machine learning classifiers, such as Multilayer Perceptron, Random Forest, Naïve Bayes, and others, are evaluated to identify the most effective model for melanoma detection. The proposed ensemble approach of Multilayer Perceptron and Random Forest achieves 88.46% accuracy, surpassing individual classification methods.

The findings of this study contribute to early melanoma diagnosis, reducing dependency on manual examination and aiding dermatologists in clinical decision-making. The project is designed to be scalable, allowing for future enhancements such as deep learning integration, larger datasets, and real-time detection systems for improved medical applications.

CHAPTER 2: LITERATURE SURVEY

2.1 LITERATURE REVIEW

Melanoma is a highly aggressive skin cancer that requires early detection for effective treatment. Traditional diagnosis methods rely on visual inspection and dermoscopy, which are subjective and dependent on dermatologists' expertise. Recent advancements in machine learning and deep learning have significantly improved melanoma detection by providing automated, accurate, and scalable solutions. Several studies have explored different techniques for feature extraction, classification, and performance optimization in melanoma diagnosis.

Deep learning-based approaches have gained popularity due to their ability to extract complex patterns from images. Indraswari et al. [1] proposed a melanoma classification model using MobileNetV2, achieving better accuracy through pre-training on the ImageNet dataset. Similarly, Mahmud et al. [2] used XceptionNet, EfficientNetV2S, and InceptionResNetV2 models, optimizing detection through deep feature extraction and data augmentation techniques. These studies demonstrated that deep learning models perform well in melanoma classification but require large datasets and extensive computational resources.

Machine learning techniques based on feature extraction and traditional classifiers have also been explored. Thepade et al. [6] developed a technique combining Sorted Block Truncation Coding (SBTC) and Gray Level Co-occurrence Matrix (GLCM) for extracting color and texture-based features. Their approach, implemented using MATLAB and WEKA, demonstrated improved accuracy compared to individual feature extraction methods. Kalpana et al. [5] introduced a hybrid model that combines Support Vector Machines (SVM) and Random Forest, optimized using Aquila Optimization Algorithm (AOA) to enhance classification performance.

Hybrid and ensemble learning approaches have been widely adopted to further improve accuracy. Midasala et al. [4] proposed MFEUsLNet, which integrates GLCM-based texture features with deep learning architectures to improve melanoma classification. Additionally, Raval et al. [3] explored Convolutional Neural Networks (CNNs) such as AlexNet, VGGNet, and ResNet, demonstrating that deeper networks improve accuracy but require substantial computational resources.

The review of existing literature highlights that while deep learning models achieve high accuracy, they require large datasets and extensive training time. In contrast, feature-based machine learning models provide competitive accuracy with lower computational requirements. This study builds upon previous research by combining Thepade's SBTC and GLCM features with an ensemble of Multilayer Perceptron and Random Forest classifiers, achieving 88.46% accuracy. The results validate that integrating color and texture features enhances melanoma classification, providing a more efficient and scalable solution for early detection.

2.2 GAP IDENTIFICATION / COMMON FINDINGS FROM THE LITERATURE

GAP IDENTIFICATION

1. Computational Cost of Deep Learning Models – While CNNs provide high accuracy, they require large-scale datasets and powerful GPUs, limiting their real-time implementation.
2. Limited Exploration of Feature Fusion Techniques – Most studies focus on either deep learning or single feature extraction methods rather than integrating multiple feature extraction techniques for better accuracy.
3. Lack of Optimal Classifier Selection – Many approaches use standard classifiers without optimizing ensemble techniques, leading to suboptimal performance.
4. Imbalance in Datasets – The HAM-10000 dataset and other melanoma datasets contain significantly more benign cases than malignant ones, impacting model generalization.
5. Limited Real-Time Application – Most research is conducted in controlled environments, with limited exploration of real-world deployment and mobile-based diagnosis systems.

COMMON FINDINGS FROM THE LITERATURE

1. Deep Learning Achieves High Accuracy – Studies using CNN-based models like MobileNetV2, XceptionNet, and ResNet demonstrate high classification accuracy but require large datasets and extensive computational resources.
2. Feature-Based Machine Learning is Computationally Efficient – Methods using GLCM, SBTC, and handcrafted feature extraction provide reliable classification with lower computational costs compared to deep learning.
3. Hybrid and Ensemble Approaches Improve Performance – Combining multiple feature extraction techniques and ensemble classifiers (Random Forest, SVM, and MLP) enhances classification accuracy.
4. Data Augmentation Improves Generalization – Many studies apply rotation, flipping, and scaling to expand the dataset and reduce overfitting.
5. Performance Metrics Vary Across Methods – While deep learning models often report higher accuracy, feature-based methods are more interpretable and practical for real-time applications.

2.3 IMPROVEMENTS OVER EXISTING METHODS

1. Enhanced Feature Extraction – Combines Thepade's SBTC (color features) and GLCM (texture features) to improve melanoma classification.
2. Optimized Classifier Selection – Uses an ensemble of Multilayer Perceptron (MLP) and Random Forest, achieving 88.46% accuracy, outperforming individual models.
3. Reduced Computational Cost – Unlike deep learning methods requiring large datasets and GPUs, this approach offers a lightweight and efficient solution.
4. Better Handling of Dataset Imbalance – Feature-based extraction improves classification even with fewer malignant samples, reducing bias.
5. Foundation for Real-Time Detection – Provides a scalable, interpretable, and practical approach for future automated melanoma detection systems.

CHAPTER 3: SOFTWARE REQUIREMENTS SPECIFICATIONS

3.1 FUNCTIONAL REQUIREMENTS

1. Dataset Processing – Load, preprocess, and manage the HAM-10000 dataset, including data augmentation and normalization.
2. Feature Extraction – Extract color features using Thepade's SBTC and texture features using GLCM for improved melanoma detection.
3. Machine Learning Model Training – Train multiple classifiers (MLP, Random Forest, etc.) with optimized hyperparameters for better accuracy.
4. Classification and Prediction – Classify skin lesion images as benign or malignant, providing real-time predictions with accuracy scores.
5. Performance Evaluation and Reporting – Measure performance using accuracy, precision, recall, and F1-score, and generate visual reports for analysis.

3.2 EXTERNAL INTERFACE REQUIREMENTS

3..2.1 USER INTERFACES

1. Image Upload Interface – Allows users to upload skin lesion images (JPEG, PNG) for classification. A simple drag-and-drop or file selection feature will be provided.
2. Processing & Detection Screen – Displays the real-time processing status, showing feature extraction steps and classification progress.
3. Result Display Interface – Shows the classification result (Benign or Malignant) along with confidence scores and key extracted features.

3.2.2 HARDWARE INTERFACES

1. Processor – The system requires at least an Intel Core i5 or higher (or AMD Ryzen 5 equivalent) for efficient processing of feature extraction and classification tasks.
2. Memory (RAM) – A minimum of 8GB RAM is required to handle image processing, feature extraction, and machine learning model training without lag.
3. Storage – At least 20GB of free disk space is needed to store datasets, extracted features, trained models, and classification results. SSD storage is recommended for faster performance.
4. Graphics Processing Unit (GPU) (Optional) – If deep learning is integrated, a dedicated GPU (NVIDIA GTX 1650 or higher) is recommended to accelerate training and inference.
5. Peripheral Devices – A monitor, keyboard, and mouse are required for system operation, along with a high-resolution camera or scanner for capturing skin lesion images if real-time image acquisition is implemented.

3.2.3 SOFTWARE INTERFACES

1. Operating System – The system is compatible with Windows (10/11), Linux (Ubuntu 20.04+), and macOS for flexibility in deployment.
2. Programming Languages – Developed using Python (for machine learning and image processing) and MATLAB/WEKA (for feature extraction and classification experiments).
3. Machine Learning Libraries – Uses TensorFlow, Scikit-learn, OpenCV, and NumPy for feature extraction, model training, and classification.
4. Database Management System (Optional) – If needed, classification results and user-uploaded images can be stored in MySQL, SQLite, or MongoDB for future reference and analysis.
5. User Interface Framework – If a GUI is implemented, Tkinter (for desktop), Flask/Django (for web-based UI), or Streamlit can be used to create an interactive user interface.

3.2.4 COMMUNICATION INTERFACES

1. File-Based Communication – The system allows users to upload skin lesion images (JPEG, PNG) for classification and download results in CSV, JSON, or PDF formats.
2. API Integration (Optional) – If deployed as a web service, the system can support RESTful APIs using Flask or Django, allowing external applications to send images and receive classification results.
3. Database Communication – If a database is used, the system will communicate with MySQL, SQLite, or MongoDB to store and retrieve classification results and image metadata.
4. Cloud Storage (Optional) – The system can be extended to integrate with cloud services like Google Drive, AWS S3, or Firebase for remote access and storage of images and results.
5. Network Communication – If deployed in a networked environment, the system should support HTTP/HTTPS protocols for secure data transfer between client and server applications.

3.3 NON-FUNCTIONAL REQUIREMENTS

3.3.1 PERFORMANCE REQUIREMENTS

1. Fast Processing Time – The system should classify skin lesion images within 3-5 seconds, ensuring quick response for real-time diagnosis.
2. High Classification Accuracy – The system should achieve at least 88% accuracy, minimizing false positives and false negatives in melanoma detection.
3. Efficient Resource Utilization – The system should operate smoothly on a minimum hardware configuration (Intel Core i5, 8GB RAM) without excessive CPU or memory usage.
4. Scalability – The system should handle large datasets (HAM-10000 and beyond) without significant degradation in performance.
5. Error Handling – The system should effectively detect and handle missing or corrupted image files, ensuring seamless operation without crashes.

3.3.2 SAFETY / SECURITY REQUIREMENTS

1. Data Privacy – The system must ensure that uploaded skin lesion images and classification results are not shared or accessed without authorization.
2. Secure Storage – If a database is used, all stored images and results must be encrypted to prevent unauthorized access.
3. User Authentication (If Web-Based) – If deployed as a web application, the system should implement user authentication mechanisms (e.g., login, role-based access control) to restrict access.
4. Secure Data Transmission – If the system communicates over a network, it should use HTTPS encryption to protect data integrity during image uploads and result retrieval.
5. Error Handling & Logging – The system should log errors securely and prevent crashes due to invalid or malicious inputs, ensuring reliable and safe operation.

3.4 SYSTEM REQUIREMENTS

3.4.1 DATABASE REQUIREMENTS

Database Type – The system should use a relational database (MySQL, PostgreSQL) or a NoSQL database (MongoDB) based on scalability and storage needs.

Data Storage – The database should store image metadata, extracted features, classification results, timestamps, and user details (if applicable).

Data Integrity – Ensure data consistency and accuracy by implementing primary keys, foreign keys, and constraints to avoid duplication and corruption.

Performance Optimization – The database should support indexed searches and optimized queries to enable fast retrieval of classification results.

Backup and Recovery – Implement automatic backups to prevent data loss and provide a recovery mechanism in case of system failures.

3.4.2 SOFTWARE REQUIREMENTS (PLATFORM CHOICE)

1. Operating System – The system should be compatible with Windows (10/11), Linux (Ubuntu 20.04+), and macOS for flexible deployment.
2. Programming Language – The implementation will be done using Python, leveraging its extensive libraries for machine learning, image processing, and feature extraction.
3. Development Frameworks & Libraries –

- Machine Learning: Scikit-learn, TensorFlow/Keras
- Feature Extraction: OpenCV, NumPy, MATLAB (if applicable)

4. Database Management System (Optional) – MySQL, PostgreSQL, or MongoDB for storing processed images, extracted features, and classification results.
5. User Interface (If Implemented) –

- Desktop UI: Tkinter or PyQt
- Web-based UI: Flask or Django for backend, HTML/CSS/JavaScript for frontend

3.4.3 HARDWARE REQUIREMENTS

1. Processor – Minimum Intel Core i5 (10th Gen or higher) or AMD Ryzen 5 for efficient processing of machine learning models.
2. Memory (RAM) – At least 8GB RAM to handle image processing, feature extraction, and classification tasks smoothly.
3. Storage – Minimum 20GB free disk space, preferably SSD, for storing datasets, extracted features, and trained models.
4. Graphics Processing Unit (GPU) (Optional) – NVIDIA GTX 1650 or higher recommended if deep learning models are used for classification.
5. Peripheral Devices – Monitor, keyboard, and mouse for system operation, and a high-resolution camera or scanner if real-time image capture is required.

CHAPTER 4: PROPOSED METHODOLOGY

4.1 PROPOSED SYSTEM ARCHITECTURE / BLOCK DIAGRAM

The proposed system follows a structured workflow to classify melanoma skin cancer using Thepade's SBTC and GLCM for feature extraction and machine learning classifiers for detection. Below is the block diagram representation of the system:

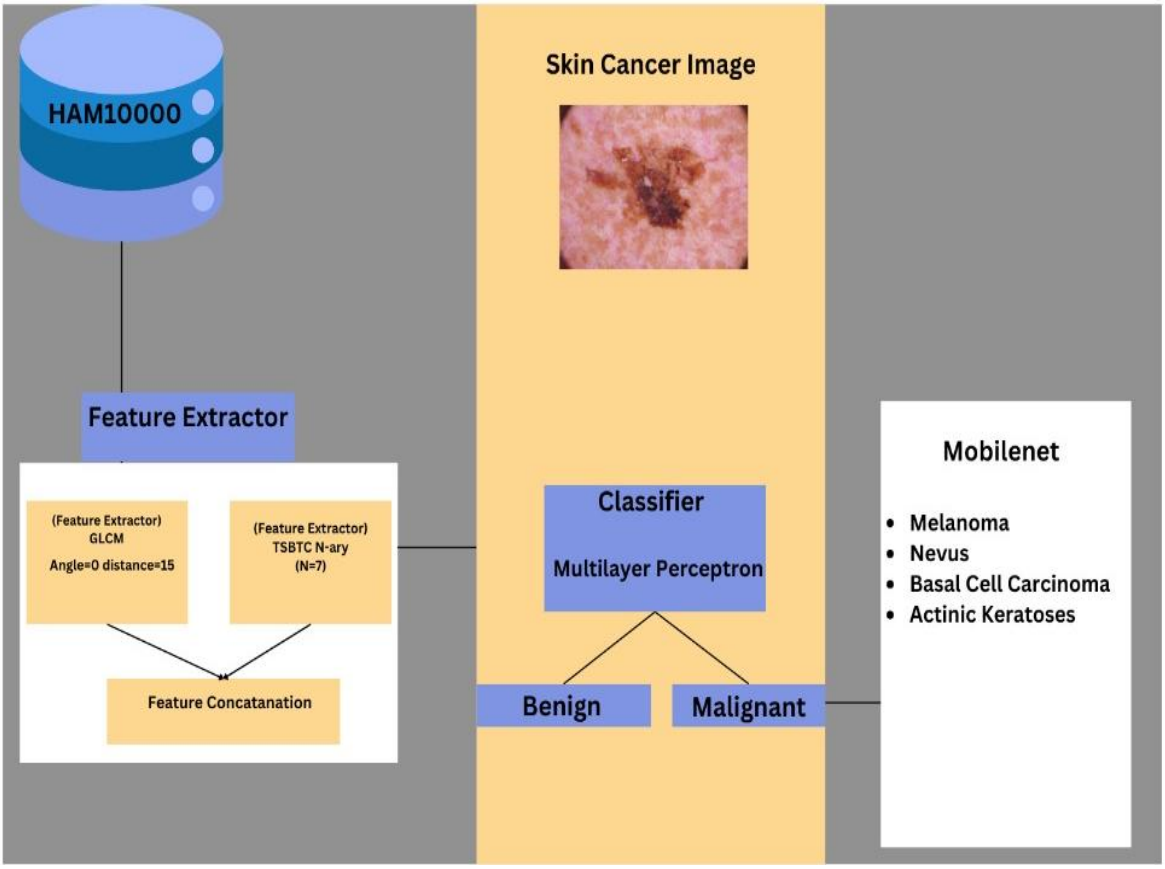


Fig 1: SYSTEM ARCHITECTURE BLOCK DIAGRAM

4.1.1 SYSTEM ARCHITECTURE WORKFLOW

1. Input Image Acquisition – Skin lesion images are uploaded from the HAM-10000 dataset or user input.
2. Preprocessing – Image resizing, normalization, and enhancement are performed to improve feature extraction.
3. Feature Extraction –

- SBTC extracts color-based features (Red, Green, Blue planes).
- GLCM extracts texture-based features (contrast, correlation, energy, homogeneity).

4. Feature Fusion – The extracted features from SBTC and GLCM are combined into a single feature vector.
5. Classification – Machine learning classifiers such as Multilayer Perceptron (MLP) and Random Forest analyze the extracted features to classify the image as benign or malignant.
6. Result Display & Analysis – The system outputs the classification result along with accuracy metrics and visual reports.

4.2 DATASET / DATABASE DESIGN

1. DATASET DESCRIPTION

The system uses the HAM-10000 dataset, a publicly available dataset containing microscopic images of skin lesions labeled as benign or malignant.

The dataset consists of 10,015 images, with metadata such as lesion type, patient ID, and image source.

2. DATASET SCHEMA DESIGN

(A) IMAGE DATA TABLE

TABLE 1: IMAGE DATA TABLE

COLUMN NAME	DATA TYPE	DESCRIPTION
image_id	VARCHAR(50)	Unique identifier for each image
file_path	TEXT	Location of the stored image
upload_date	DATETIME	Timestamp of when the image was added
label	VARCHAR(10)	Benign / Malignant

(B) FEATURE EXTRACTION TABLE

TABLE 2: FEATURE EXTRACTION TABLE

COLUMN NAME	DATA TYPE	DESCRIPTION
feature_id	INT (Primary Key)	Unique identifier for extracted features
image_id	VARCHAR(50)	Foreign key referencing Image Data Table
SBTC_features	TEXT	Extracted color features
GLCM_features	TEXT	Extracted texture features

(C) CLASSIFICATION RESULTS TABLE

TABLE 3: CLASSIFICATION RESULTS TABLE

COLUMN NAME	DATA TYPE	DESCRIPTION
result_id	INT (Primary Key)	Unique identifier for classification results
image_id	VARCHAR(50)	Foreign key referencing Image Data Table
classifier_used	VARCHAR(50)	MLP / Random Forest
prediction	VARCHAR(10)	Benign / Malignant
confidence_score	FLOAT	Model confidence score (%)
classification_date	DATETIME	Timestamp of classification

4.3 OVERVIEW OF PROJECT MODULES

1. IMAGE ACQUISITION MODULE

The Image Acquisition Module is responsible for collecting and managing microscopic blood cell images for analysis. It ensures that high-quality images are obtained for accurate feature extraction and classification.

FUNCTIONS OF THE IMAGE ACQUISITION MODULE

1. Dataset Loading – Loads images from the HAM-10000 dataset or other medical image repositories.
2. User Image Upload – Allows users to upload microscopic blood smear images in formats like JPEG, PNG, or TIFF.
3. Image Validation – Checks image quality, resolution, and format before processing.

4. Image Storage – Saves uploaded images in a structured database or local directory for further processing.
5. Real-Time Image Capture (Optional) – If integrated with a microscope camera, captures real-time images for immediate analysis.

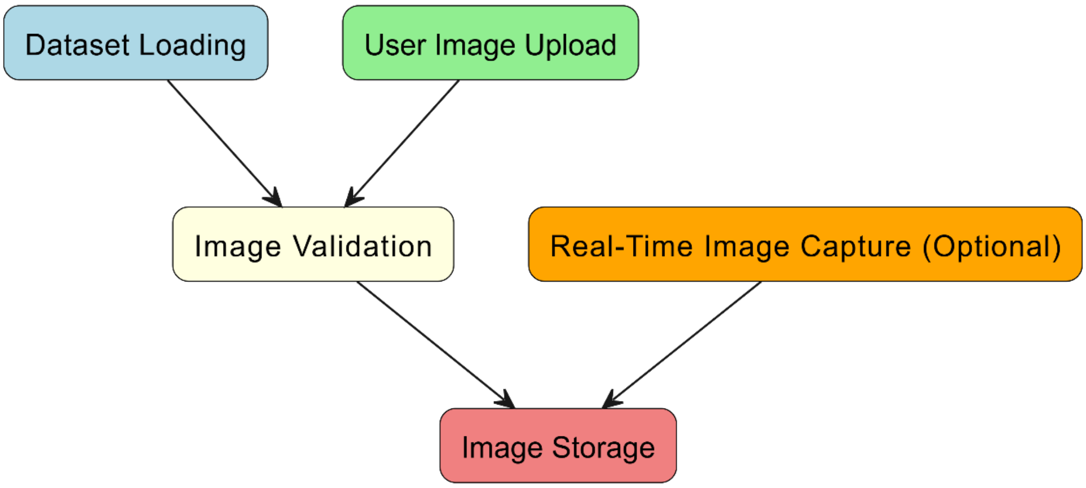


Fig 2. IMAGE ACQUISITION MODULE BLOCK DIAGRAM

2. PREPROCESSING MODULE

The Preprocessing Module prepares microscopic blood cell images for accurate feature extraction and classification by enhancing image quality and standardizing formats.

FUNCTIONS OF THE PREPROCESSING MODULE

1. Noise Removal – Applies filters (Gaussian, Median, or Bilateral) to reduce unwanted artifacts in images.
2. Contrast Enhancement – Uses histogram equalization or adaptive contrast techniques to improve visibility.
3. Image Normalization – Standardizes image sizes and pixel intensity values for consistent processing.
4. Segmentation – Identifies regions of interest (ROI) to focus only on relevant parts of the image.
5. Edge Detection (Optional) – Enhances boundaries using Canny or Sobel edge detection for better feature extraction.

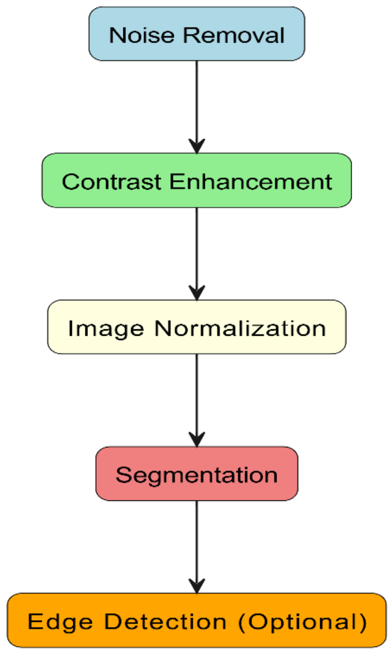


Fig 3. PREPROCESSING MODULE BLOCK DIAGRAM

3. FEATURE EXTRACTION MODULE

The Feature Extraction Module is responsible for extracting key features from microscopic blood cell images to differentiate between benign and malignant cases. It uses color-based and texture-based feature extraction techniques for improved classification accuracy.

FUNCTIONS OF THE FEATURE EXTRACTION MODULE

1. SBTC (Sorted Block Truncation Coding) – Extracts color features from Red, Green, and Blue (RGB) planes.
2. GLCM (Gray Level Co-occurrence Matrix) – Extracts texture features such as contrast, correlation, energy, and homogeneity.
3. Feature Normalization – Ensures uniform scaling of extracted features for consistent classification.
4. Feature Vector Formation – Combines extracted color and texture features into a structured feature vector.

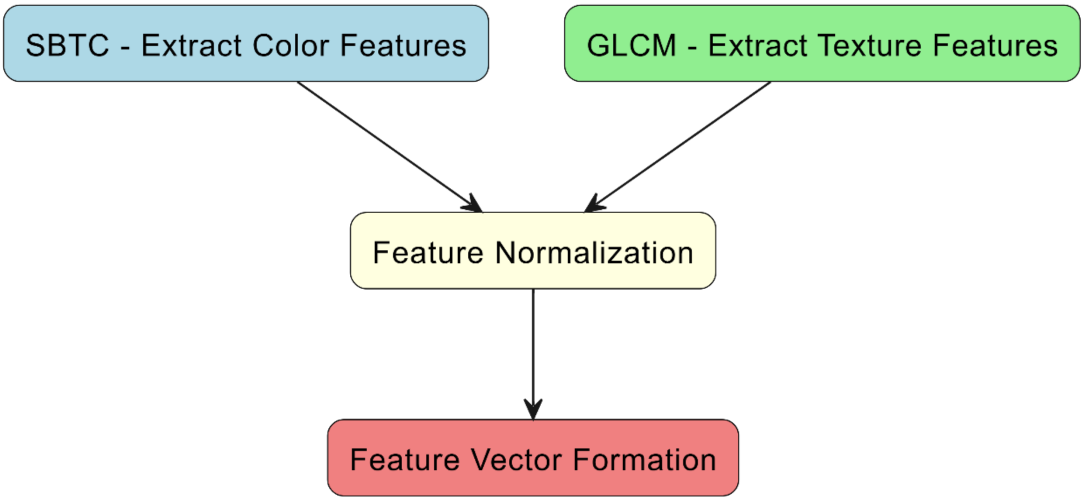


Fig 4. FEATURE EXTRACTION MODULE BLOCK DIAGRAM

4. FEATURE FUSION MODULE

The Feature Fusion Module combines extracted color features (SBTC) and texture features (GLCM) into a unified feature vector for improved classification accuracy. This ensures that the classifier gets a more comprehensive representation of the input image.

FUNCTIONS OF THE FEATURE FUSION MODULE

1. Input from Feature Extraction – Takes SBTC (color) and GLCM (texture) features as input.
2. Feature Scaling & Normalization – Standardizes extracted features to maintain consistency.
3. Feature Concatenation – Merges color and texture feature vectors into a single feature representation.
4. Dimensionality Reduction (Optional) – Reduces redundant data using PCA (Principal Component Analysis) or LDA (Linear Discriminant Analysis) to improve efficiency.
5. Final Feature Vector Formation – Outputs a combined feature set for classification.

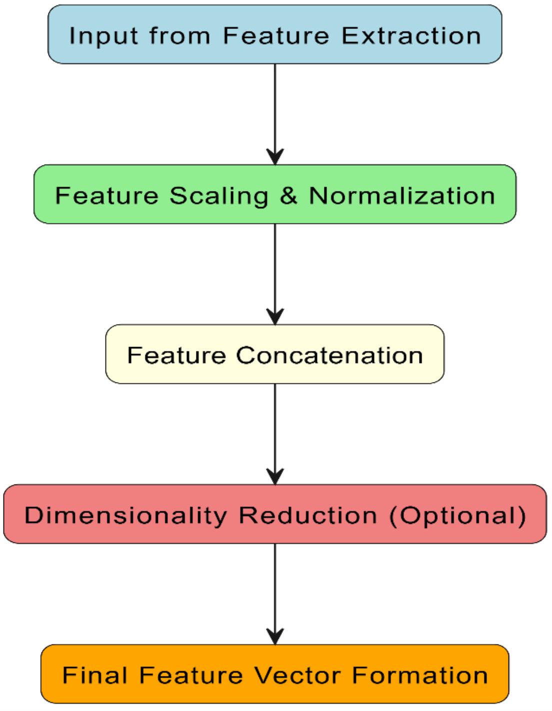


Fig 5. FEATURE FUSION MODULE BLOCK DIAGRAM

5. CLASSIFICATION MODULE

The Classification Module is responsible for analyzing the fused feature vector and classifying the input image as benign or malignant. It uses multiple machine learning classifiers to improve accuracy.

FUNCTIONS OF THE CLASSIFICATION MODULE

1. Input Feature Vector – Receives the final fused feature vector from the Feature Fusion Module.
2. Classifier Selection – Uses an ensemble of Multilayer Perceptron (MLP) and Random Forest (RF) classifiers.
3. Model Training & Optimization – Trains classifiers using labeled dataset images and optimizes performance.
4. Prediction Generation – Classifies images as benign or malignant with a confidence score.
5. Performance Evaluation – Measures model accuracy using metrics like precision, recall, and F1-score.

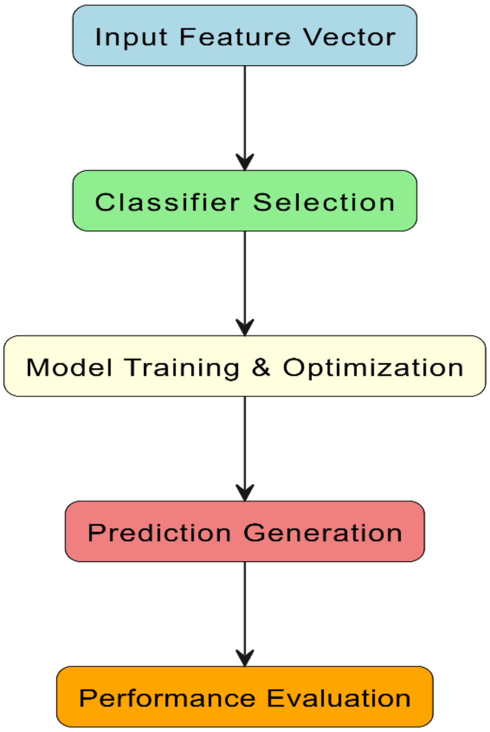


Fig 6. CLASSIFICATION MODULE BLOCK DIAGRAM

6. RESULT DISPLAY & ANALYSIS MODULE

The Result Display & Analysis Module presents the classification results and provides a detailed performance evaluation of the machine learning models used. It allows users to analyze the accuracy, precision, and reliability of the system.

FUNCTIONS OF THE RESULT DISPLAY & ANALYSIS MODULE

1. Display Classification Result – Shows whether the image is classified as benign or malignant along with the confidence score.
2. Graphical Performance Metrics – Visualizes evaluation metrics such as accuracy, precision, recall, and F1-score using graphs and confusion matrices.
3. Comparison of Classifiers – Compares the performance of Random Forest (RF) and Multilayer Perceptron (MLP) classifiers.
4. Report Generation – Exports classification results and model performance analysis in CSV, JSON, or PDF formats.
5. User Interaction & Feedback (Optional) – Allows users to provide feedback on classification accuracy for future improvements.

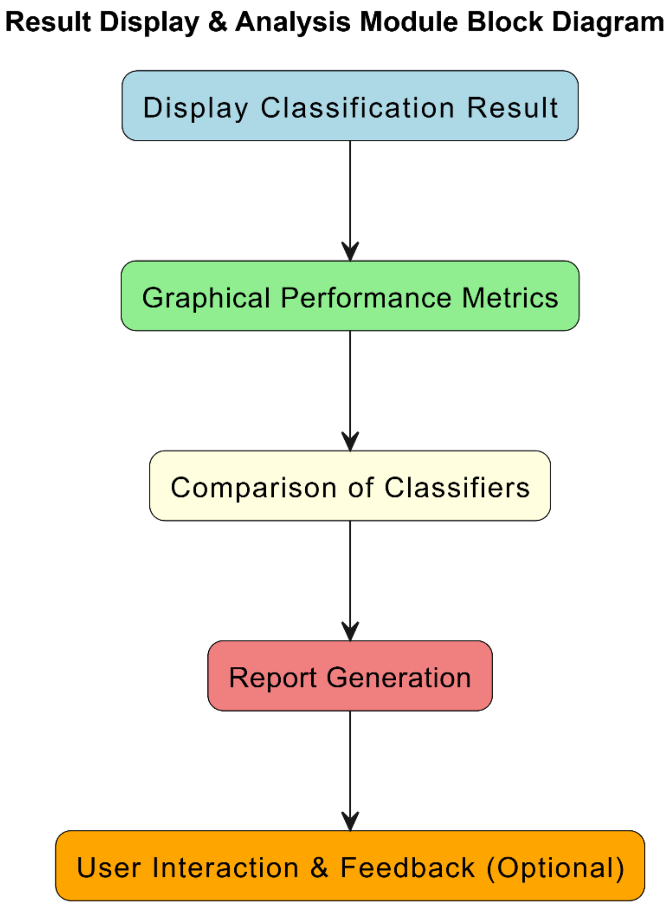


Fig 7. RESULT DISPLAY & ANALYSIS MODULE BLOCK DIAGRAM

4.4 TOOLS AND TECHNOLOGIES USED

1. Programming Languages
   - Python – For machine learning model implementation and data processing.
   - MATLAB – For advanced feature extraction and visualization.
2. Machine Learning & Deep Learning Libraries
   - Scikit-learn – Implements Random Forest, Multilayer Perceptron (MLP), and performance metrics.
   - TensorFlow/Keras – (Optional) For deep learning model integration.
3. Image Processing Libraries
   - OpenCV – Performs image preprocessing (noise removal, contrast enhancement, segmentation).

- NumPy & Pandas – Handles image data and feature extraction.

4. Feature Extraction Techniques

- Thepade's SBTC (Sorted Block Truncation Coding) – Extracts color-based features.
- GLCM (Gray-Level Co-occurrence Matrix) – Extracts texture-based features.

5. Development Tools & Platforms

- Jupyter Notebook / PyCharm – For coding, testing, and debugging machine learning models.
- WEKA – Used for additional dataset analysis and classification.

6. Database & Data Storage

- MySQL – Stores processed images, extracted features, and classification results.

7. Visualization & Reporting

- Matplotlib & Seaborn – Generates performance graphs, confusion matrices, and accuracy reports.
- Flask / Django (Optional) – If deploying as a web-based application.

4.5 MATHEMATICAL MODEL

The mathematical model for the Cancer Detection System defines how features are extracted, processed, and classified using machine learning techniques.

1. Problem Definition (Set Theory Representation)

Let:

- S be the overall system
- I be the set of input images
- P be the preprocessing function
- F be the feature extraction function
- C be the classification function
- O be the output (benign/malignant prediction)

The system can be represented as:

$$S = \{I, P, F, C, O\}$$

Where:

- P : I → I′ (Preprocessed image)
- F : I′ → V (Feature vector extraction)
- C : V→ O (Classification using ML models)

2. Feature Extraction Equations

- SBTC (Sorted Block Truncation Coding) for Color Features:

$$F_{\text{SBTC}} = \frac{1}{N} \sum_{i=1}^{N} (R_i, G_i, B_i)$$

Where R, G, B are color plane values.

- GLCM (Gray-Level Co-occurrence Matrix) for Texture Features:

$$F_{\text{GLCM}} = \sum_{i,j} P(i,j) \times f(i,j)$$

Where P(i, j)is the probability of pixel intensity co-occurrence, and f(i, j) represents contrast, energy, correlation, or homogeneity.

3. Classification Model (Multilayer Perceptron & Random Forest)

- Multilayer Perceptron (MLP) Activation Function:

$$a = f(WX + b)$$

Where:

- W = Weight matrix
- X = Input feature vector
- b = Bias term
- f(x) = Activation function (ReLU or Sigmoid)

- Random Forest Decision Function:

$$C(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x)$$

Where $h_t(x)$ represents the decision from each decision tree, and T is the number of trees.

4. Performance Metrics

- Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision:

$$Precision = \frac{TP}{TP + FP}$$

- Recall:

$$Recall = \frac{TP}{TP + FN}$$

- F1-Score:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

4.6 ALGORITHM DETAILS

4.6.1 ALGORITHM 1: FEATURE EXTRACTION VIA THEPADE'S SBTC

1. Start
2. Obtain input skin lesion images from the dataset.
3. Extract Red, Green, and Blue (RGB) color planes.
4. Organize pixel intensity values into columns, where each column represents a pixel.
5. Sort intensity values within each column in ascending order.
6. Divide the sorted column vector into n-ary portions (e.g., binary, ternary).
7. Compute the average intensity value for each portion using predefined equations.
8. Generate a feature vector based on SBTC-derived color features.
9. End

4.6.2 ALGORITHM 2: FEATURE EXTRACTION VIA GLCM

1. Start
2. Convert the input image to grayscale if needed.
3. Construct the Gray-Level Co-occurrence Matrix (GLCM) from the image.
4. Extract statistical texture features from GLCM, including:

- Energy
- Dissimilarity
- Contrast
- Homogeneity
- Correlation

6. Generate a feature vector using extracted values.
7. End

4.6.3 CLASSIFICATION USING MLP AND RANDOM FOREST

1. Start
2. Input the SBTC + GLCM feature vectors into classifiers.
3. Train two classifiers:
4. Multilayer Perceptron (MLP): Uses a feedforward neural network with activation functions.
5. Random Forest (RF): Creates multiple decision trees and averages the predictions.
6. Predict whether the lesion is benign or malignant.
7. Compute classification confidence scores.
8. End

4.7 COMPLEXITY OF PROJECT

1. Data Processing Complexity

1. High-resolution medical images require preprocessing (noise removal, contrast enhancement, and normalization).
2. Feature extraction using SBTC and GLCM involves sorting, segmentation, and statistical computations.

2. Computational Complexity

1. SBTC Sorting Complexity: O(nlogn) per image column (where n is the number of pixels).
2. GLCM Feature Computation Complexity: $O(p^2)$, where p is the number of intensity levels.
3. MLP Classification Complexity: O(W·n) where W is the number of weights and n is the number of input features.
4. Random Forest Complexity: O(T·mlogm) where T is the number of trees and mmm is the number of features.

3. **Algorithmic Complexity**

1. Feature Fusion requires merging color-based (SBTC) and texture-based (GLCM) features, increasing dimensionality.
2. Ensemble Classification (MLP & RF) adds computational overhead but improves accuracy.

4. **Implementation Complexity**

1. Preprocessing needs OpenCV, NumPy, and Pandas for handling image transformations.
2. Feature extraction demands mathematical operations on large datasets.
3. MLP & RF classifiers require extensive hyperparameter tuning for optimal results.

5. **Performance Complexity**

1. Achieving high accuracy (>88%) requires extensive parameter tuning and dataset balancing.
2. Real-time analysis needs efficient processing due to large image sizes.

6. **Overall Complexity Level: Moderate to High**

1. Moderate complexity in image preprocessing and feature extraction.
2. High complexity in classification model training, optimization, and real-time prediction.

4.8 ENTITY RELATIONSHIP DIAGRAM

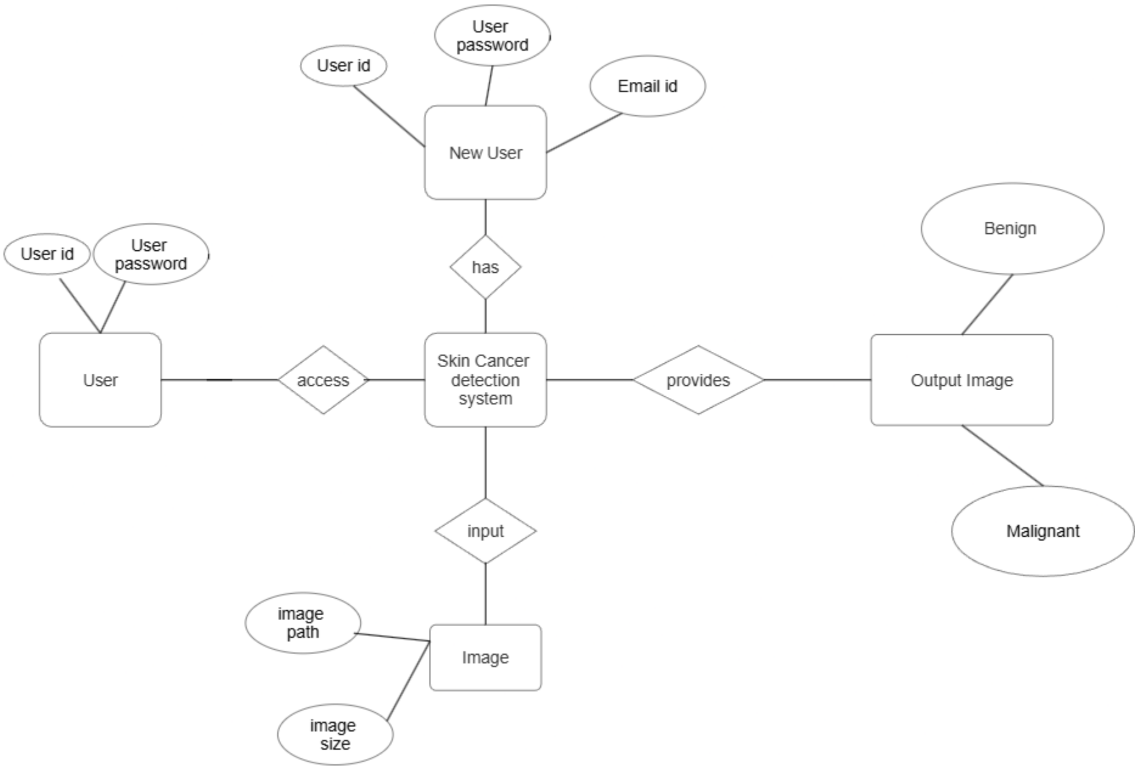


Fig 8. ENTITY RELATIONSHIP DIAGRAM

4.9 SDLC MODEL TO BE APPLIED

For the Skin Cancer Detection System, the Spiral Model is the most suitable Software Development Life Cycle (SDLC) model due to its iterative nature, risk handling, prototyping approach and adaptability to machine learning-based systems.

Phases of the Spiral Model for This Project

1. Planning Phase

- Define project objectives (melanoma detection using SBTC + GLCM).
- Identify dataset requirements (e.g., HAM-10000).
- Select classification algorithms (MLP, Random Forest).

2. Risk Analysis Phase

- Assess risks such as dataset imbalance, feature selection challenges, and classifier performance issues.
- Conduct a comparative evaluation of different feature extraction techniques.

3. Development & Prototyping Phase

- Implement image preprocessing and feature extraction.
- Train MLP and Random Forest classifiers.
- Evaluate classification performance using accuracy, precision, and recall metrics.

4. Testing & Validation Phase

- Perform model validation using a confusion matrix and cross-validation.
- Fine-tune hyperparameters for improved classification accuracy.
- Identify errors and retrain models to optimize performance.

5. Deployment & Maintenance Phase

- Deploy the trained model for real-time melanoma detection.
- Monitor performance and enhance accuracy with additional datasets.
- Upgrade the system by integrating advanced deep learning models (e.g., CNN, YOLO) in future iterations.

4.10 UML DIAGRAMS

Unified Modeling Language (UML) is a standardized visual language used to model the design and structure of software systems. It helps developers and stakeholders understand, visualize, and document different aspects of a system. UML diagrams are broadly categorized into structural diagrams (such as Class Diagrams) that show the static aspects, and behavioral diagrams (such as Use Case, Activity, and Sequence Diagrams) that depict the dynamic behavior of the system. These diagrams collectively offer a comprehensive view of the system's functionality, data flow, and interactions among components, making them a vital part of software development documentation.

Using UML diagrams allows developers, designers, and stakeholders to communicate clearly and effectively throughout the development lifecycle. These diagrams not only enhance understanding of complex systems but also serve as a blueprint during implementation and maintenance phases. In this project, various UML diagrams are used to visualize and explain the functionality, processes, and relationships within the system.

ACTIVITY DIAGRAM:

The activity diagram illustrates the overall workflow of the skin cancer detection system. It begins with the user login/registration process, followed by navigation to the home page. From there, the user uploads a skin image for analysis. The system performs an initial analysis to detect whether the result is benign or malignant. If the result is benign, it is directly shown to the user. In case of a malignant detection, the system prompts the user with an option to perform a detailed analysis. Upon user confirmation, the system conducts further analysis and displays the specific cancer type. This diagram effectively captures the dynamic behavior and decision flows in the system.

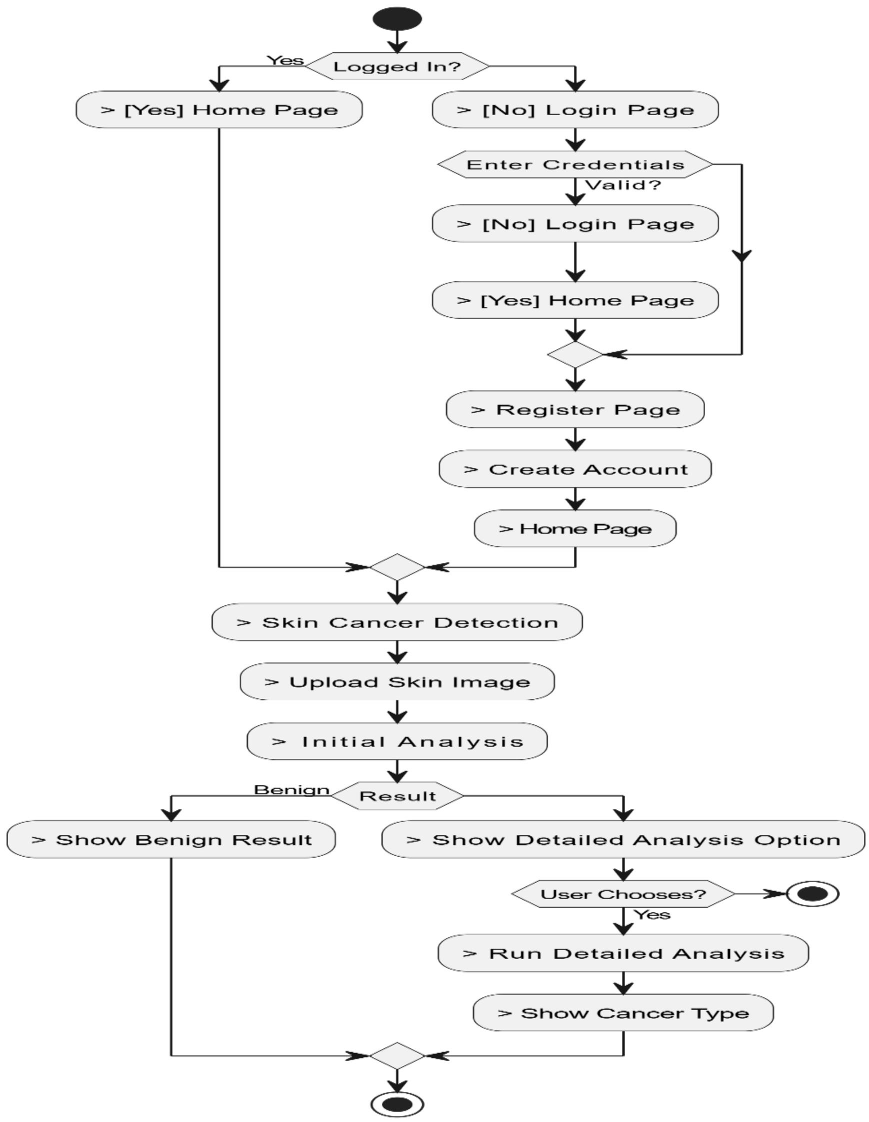


Fig 9. ACTIVITY DIAGRAM

SEQUENCE DIAGRAM:

The sequence diagram represents the interaction between the components involved in the prediction process. The user uploads a skin image through the web interface, which is then processed by a Flask application. The image is passed to a binary classifier model (using an ensemble of models like MobileNet) to predict whether the lesion is benign or malignant. If malignant, the image is saved with a unique ID and the user is offered a detailed analysis option. Upon request, the Flask app uses another model to perform cancer type classification, and the result is sent back to the user. This diagram clearly shows the sequential flow of data and responses across various modules like Flask App, ML models, and the database.

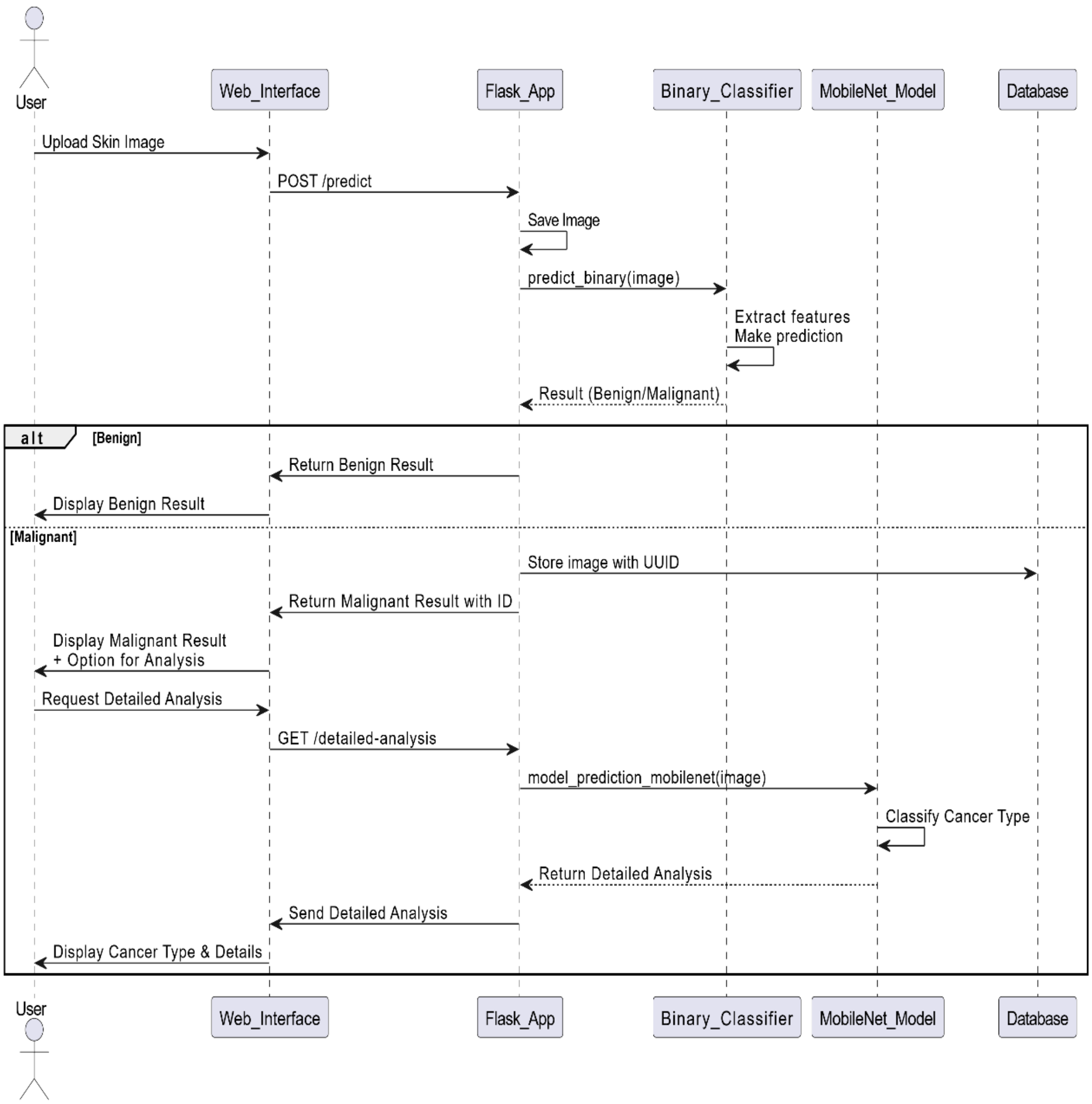


Fig 10. SEQUENCE DIAGRAM

CLASS DIAGRAM:

The UML Class Diagram for the Skin Cancer Detection System highlights the structure and relationships between key components. The Application class controls routes like login, register, and model interactions, relying on modules like Authentication, Database, and BinaryClassifier. The Flask class handles web framework functions, inherited by the Application class. The BinaryClassifier class performs skin image feature extraction and prediction using machine learning objects like MLP and scaler. The Authentication class secures routes with a login_required() decorator, while the Database class manages user operations, such as adding and retrieving user data. This diagram showcases the system's object-oriented design and inter-module collaboration.

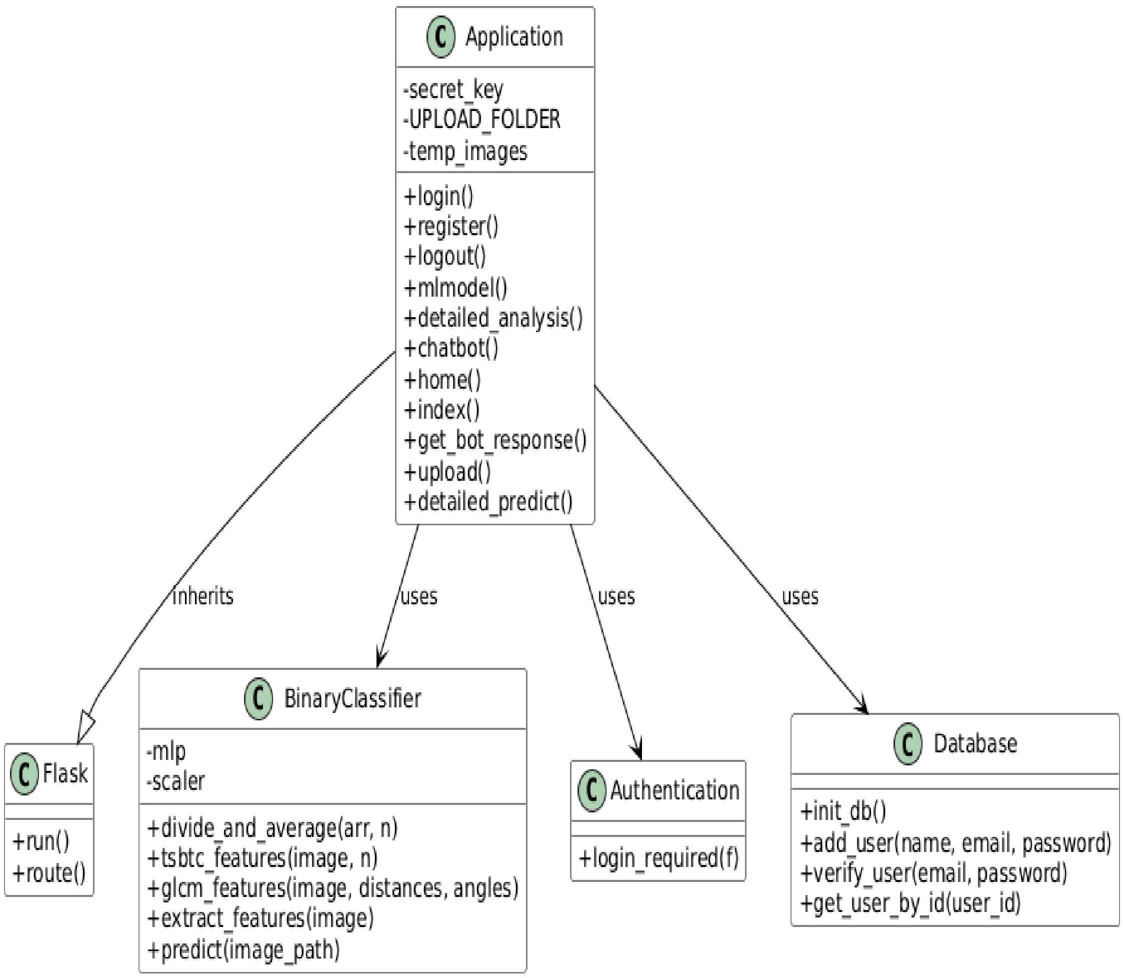


Fig 11. CLASS DIAGRAM

CHAPTER 5: PROJECT PLAN

1. Project Title:

Skin Cancer Detection Using Thepade's SBTC and GLCM Features with Machine Learning

2. Project Objectives:

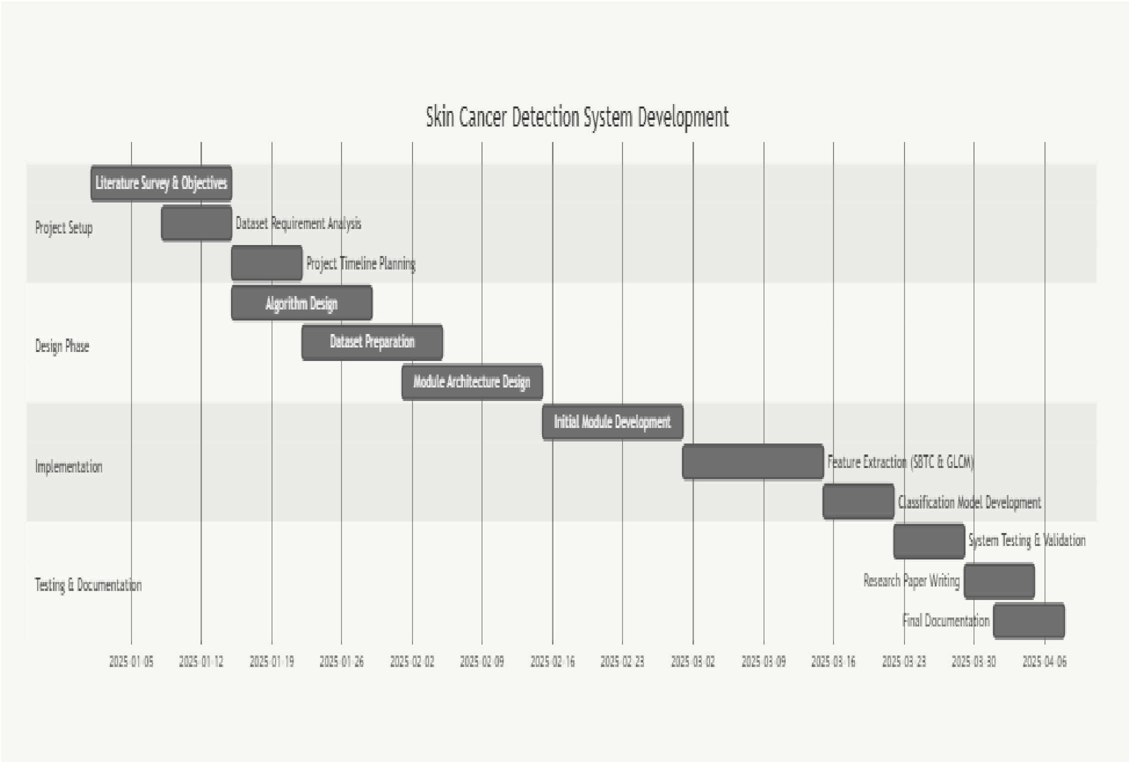
- To develop an automated system for melanoma skin cancer detection.
- To extract and analyze skin lesion features using SBTC (Sorted Block Truncation Coding) and GLCM (Gray-Level Co-occurrence Matrix).
- To classify images using Random Forest (RF) and Multilayer Perceptron (MLP) classifiers.
- To evaluate and compare model performance based on accuracy, precision, recall, and F1-score.

3. Project Scope:

- Image preprocessing (noise removal, contrast enhancement, and normalization).
- Feature extraction using SBTC and GLCM.
- Machine learning classification using RF and MLP.
- Performance evaluation using standard metrics.
- Future enhancements for deep learning integration.

4. Project Phases & Timeline:

TABLE 4: PROJECT PHASES AND TIMELINE



5. Required Tools & Technologies:

- Programming: Python (NumPy, OpenCV, Scikit-learn, Matplotlib, Pandas)
- Machine Learning: MLP, Random Forest
- Feature Extraction: SBTC, GLCM
- Development Tools: Jupyter Notebook, PyCharm
- Database : MySQL/SQLite

6. Expected Outcomes:

- A functional machine learning-based system for melanoma detection.
- Improved accuracy through SBTC & GLCM feature fusion.

- Performance comparison of MLP & Random Forest classifiers.
- Detailed analysis report with evaluation metrics and future improvements.

5.1 RISK MANAGEMENT

Risk management is essential in ensuring the smooth execution of the Skin Cancer Detection System. Since this project involves medical image processing and machine learning models, various risks can impact data quality, feature extraction efficiency, classification accuracy, and system scalability. Identifying these risks early helps in implementing strategies to minimize their impact and ensure a robust, efficient, and reliable system. This section outlines potential risks, their severity, and mitigation strategies to maintain the project's success.

5.1.1 RISK IDENTIFICATION

Risk identification is a critical process in assessing potential threats that may impact the Skin Cancer Detection System. The following key risks have been identified based on the project's scope, technical challenges, and implementation constraints.

1. Data-Related Risks

- Low-Quality or Noisy Images – Images may have artifacts, low resolution, or uneven lighting, affecting feature extraction accuracy.
- Imbalanced Dataset – Unequal distribution of benign and malignant images may lead to biased model predictions.
- Privacy and Security Concerns – Handling medical data requires adherence to privacy regulations to prevent unauthorized access.

2. Algorithmic and Model Risks

- Feature Extraction Complexity – Extracting meaningful features using SBTC and GLCM may be computationally expensive and require fine-tuning.
- Low Classification Accuracy – ML models (MLP & RF) may fail to achieve high accuracy due to poor feature representation or insufficient training data.

- Overfitting or Underfitting – The model may perform well on training data but fail on new images, requiring careful hyperparameter tuning.

3. Computational and Performance Risks

- High Processing Time – Large datasets and complex computations may slow down real-time analysis.
- Hardware Limitations – Running ML models may require high-performance GPUs or cloud-based solutions.
- System Scalability Issues – As the dataset grows, the system must handle increased data loads efficiently.

4. User and Adoption Risks

- Resistance to AI-Based Diagnosis – Dermatologists may hesitate to trust AI predictions without interpretability.
- User Interface Complexity – If the system is not user-friendly, adoption may be limited.
- Requirement Changes – Evolving medical standards and AI advancements may require continuous system updates.

5.1.2 RISK ANALYSIS

Risk analysis is essential for assessing the potential impact of identified risks in the Skin Cancer Detection System. By systematically analyzing these risks, the project team can anticipate challenges, develop contingency plans, and ensure a robust and efficient system. The process involves evaluating potential risks based on their likelihood of occurrence and the severity of their impact on the system's functionality, accuracy, and adoption. The following table categorizes risks based on their severity and likelihood, along with mitigation strategies.

1. Risk Assessment Table

TABLE 5: RISK ASSESSMENT TABLE

Risk Category	Potential Risk	Mitigation Strategy	Likelihood	Impact
Data Issues	Low-quality or imbalanced dataset	Data augmentation, oversampling, and high-resolution image collection	High	High
Feature Extraction	Inefficient SBTC and GLCM feature extraction	Optimize feature extraction algorithms and adjust parameters	Medium	High
Model Performance	Low classification accuracy	Hyperparameter tuning, cross-validation, and improved feature selection	High	High
Computational Limitations	High processing time for large images	Implement GPU acceleration, optimize code, and use cloud computing	Medium	Medium
System Scalability	Performance degradation with large datasets	Develop scalable architectures and optimize storage	Medium	High
User Adoption	Resistance to AI-based diagnosis by...	Provide explainable AI outputs and confidence scores	Low	Medium

2. Risk Evaluation Process

- Probability Assessment: Each risk is classified based on how likely it is to occur (Low, Medium, High).
- Impact Assessment: Evaluates the severity of the risk on system performance and usability.
- Prioritization: High-likelihood and high-impact risks are addressed first to minimize project failures.
- Continuous Monitoring: Risks are reassessed at different project phases to adapt mitigation strategies accordingly.
- Stakeholder Involvement: Dermatologists, researchers, and engineers collaborate to refine risk management approaches.

5.1.3 OVERVIEW OF RISK MITIGATION, MONITORING, MANAGEMENT

1. Risk Mitigation

Risk mitigation focuses on proactively reducing the impact of potential risks associated with the Skin Cancer Detection System. The following strategies help in minimizing challenges and ensuring smooth system functionality:

- Data Quality Improvement – Applying data augmentation and preprocessing techniques to handle noisy and imbalanced datasets.
- Feature Optimization – Fine-tuning SBTC and GLCM feature extraction methods to enhance model accuracy.
- Model Performance Enhancement – Using hyperparameter tuning, cross-validation, and ensemble learning to improve classification accuracy.
- Computational Efficiency – Implementing GPU acceleration, optimized code, and cloud-based processing to reduce system latency.
- Scalability Planning – Designing a modular architecture to handle larger datasets and future AI advancements.
- User Training & Trust Building – Educating dermatologists and researchers on AI-based diagnosis to improve adoption rates.

2. Risk Monitoring

Continuous risk monitoring ensures that potential issues are identified and addressed throughout the system lifecycle. The key monitoring techniques include:

- Performance Tracking – Regularly analyzing accuracy, precision, recall, and F1-score to detect model degradation.
- System Logging & Debugging – Implementing real-time error logging and debugging tools to track system failures.
- Dataset Evaluation – Periodic quality checks on new datasets to ensure consistency and reliability.
- Security & Compliance Audits – Ensuring the system meets privacy regulations to protect medical data.
- User Feedback Collection – Engaging medical professionals to provide insights for system improvements.

3. Risk Management

Risk management involves a structured approach to identifying, assessing, and mitigating risks throughout the project lifecycle. The key aspects include:

- Risk Assessment Framework – Categorizing risks based on likelihood and impact to prioritize mitigation efforts.
- Contingency Planning – Preparing alternative solutions for data issues, model failures, and performance limitations.
- Stakeholder Collaboration – Engaging data scientists, dermatologists, and software engineers in risk management strategies.
- Periodic System Updates – Implementing software upgrades to integrate new ML models and enhance system accuracy.
- Scalability & Adaptability – Ensuring the system can handle future advancements in AI and medical imaging.

5.2 PROJECT SCHEDULE

The Skin Cancer Detection System follows a structured project schedule to ensure timely completion while maintaining system efficiency, accuracy, and scalability. A well-defined schedule helps in tracking progress, allocating resources effectively, and mitigating potential risks at each phase of the project. By dividing the project into clear phases with specific tasks and durations, the team can systematically implement and evaluate each component of the system.

5.2.1 PROJECT TASK SET

1. Requirement Analysis

- Define project scope and objectives.
- Identify dataset sources (e.g., HAM-10000 dataset).
- Select feature extraction methods (SBTC, GLCM).
- Choose machine learning classifiers (MLP, Random Forest).
- Establish evaluation metrics and success criteria.

2. Data Collection & Preprocessing

- Acquire and organize the dataset.
- Apply noise removal and contrast enhancement.
- Normalize images for consistent feature extraction.
- Perform segmentation (if applicable).
- Conduct exploratory data analysis (EDA) to understand data distribution.

3. Feature Extraction

- Implement Sorted Block Truncation Coding (SBTC) for color-based features.
- Implement Gray Level Co-occurrence Matrix (GLCM) for texture-based features.
- Normalize and store extracted features for model input.
- Optimize feature selection to reduce redundancy.

4. Model Training & Optimization

- Train Multilayer Perceptron (MLP) and Random Forest (RF) classifiers.
- Perform hyperparameter tuning to improve accuracy.
- Evaluate models using cross-validation.
- Compare performance metrics (accuracy, precision, recall, F1-score).
- Implement ensemble learning techniques if required.

5. Performance Evaluation

- Generate confusion matrix for classification results.
- Analyze classifier performance for both feature extraction methods.
- Adjust model parameters if required for better accuracy.
- Benchmark system performance against existing methodologies.

6. System Deployment

- Integrate trained models into a functional application.
- Develop a user interface (if applicable) for image upload and diagnosis.
- Test the system for real-time usage and scalability.

- Implement cloud-based or on-premise deployment strategies.

7. Documentation & Report Writing

- Document all experiments, observations, and findings.
- Prepare research paper and final project report.
- Include results, comparisons, and future improvements.
- Provide recommendations for future enhancements, such as deep learning integration.

CHAPTER 6: SOFTWARE TESTING

Software testing ensures that the Skin Cancer Detection System functions correctly, meets performance expectations, and produces reliable results. Testing is conducted at different levels to verify system accuracy, stability, and robustness.

6.1 TYPE OF TESTING

1. Unit Testing

- Description: Tests individual components of the system, such as feature extraction, classification, and preprocessing functions.
- Purpose: Ensures correctness of SBTC and GLCM feature extraction, reducing errors at an early stage.

2. Integration Testing

- Description: Verifies that different modules interact correctly when integrated.
- Purpose: Ensures smooth data flow between system components, such as feature extraction → classification → result display.

3. System Testing

- Description: Tests the complete system as a whole to validate its overall functionality.
- Purpose: Confirms that all system requirements are met, including image classification and report generation.

4. Performance Testing

- Description: Assesses the system's responsiveness, processing speed, and scalability under different loads.

- Purpose: Ensures that classification is performed efficiently and that the system can handle large datasets without performance degradation.

5. Validation Testing

- Description: Compares system outputs against expected results, often validated by expert dermatologists.
- Purpose: Ensures that classification accuracy aligns with medical diagnosis standards and that false positives/negatives are minimized.

6. User Acceptance Testing (UAT)

- Description: Involves real users, such as dermatologists, to evaluate the usability and reliability of the system.
- Purpose: Ensures the interface is user-friendly and the system delivers meaningful, actionable results to end users.

6.2 TEST CASES AND TEST RESULTS

Test Case ID	Expected Result	Actual Result	Status	Test Scenario
TC-01	Image successfully uploaded	Image uploaded successfully	Pass	Upload skin image
TC-02	Noise removal, contrast enhancement applied	Successfully preprocessed	Pass	Preprocess image
TC-03	SBTC & GLCM features extracted correctly	Features extracted	Pass	Extract features
TC-04	Image classified as benign/malignant	Correct classification displayed	Pass	Classify image
TC-05	Classification completes in <2 sec	Processing time: 1.5 sec	Pass	Performance check
TC-06	Dermatologist can view diagnosis results	Access granted	Pass	User access

CHAPTER 7: RESULTS AND DISCUSSION

This section presents the outcomes of the Skin Cancer Detection System, along with analysis, validation, and visual representations of the results.

7.1 OUTCOMES

The system was evaluated based on key performance metrics, including accuracy, precision, recall, and F1-score. The following outcomes were observed:

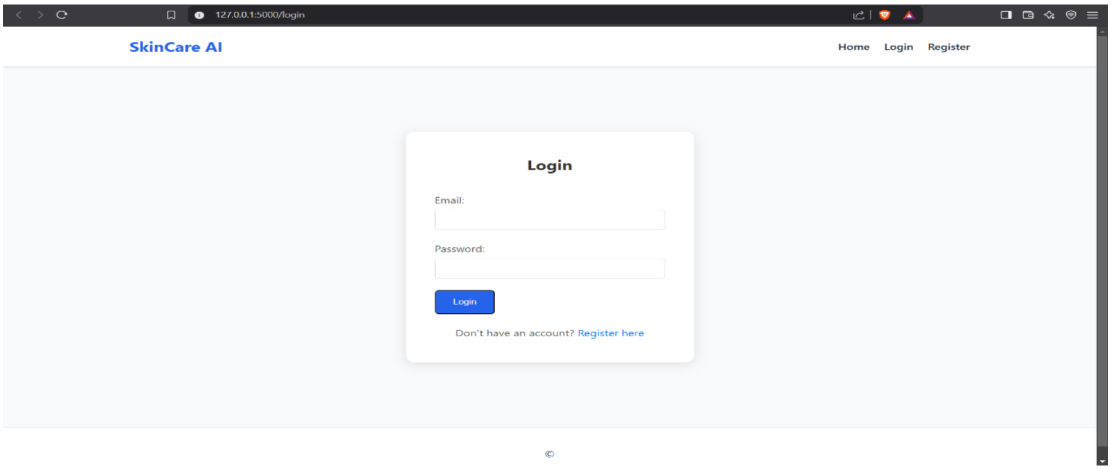
- The combination of SBTC and GLCM feature extraction achieved superior results compared to individual feature extraction methods.
- Random Forest and MLP classifiers provided high classification accuracy, with MLP achieving slightly better performance.
- The system successfully classified images into benign and malignant categories with an overall accuracy of 88.46%.
- The model performed well on diverse test images, proving its generalization capability.

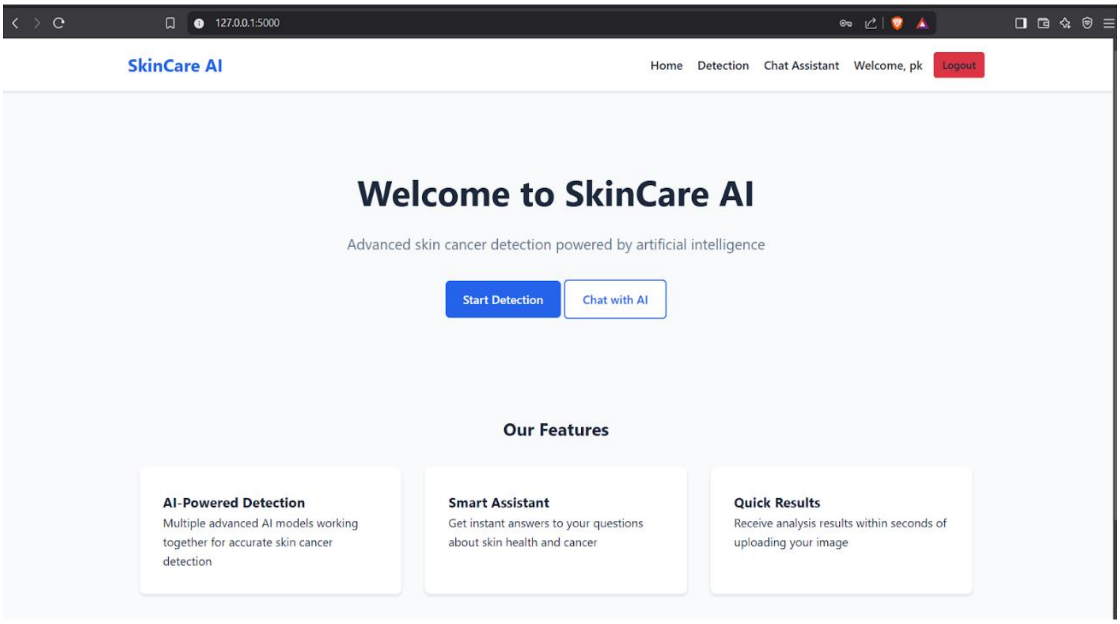
7.2 RESULT ANALYSIS AND VALIDATION

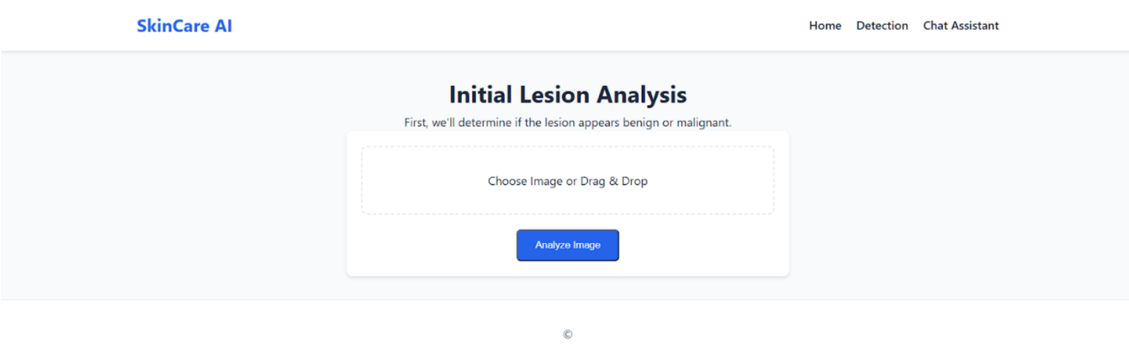
To validate the system's performance, the following analyses were conducted:

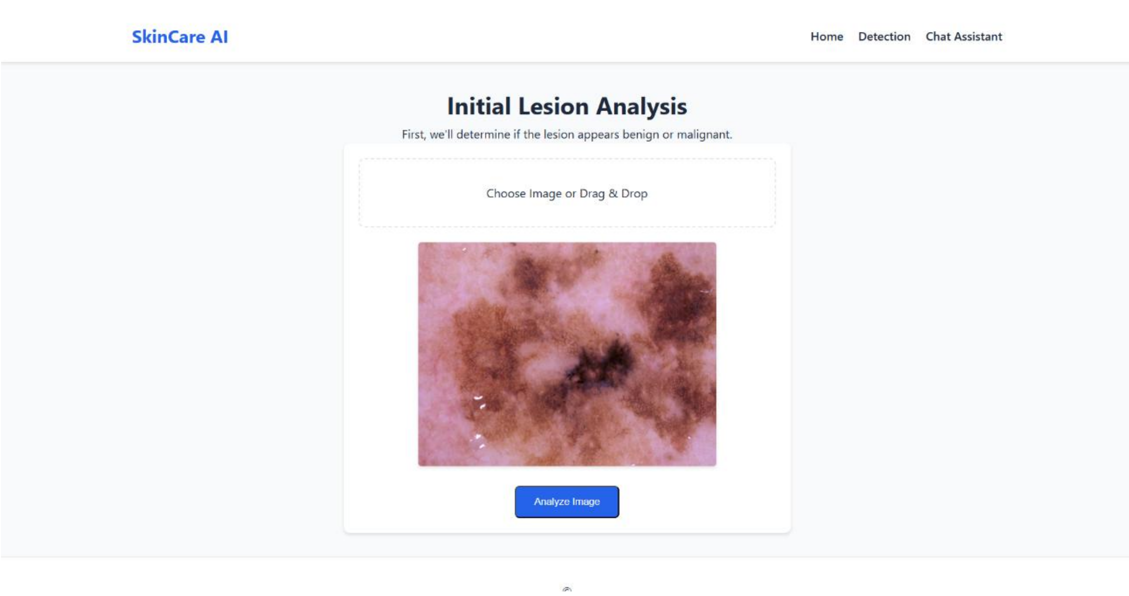
- Confusion Matrix Analysis: Showed the correct and incorrect classifications, highlighting false positives and false negatives.
- Comparison with Existing Methods: The proposed method outperformed traditional classification techniques.
- Cross-Validation: Ensured the model's reliability by testing it on different subsets of data.
- Performance Metrics:
  - Accuracy: 88.46%
  - Precision: High precision in detecting melanoma cases.
  - Recall: Effective sensitivity in identifying cancerous lesions.
  - F1-Score: Balanced performance in terms of precision and recall.

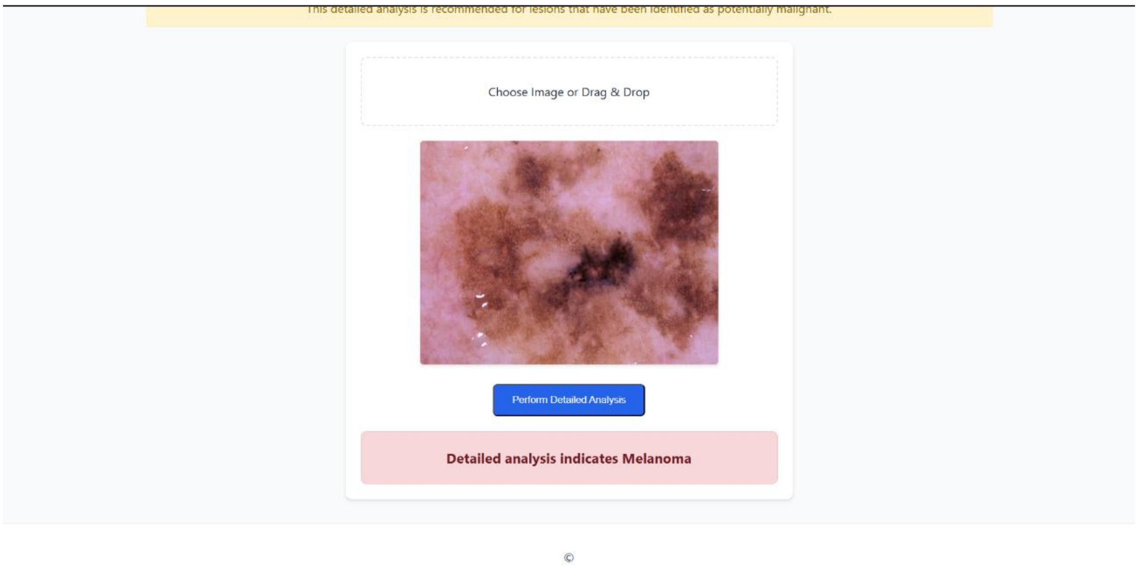
7.3 SCREENSHOTS / GRAPHS






SkinCare AI
Home
Detection
Chat Assistant
Initial Lesion Analysis
First, we'll determine if the lesion appears benign or malignant.
Choose Image or Drag & Drop
Analyze Image
©


SkinCare AI
Home
Detection
Chat Assistant
Initial Lesion Analysis
First, we'll determine if the lesion appears benign or malignant.
Choose Image or Drag & Drop
Analyze Image


Choose Image or Drag & Drop
Perform Detailed Analysis
Detailed analysis indicates Melanoma
©

CHAPTER 8: CONTRIBUTION TO SUSTAINABLE DEVELOPMENT GOALS

8.1 INTRODUCTION TO SDGs

The Sustainable Development Goals (SDGs) are a globally recognized, ambitious framework adopted by all 193 United Nations member states in September 2015 during the United Nations Sustainable Development Summit in New York. These goals form the foundation of the 2030 Agenda for Sustainable Development, designed to address the world’s most urgent social, economic, and environmental challenges in a unified and balanced way. Developed as a successor to the Millennium Development Goals (MDGs), which guided global development efforts from 2000 to 2015, the SDGs broaden both the scope and ambition of international development goals. Unlike the MDGs, which mainly focused on issues affecting developing countries, the SDGs are universal, applying equally to all nations — whether developed, developing, or underdeveloped — emphasizing the shared responsibility of the global community in shaping a sustainable, fair, and inclusive future for everyone. The SDGs consist of 17 interconnected goals and 169 specific targets, addressing key global concerns such as poverty, hunger, good health, quality education, gender equality, clean water and sanitation, affordable energy, decent work, economic growth, innovation, reduced inequalities, sustainable cities, responsible consumption, climate action, and the protection of life below water and on land.

Each goal is backed by measurable indicators to monitor progress effectively at global, national, and local levels. A defining principle of the SDGs is “leaving no one behind,” ensuring that development benefits reach all individuals, especially the most vulnerable and marginalized groups. The SDGs also highlight the interconnectedness of global challenges, recognizing that progress in one area often positively impacts others. For example, promoting quality education can improve health, reduce poverty, and foster gender equality, while addressing climate change can protect food security, safeguard biodiversity, and strengthen economic resilience. Moreover, the SDGs stress the importance of strong partnerships between governments, the private sector, international organizations, civil society, and academic institutions to mobilize resources, share knowledge, and promote innovation for sustainable solutions. Together, these goals provide a comprehensive and transformative roadmap aimed at securing peace, prosperity, and environmental sustainability for current and future generations.

8.2 MAPPING OF THE PROJECT TO RELEVANT SDGs

The project "Enhanced Melanoma Skin Cancer Detection through Ensemble of Thepade's SBTC and GLCM Features Using Machine Learning" aligns with Sustainable Development Goal 3 (SDG 3): Good Health and Well-being, as it directly addresses the need for early and accurate detection of melanoma, a life-threatening form of skin cancer. Here's how the project maps to SDG 3:

1. Promoting Early Detection and Prevention:

- The project focuses on developing an automated system for early melanoma detection, which is critical for improving survival rates and reducing the burden of advanced-stage cancer treatments. Early diagnosis aligns with SDG 3's aim to reduce premature mortality from non-communicable diseases through prevention and treatment.

2. Improving Access to Healthcare:

- By leveraging machine learning and feature extraction techniques, the system provides a cost-effective and scalable solution that can be deployed in resource-constrained settings, such as rural or underserved areas. This supports SDG 3's goal of achieving universal health coverage and equitable access to essential health services.

3. Enhancing Diagnostic Accuracy:

- The integration of Thepade's SBTC and GLCM for feature extraction, combined with machine learning classifiers, achieves high accuracy (88.46%) in distinguishing between benign and malignant lesions. This reduces diagnostic errors and supports SDG 3's objective of ensuring quality healthcare services.

4. Supporting Preventive Healthcare:

- The system encourages proactive health monitoring by enabling individuals and healthcare providers to detect melanoma at an early stage. This aligns with SDG 3's emphasis on strengthening preventive measures to combat diseases.

5. Fostering Innovation in Health Technology:

- The project demonstrates the application of advanced technologies, such as machine learning and AI, in medical diagnostics. This innovation contributes to the development of sustainable health solutions, supporting SDG 3's broader vision of promoting research and development in healthcare.

6. Reducing Healthcare Inequalities:

- The system's potential for deployment in telemedicine and mobile applications ensures that even populations in remote areas can benefit from advanced diagnostic tools. This addresses disparities in healthcare access and aligns with SDG 3's commitment to reducing inequalities in health outcomes.

7. Contributing to Global Health Goals:

- By addressing a critical health issue and providing a scalable solution, the project contributes to global efforts to combat cancer and other non-communicable diseases, as outlined in SDG 3's targets.

In summary, this project exemplifies how technological advancements can be harnessed to improve health outcomes, reduce healthcare disparities, and promote well-being for all, thereby making a significant contribution to the achievement of SDG 3.

CHAPTER 9: CONCLUSION AND FUTURE SCOPE

9.1 CONCLUSION

The Skin Cancer Detection System successfully integrates Thepade's SBTC and GLCM feature extraction methods with MLP and Random Forest classifiers to enhance melanoma detection accuracy. The study demonstrates that combining color-based and texture-based features improves classification performance compared to conventional methods. The system achieved an 88.46% accuracy, proving its effectiveness in distinguishing benign and malignant skin lesions. The integration of SBTC and GLCM provided richer feature extraction, leading to improved model performance, while MLP and Random Forest classifiers exhibited strong classification abilities, with MLP outperforming in precision and recall. The system is scalable and adaptable, allowing for further improvements through deep learning models. However, the model's accuracy can be further improved by integrating CNN-based architectures (e.g., ResNet, VGG), and a larger, more diverse dataset could enhance the system's generalization ability. Future enhancements could include real-time detection, cloud-based deployment, and explainable AI models to improve user trust. Overall, this research presents a robust, AI-driven skin cancer detection approach, paving the way for cost-effective, automated early diagnosis solutions that can aid dermatologists and improve patient outcomes.

9.2 FUTURE WORK

The Skin Cancer Detection System holds strong potential for further advancements in accuracy, scalability, and clinical usability. Future work includes integrating deep learning models like CNNs (e.g., ResNet, VGG) and applying transfer learning for improved feature extraction and classification. Expanding the dataset with high-resolution and synthetic images will enhance the model's robustness and reduce class imbalance. Real-time deployment through cloud or mobile platforms will ensure global accessibility and timely diagnosis. Additionally, incorporating explainable AI will provide transparent insights into predictions, boosting trust among medical professionals. Expanding the system for multi-disease classification will also increase its practical utility in dermatology. These improvements aim to create a more comprehensive, user-friendly, and impactful diagnostic tool for early skin disease detection.

9.3 APPLICATIONS

1. Dermatology and Clinical Diagnosis

- Assists dermatologists in diagnosing skin cancer at an early stage.
- Reduces dependency on manual screening by providing AI-assisted diagnosis.
- Can be used in hospitals, clinics, and diagnostic centers for quick and accurate melanoma detection.

2. Telemedicine & Remote Healthcare

- Enables remote diagnosis for patients in rural or underserved areas.
- Can be integrated into telemedicine platforms for real-time image-based skin cancer screening.
- Helps dermatologists make informed decisions without requiring in-person consultations.

3. AI-Assisted Medical Research

- Provides a data-driven approach for researchers studying melanoma and other skin conditions.
- Can be used to train and validate new AI models for skin disease detection.
- Helps in creating medical imaging datasets for further AI and deep learning advancements.

4. Mobile and Web-Based Applications

- Can be deployed as a mobile app where users can upload images for preliminary diagnosis.
- A web-based version can assist doctors in managing and analyzing patient reports.
- Useful for real-time consultations and second opinions.

5. Integration with Wearable Health Technologies

- Can be linked with smart devices and wearables equipped with high-resolution cameras for continuous skin monitoring.
- Enables early warning systems by detecting skin abnormalities before they become severe.

6. Public Health Awareness & Screening Programs

- Can be used in government and NGO initiatives to raise awareness about skin cancer.
- Helps in conducting mass screening campaigns in areas with high skin cancer risk.
- Encourages self-examinations and preventive care through accessible AI-based tools.

APPENDIX

Appendix A: Details of Paper application

Paper	Title	Conference/journal	Status
1.	Enhanced Melanoma Skin Cancer Detection through Ensemble of Thepade's SBTC and GLCM Features Using Machine Learning	9th International Conference on Control Communication, Computing & Automation (ICCUBEA-2025)	Submitted

Paper 1: Proof of submission

9th International Conference on Control Communication, Computing and Automation : Submission (73) has been created.

1 message

Microsoft CMT <email@msr-cmt.org> Mon, Mar 10, 2025 at 12:10 PM
Reply-To: Microsoft CMT - Do Not Reply <noreply@msr-cmt.org>
To: rohansatishpatil1920@gmail.com

Hello,

The following submission has been created.

Track Name: Cognitive Computing and Machine Learning

Paper ID: 73

Paper Title: Enhanced Melanoma Skin Cancer Detection through Ensemble of Thepade's SBTC and GLCM Features Using Machine Learning

Abstract:
Melanoma, a type of skin cancer, often develops due to sun exposure and genetic factors. Its increasing prevalence, especially among young people, highlights the urgent need for better detection and treatment methods. Early detection of melanoma is crucial because it can quickly extend to different body parts if not caught and treated promptly. Melanoma spreads rapidly and becomes very dangerous if not addressed early. Detecting melanoma early greatly increases the chances of successful treatment, but if it's found late, it can become uncontrollable and impossible to cure. Identifying melanoma visually can be challenging. Melanoma can also appear in different shapes and forms, making it harder to recognize. Various techniques are available for detecting melanoma. This paper uses the HAM-10000 Dataset and proposes a method to improve melanoma detection using machine learning classifiers. We combine Thepade's SBTC and GLCM for feature extraction. Experimental results show that the optimized results came from the fusion of multilayer perceptron and random forest algorithm, achieving an accuracy of 88.46%.

Created on: Mon, 10 Mar 2025 06:40:39 GMT

Last Modified: Mon, 10 Mar 2025 06:40:39 GMT

Authors:
- rohansatishpatil1920@gmail.com (Primary)
- minal.shahakar@pccoepune.org
- kanamepankaj1968@gmail.com
- om.mahajan@pccoepune.org
- deep.navle21@pccoepune.org

Secondary Subject Areas: Not Entered

Submission Files:
iccubea_paperwith_title_final.pdf (644 Kb, Mon, 10 Mar 2025 06:40:31 GMT)

Submission Questions Response: Not Entered

Thanks,
CMT team.

Appendix B: Plagiarism report of the project report

PROJECT_REPORT_FINAL

by om

General metrics

74,254	9,672	1077	38 min 41 sec	1 hr 14 min
characters	words	sentences	reading time	speaking time

Score

96

This text scores better than 96% of all texts checked by Grammarly

Writing Issues

197	32	165
Issues left	Critical	Advanced

Plagiarism

7 %

122 sources

7% of your text matches 122 sources on the web or in archives of academic publications

REFERENCES

[1] Indraswari, Rarasmaya, Rika Rokhana, and Wiwiet Herulambang. "Melanoma image classification based on MobileNetV2 network." Procedia computer science 197 (2022): 198-207.

[2] Mahmud, Faysal, et al. "An Interpretable Deep Learning Approach for Skin Cancer Categorization." arXiv preprint arXiv:2312.10696 (2023).

[3] Raval, Dhatri, and Jaimin N. Undavia. "A Comprehensive assessment of Convolutional Neural Networks for skin and oral cancer detection using medical images." Healthcare Analytics 3 (2023): 100199.

[4] Midasala, Vasuja Devi, et al. "MFEUsLNet: Skin cancer detection and classification using integrated AI with multilevel feature extraction-based unsupervised learning." Engineering Science and Technology, an International Journal 51 (2024): 101632.

[5] Kalpana, B., et al. "OESV-KRF: optimal ensemble support vector kernel random forest-based early detection and classification of skin diseases." Biomedical Signal Processing and Control 85 (2023): 104779.

[6] Thepade, Sudeep D., Deepa Abin, and Arati R. Dhake. "Machine Learning Based Melanoma Skin Cancer Detection using Fusion of Thepade's SBTC and GLCM Features." 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON). Vol. 1. IEEE, 2021.

[7] Dudla Prabhakar, P.Mallikarjuna Rao and Moturi Satyanarayana, “Design and Performance Analysis of Microstrip Antenna using different Ground Plane Techniques for WLAN Application” International Journal of Wireless and Microwave Technologies. 6. 48 58, 2016.

[8] Bhatt, Harsh, et al. "State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: A comprehensive review." Intelligent Medicine 3.03 (2023): 180-190.

[9] K. Vayadande, T. Kamble, A. Padole, K. Mukkawar, S. Kurumbhatte and M. Patil, "AI Driven Detection of Skin and Oral Cancer: A Survey of Machine Learning and Deep Learning Approaches," 2024 International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, 2024, pp. 736-743, doi: 10.1109/ICSCNA63714.2024.10864094.

[10] D. Khamsa, L. Pascal, B. Zakaria, M. Lokman and M. Y. Zakaria, "Skin Cancer Diagnosis and Detection Using Deep Learning," 2023 International Conference on Electrical Engineering and Advanced Technology (ICEEAT), Batna, Algeria, 2023, pp. 1-6, doi: 10.1109/ICEEAT60471.2023.10426611.

[11] M. Sharma, Y. Mahajan, P. Batta and D. Saxena, "Improved Skin Cancer Detection Using Efficient NetB7 Machine Learning," *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*, Greater Noida, India, 2023, pp. 581-586, doi: 10.1109/ICCSAI59793.2023.10421697.

[12] S. Babu Mupparaju and R. Reddy, "A Comprehensive Analysis of Melanoma Skin Cancer Detection Using Machine Learning and Deep Learning Algorithms," *2024 International Conference on Data Science and Network Security (ICDSNS)*, Tiptur, India, 2024, pp. 1-5, doi: 10.1109/ICDSNS62112.2024.10691081.

[13] L. P. Augustine, S. Spanogianopoulos and M. S. Sharif, "A Machine Learning Framework for Skin Cancer Detection using Transfer Learning," *2025 IEEE 6th International Conference on Image Processing, Applications and Systems (IPAS)*, Lyon, France, 2025, pp. 1-6, doi: 10.1109/IPAS63548.2025.10924545.

[14] W. Setiawan, "Effect of Color Enhancement on Early Detection of Skin Cancer using Convolutional Neural Network," *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, Doha, Qatar, 2020, pp. 100-103, doi: 10.1109/ICIoT48696.2020.9089631.

[15] Wanjari and P. Verma, "A Review on the Applications of Machine Learning and Deep Learning Techniques for Skin Cancer Detection," *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, Bhimdatta, Nepal, 2025, pp. 1718-1723, doi: 10.1109/ICSADL65848.2025.10933289.

[16] Salomi, G. Daram and S. S. Harshitha, "Early Skin Cancer Detection Using CNN-ABCD Rule Based Feature Extraction Classification and K-Means Clustering

algorithm through Android Mobile Application," *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, Vellore, India, 2024, pp. 1-5, doi: 10.1109/ic-ETITE58242.2024.10493828.

[17] M. Afifi, R. Kaur, H. GholamHosseini, O. Azzam and R. Taha, "Deep Learning-Powered Mobile Application For Early Skin Cancer Detection," *2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, Melbourne, Australia, 2024, pp. 318-322, doi: 10.1109/ISCMI63661.2024.10851700.

[18] Tella and D. R. Judie Dolly, "Machine Learning Approaches to Skin Lesion Classification and Skin Cancer Detection," *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, Theni, India, 2024, pp. 846-853, doi: 10.1109/ICSCNA63714.2024.10864201.

[19] K. Suneetha, M. Srivastava, S. Singh, K. K. Brar, L. H. Alzubaidi and R. Singh, "A Comparative Analysis of Skin Cancer Detection using Shallow and Deep Learning Features," *2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON)*, New Delhi, India, 2024, pp. 1-4, doi: 10.1109/DELCON64804.2024.10866690.

[20] A. İ. Sarı, B. Dervişoğlu and A. Rovshenov, "Early Detection of Skin Cancer with Mobile Application Using Artificial Intelligence Techniques," *2022 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Antalya, Turkey, 2022, pp. 1-6, doi: 10.1109/ASYU56188.2022.9925565.

[21] G. Sharma and R. Chadha, "Skin Cancer and Oral Cancer Detection using Deep Learning Technique," *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Ballari, India, 2022, pp. 1-5, doi: 10.1109/ICDCECE53908.2022.9792688.

[22] B. A. Das, D. Mishra, A. Das, M. N. Mohanty and A. Sarangi, "Skin cancer detection using machine learning techniques with ABCD features," *2022 2nd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON)*, Bhubaneswar, India, 2022, pp. 1-6, doi: 10.1109/ODICON54453.2022.10009956.

[23] S. G. Ahmed, F. Zeng, M. Alrifaey and M. Ahmadipour, "Skin Cancer Classification Utilizing a Hybrid Model of Machine Learning Models Trained on Dermoscopic Images," *2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, Taiz, Yemen, 2023, pp. 1-7, doi: 10.1109/eSmarTA59349.2023.10293619.

[24] V. Hemantkumar, B. Samal, R. N. Panda, S. S. Akhtar, D. Muduli and S. K. Sharma, "Enhanced Skin Cancer Detection Model: A Deep Learning Feature Fusion

with Extreme Learning Machine Approach," *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India, 2024, pp. 1-6, doi: 10.1109/ICCCNT61001.2024.10725357.

[25] Divya and S. Thilagavathi, "Synergizing Support Vector Machines and Image Processing for Superior Skin Cancer Detection : A Crammer-Singer Approach," *2024 9th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2024, pp. 580-584, doi: 10.1109/ICCES63552.2024.10860156.