# A Machine Learning Driven Clinical Decision Support System for Skin Cancer Detection Using Deep Learning and Ensemble Methods

***Dr. Ketki C Pathak***
Electronics and communication Department,
Sarvajanik College of Engineering and Technology, Surat, India

***Mahesuriya Fenil Rameshkumar***
Information and Communication Technology Department,
Adani University,
Ahmedabad, India

***Abstract— This paper presents a comprehensive machine learning-driven clinical decision support system for automated skin cancer detection and classification. Our approach leverages deep learning feature extraction using pre-trained ResNet50 architecture combined with ensemble machine learning classifiers including Support Vector Machine (SVM) and Random Forest (RF). The system integrates dermoscopic image analysis with patient metadata (age, gender, lesion location) and employs advanced techniques such as Principal Component Analysis (PCA) for dimensionality reduction and Synthetic Minority Oversampling Technique (SMOTE) for class imbalance handling. Results demonstrate promising accuracy rates with Random Forest achieving 98% accuracy and SVM achieving 97% accuracy on balanced multi-class skin lesion classification. The system incorporates explainable AI techniques including Grad-CAM, SHAP, and LIME to provide interpretable predictions, making it suitable for clinical deployment as a decision support tool for dermatologists.***

## **1.** Introduction

The word "cancer" has its roots originating from "καρκίνος" which means 'crab' or 'tumor' in the ancient Greek language [1]. Cancer was first introduced to the world of medical sciences in 1600 BCE. Research into the causes of cancer and its diagnosis, treatment, cure, and prevention are cancer research and the branch of medicine that deals with cancer is called Oncology. When normal cells undergo a series of mutations, they may become cancerous; these cancerous cells grow and divide continuously leading to the formation of solid lumps or tumors [3]. They crowd normal cells by growing out of control instead of dying when they should, making it difficult for the body to function properly. Though abnormal tissue growth may be cancerous, not all tissue growths turn out to be cancer.

According to the World Health Organization (WHO), over 2-3 million non-melanoma skin cancers and 1,32,200 melanoma skin cancers occur worldwide every year.

Over the past decades, continuous evolution related to cancer research has been performed. Scientists applied different methods, such as screening in early stage, in order to find types of cancer before they cause symptoms. Moreover, they have developed new strategies for the early prediction of cancer treatment outcome. Cancer is a fatal illness frequently caused by a variety of obsessive changes and genetic disorders. Cancer cells known as abnormal cells can grow in any part of the human body. A preliminary diagnosis of cancer is necessary as cancer is one of the most alarming diseases. Detecting cancer and treating it in the initial stage can decrease the death rate.

In recent years, the integration of Machine Learning (ML) in healthcare has revolutionized the way clinical decisions are made. One of the most impactful applications of this integration is the development of Clinical Decision Support Systems (CDSS) powered by ML algorithms.

Recent advances in machine learning (ML) and deep learning (DL) have demonstrated enormous promise in the medical field, particularly in image-based diagnostics. These technologies are capable of learning complex patterns from large datasets,

allowing for expert-level automated classification of skin lesions.

A Machine Learning Driven Clinical Decision Support System (CDSS) for skin cancer detection leverages these capabilities to assist clinicians in identifying malignant lesions—such as melanoma and basal cell carcinoma, and squamous cell carcinoma—more accurately and efficiently.

Such a system typically integrates dermoscopic image analysis with patient metadata (e.g., age, sex, lesion location) and employs robust models like Convolutional Neural Networks (CNNs)- ResNet50, Principal Component Analysis (PCA), Logistic Regression, Support Vector Machines (SVMs), and Random Forests. Moreover, to ensure transparency and clinical trust, explainable AI techniques such as Synthetic Minority Oversampling Technique (SMOTE) for class imbalance, Grad-CAM, SHAP, and LIME are incorporated, offering visual and feature-based justifications for model predictions.

Stage 1 and Stage 2 cancers are usually treated with radiation therapy and local surgery. Earlier, cancer treatment mainly involved using and upgrading surgical techniques for removing tumors while radiation therapy took a rise in the 1900s. In some cases, other drugs and chemotherapy may be used. A proper diet consisting of vegetables, fruits, and fish is also recommended for a speedy recovery from cancer whereas red meat and processed food should be avoided. Sometimes, cancer which is thought to be cured comes back after many years. This is why doctors say that cancer is in remission; it means that there are very little to no symptoms of cancer [1].

Cancer at stage 3 means that it has locally advanced whereas stage 4 cancer means that cancer has already spread to distant parts of the body through the bloodstream or lymphatic system. The grave nature of cancer cases makes it important to accurately diagnose them to increase the chances of survival.

The present work introduces a comprehensive clinical decision support system for skin cancer detection comprising three key innovations:

- **Deep Learning Feature Extraction**: Utilizing pre-trained ResNet50 architecture to extract high-level visual features from dermoscopic images, capturing texture, shape, and color patterns indicative of malignancy.
- **Ensemble Machine Learning Classification**: Implementing multiple ML algorithms (SVM, Random Forest, Logistic Regression) with advanced preprocessing techniques including PCA dimensionality reduction and SMOTE class balancing to handle real-world data challenges.
- **Explainable AI Integration**: Incorporating interpretability techniques such as Grad-CAM, SHAP, and LIME to provide visual and quantitative explanations for model predictions, ensuring clinical trust and transparency.

This integrated approach addresses the critical need for accurate, scalable, and interpretable skin cancer detection systems that can assist dermatologists in clinical decision-making while being deployable in various healthcare settings from specialized clinics to remote screening programs.

## 2. Literature Review

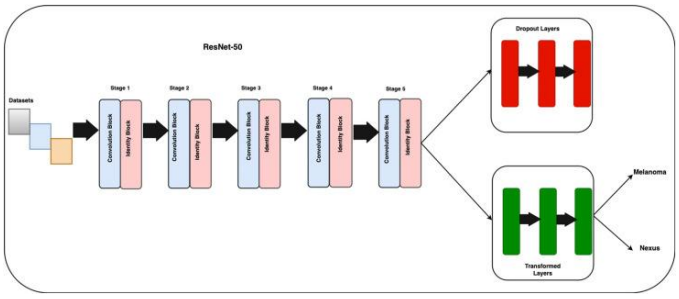
Multiple investigations have been done in cancer research. A large number of research works based on ML and DL methods have been proposed in the literature to predict the risk of various types of cancers. Most researchers have used standard databases to predict cancer in their work. In their experimental study, authors achieved a higher degree of prediction accuracy and other metrics by using different ML and DL algorithms.

A Systematic Review of Artificial Intelligence Techniques in Cancer Prediction and Diagnosis [1] show us a detailed examination of the machine learning and deep learning models which are used in medical imaging to detect cancer at early stages. AI-Based techniques which use machine learning and deep learning to extract the disease features and

classify them are proven to be helpful in early cancer prognosis and detection.

Cancer is caused by the growth of cells that have the ability to spread to other parts of the body [1]. Skin cancer is a highly malignant and perilous form of cancer. The only time skin cancer can be cured is when it is diagnosed in its early stages. The skin envelops all anatomical structures of the body, such as muscles and bones, and is indispensable to the human body. A minor modification in the functioning of the skin has a significant influence on every system in the body, establishing it as a prominent participant. A lesion area refers to the affected region of the skin. There is a wide range of different types of skin lesions. Each lesion is categorized into categories according to the specific type of skin cells from which it originated. Melanocytes generate melanocytic lesions, which bear resemblance to melanoma. Melanin is a protein pigment synthesized by melanocytes [4].

This study describes skin cancer, and the challenges associated with its early detection. It explores the application of deep learning techniques, focusing on quantitative methods like few-shot learning and Convolutional Neural Networks (CNNs) to enhance accuracy. While a mixed-method strategy, combining both qualitative and quantitative approaches, could theoretically offer insights into the practical implementation and real-world challenges, the study focuses purely on quantitative methods due to the nature of machine learning models and the necessity for large-scale data-driven approaches.




Utilizing deep learning techniques for the identification of skin cancer. While deep learning methods like CNNs have been successfully applied to skin cancer detection, they often struggle with the imbalance in skin lesion datasets, particularly in detecting rare but critical conditions like melanoma. Furthermore, their reliance on large datasets makes them less applicable in scenarios where data is scarce. This research addresses these shortcomings by utilizing GANs to generate synthetic data, improving model performance even in the face of small or imbalanced datasets.

Deep learning algorithms play a vital role in identifying and diagnosing skin cancer. Their nodes are connected. The neural interconnection bears similarity to the structure of the brain. Together, their nodes work together to solve problems. Through the process of training neural networks for certain activities, they develop a high level of proficiency in those particular domains.

In recent years, extensive research has explored the use of machine learning (ML) and deep learning (DL) in cancer detection, particularly in skin cancer, due to its high mortality if not identified early. Studies consistently demonstrate that AI-based models, especially CNNs and few-shot learning techniques, significantly improve diagnostic accuracy by automatically extracting key features from medical images. Standard datasets like HAM10000 have facilitated benchmarking and reproducibility in these investigations. While traditional diagnostic methods such as biopsy are time-intensive and invasive, AI-driven tools offer a faster, more accessible, and non-invasive alternative, particularly valuable in low-resource or remote settings. Despite challenges such as image artifacts and class imbalance, ML and DL models have shown promising results in differentiating malignant from benign lesions using dermoscopic features. Overall, these advancements highlight the transformative potential of AI in supporting early and reliable skin cancer diagnosis, aligning with global healthcare goals.

## 3. Problem Definition and Scope

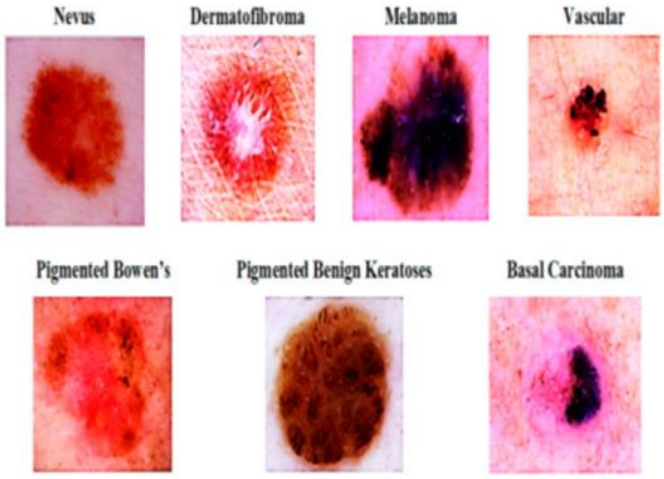
Cancer remains one of the leading causes of death worldwide, with early and accurate detection being critical for improving patient outcomes. However, traditional diagnostic methods—such as imaging analysis, biopsies, and lab reports—are often time-

consuming, prone to human error, and vary in accuracy across institutions and clinicians. Furthermore, the heterogeneity of cancer across types (e.g., breast, lung, skin, cervical, prostate) and within subtypes poses a significant challenge to standardized diagnosis. Cancer is a major global health challenge, emphasizing the critical need for early detection to enhance patient outcomes.

Skin cancer is one of the most prevalent forms of cancer worldwide, with rising incidence rates attributed to factors such as increased ultraviolet (UV) exposure, genetic predisposition, and environmental conditions. Early and accurate diagnosis plays a pivotal role in improving patient survival rates and minimizing invasive treatments. However, visual inspection by dermatologists, even when assisted by dermoscopic imaging, can be subjective and limited by human variability. Consequently, there is a pressing need for scalable, objective, and high-performance diagnostic tools to augment clinical decision-making.

Skin cancer is one of the most dangerous and fastest growing skin disease. So, the treatment for skin cancer should be done in the early stages. One of the skin cancer types is ‘Melanoma’ skin cancer which has a high death rate and treatment for this kind of cancer is costly. Manual detection of Melanoma cancer cells takes time to cure.

Here, the employed machine learning techniques to analyze parameters, such as cell size, shape, nucleus characteristics, and additional factors like cell texture, mitosis count, tumor progression, metastasis, gene expression patterns, and biological markers. This methodology is distinguished by its effective use of diverse data types and automated feature extraction to improve cancer detection and prediction accuracy. Advanced machine learning methods enhance the precision and reliability of current cancer cell classification algorithms. The research underscores the importance of timely and accurate cancer detection, which enables early intervention and significantly improves patient survival rates. The results and discussion section meticulously analyzes the findings, demonstrating the approach's effectiveness in accurately identifying cancerous cells and assessing cancer severity [2].




| Lesion Category | Label | Full Name |
|---|---|---|
| 0 | akiec | Actinic keratoses and intraepithelial carcinoma |
| 1 | bcc | Basal cell carcinoma |
| 2 | bkl | Benign keratosis-like lesions |
| 3 | df | Dermatofibroma |
| 4 | mel | Melanoma |
| 5 | nv | Melanocytic nevi |
| 6 | vasc | Vascular lesions |

- Key areas within the scope include:
    - **Automated Diagnosis**: ML models can identify and classify skin cancer types from images with performance comparable to expert dermatologists.

- **Early Detection**: By recognizing subtle patterns in early-stage lesions, ML enables timely intervention, increasing survival rates.
- **Feature Extraction & Analysis**: Advanced techniques (e.g., CNNs, PCA, Grad-CAM, SHAP) can learn visual and clinical features such as color, texture, border irregularities, and metadata (age, gender, location).
- **Handling Class Imbalance**: Techniques like SMOTE address the challenge of imbalanced datasets common in skin cancer diagnosis.
- **Scalability and Remote Screening**: ML models can be deployed in mobile apps or web platforms to assist in rural or under-resourced regions.
- **Integration with Clinical Workflows**: ML tools can act as decision support systems to reduce diagnostic error and enhance dermatologists' productivity.

Overall, the scope of ML in skin cancer detection spans **image preprocessing, lesion segmentation, classification, metadata integration, and explainability**, with the goal of improving diagnostic accuracy, consistency, and accessibility.

## 4. Methodology

Here is a comprehensive **Methodology for Skin Cancer Detection using Machine Learning**, the methodology for skin cancer detection using machine learning involves a systematic approach comprising data collection, preprocessing, feature extraction, model training, evaluation, and explainability. The goal is to build a reliable and interpretable model that can classify skin lesions accurately.

**1. Data Collection**

- **Datasets Used**: Publicly available dermoscopic image datasets like HAM10000, ISIC Archive, PH², Derm7pt, and BCN20000.
- **Metadata**: Patient information such as age, gender, and lesion location is collected when available to enhance prediction accuracy.

**2. Data Preprocessing**

- **Image Resizing**: Images are resized to a fixed size (e.g., 128x128, 224×224 or 256×256) for model input consistency.
- **Label Encoding**: Skin lesion types (e.g., melanoma, nevus) are encoded into numerical format.

**3. Handling Class Imbalance**

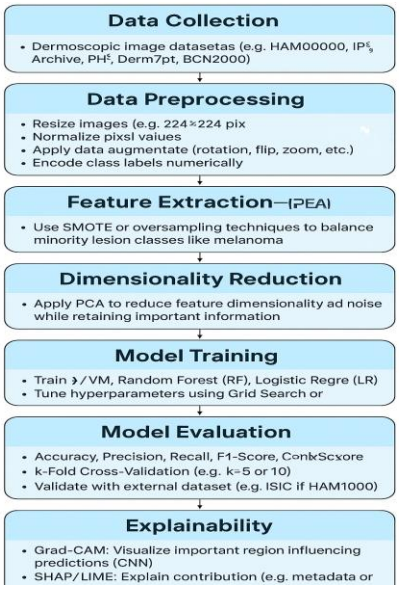
- **SMOTE (Synthetic Minority Oversampling Technique)** or **Random Oversampling** is used to address the imbalance among classes (e.g., underrepresented melanoma cases).

**4. Feature Extraction**

- **Deep Learning Features**: Pre-trained CNN models (e.g., ResNet50, VGG16, Efficient Net) are used to extract deep image features via transfer learning.
- **Handcrafted Features** *(optional)*: Shape, texture (GLCM), and color features can be extracted using OpenCV and skimage.

**5. Dimensionality Reduction**

- **PCA (Principal Component Analysis)** is applied to reduce high-dimensional feature vectors while preserving important variance.



**6. Model Selection and Training**

- **Classifiers Used**:
    - Support Vector Machine (SVM)
    - Random Forest (RF)
    - *(Optional: kNN, XGBoost, or Neural Networks)*
- **Hyperparameter Tuning**: Grid Search or Random Search is used to optimize model parameters.

**7. Model Evaluation**

- **Metrics Used**: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.
- **Cross-Validation**: K-fold cross-validation (typically k=5 or 10) ensures robustness of results.
- **External Validation**: Model is tested on a different dataset (e.g., ISIC or PH²) to evaluate generalizability.

**8. Explainability**

- **Grad-CAM**: Visual explanation highlighting important regions of the image for CNN-based models.
- **SHAP and LIME**: Used to explain predictions based on input features and identify feature importance.

**Support Vector Machine (SVM)**

- Support Vector Machine is a supervised learning algorithm primarily used for classification.
- **Linear SVM**: Finds a straight line (in 2D) or hyperplane (in nD) to separate two classes.
- **Non-linear SVM**: Uses **kernel tricks** (like RBF, polynomial) to map data into a higher-dimensional space where it becomes linearly separable.

## Advantages

- High **accuracy** and **recall** for binary classification.
- Works well in high-dimensional feature spaces.
- Good generalization when kernel and parameters are tuned correctly.

**Disadvantages**

- Not ideal for large datasets (computationally intensive).
- Requires careful hyperparameter tuning (kernel, C, gamma).
- Struggles with heavily imbalanced datasets unless paired with techniques like SMOTE.

**Random Forest (RF)**

- Random Forest is an ensemble learning method that builds a collection of decision trees, then aggregates their outputs (via majority vote for classification).
- **Bootstrap Aggregation (Bagging)**: Multiple decision trees are trained on random subsets of the training data with replacement.
- **Feature Randomness**: At each split in a tree, only a random subset of features is considered.
- **Ensemble Voting**: Each tree votes for a class, and the majority vote determines the final prediction.

**Advantages**

- Handles **large datasets** and **noisy data** well.
- Less prone to **overfitting** than a single decision tree.
- Provides **feature importance**, which helps in explainability.

**Disadvantages**

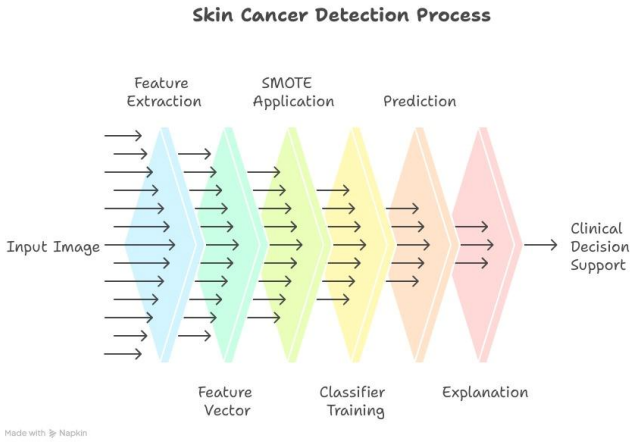
- May not achieve the same precision as SVM in finely separated binary cases.
- Slower training and inference compared to simpler models.
- Less interpretable than individual decision trees.

## 5. Implementation and Results

**Step 1: Environment Setup and Memory Management**

Purpose: Initialize the system with proper memory management to handle large image datasets efficiently.
Memory Management Functions:

- get_memory_usage (): Returns current memory percentage
- clear_memory (): Runs garbage collection and clears TensorFlow session




**Step 2: Metadata Loading and Preprocessing**

1. Load CSV file containing image metadata

**Step 3: Batch Image Processing**

**Image Preprocessing**: Resize to 128×128 pixels
**Metadata Integration**: Extract age, gender, and localization features

**Step 4: Deep Feature Extraction**

**Purpose**: Extract high-level visual features using a pre-trained ResNet50 convolutional neural network.

**Code:** resnet = ResNet50(weights='imagenet', include_top=False, pooling='avg', input_shape=(128, 128, 3))
features = resnet.predict(X_img, verbose=1, batch_size=32)

**Step 5: Feature Combination and Scaling**

**Purpose**: Combine deep learning features with patient metadata to create comprehensive feature vectors.
**Code:** print("Combining features with metadata...")
scaler = StandardScaler()
X_meta_scaled = scaler.fit_transform(X_meta)
X = np.concatenate([features.reshape(features.shape[0], -1), X_meta_scaled], axis=1)

**Step 6: Dimensionality Reduction (PCA)**

**Purpose**: Reduce computational complexity while preserving important information patterns.
**Code:** pca = PCA(n_components=100, random_state=42)
X_reduced = pca.fit_transform(X)

**Step 7: Class Balancing with SMOTE**

**Purpose**: Address class imbalance in the dataset using Synthetic Minority Oversampling Technique.
**Code:** smote = SMOTE(random_state=42)
X_balanced, y_balanced = smote.fit_resample(X_reduced, y)

**Step 8: Train-Test Split**

**Purpose**: Divide data into training and testing sets while maintaining class proportions.
Training: 80% of balanced data
Testing: 20% of balanced data
**Code:** X_train, X_test, y_train, y_test = train_test_split( X_balanced, y_balanced, test_size=0.2, stratify=y_balanced, random_state=42)

**Step 9: Model Training and Evaluation**

**Purpose**: Train multiple machine learning models and compare their performance like random forest, svm.

**Step 10: Performance Visualization**

**Purpose**: Create comprehensive visual comparisons of model performance across all classes (confusion matrix).

**Step 11: Grad-CAM Explainability**

**Purpose**: Generate visual explanations showing which image regions influenced model predictions.
**Code:** axes[1].imshow(grid)
axes[1].set_title(f"Grad-CAM Heatmap (Class {predicted_class})")
axes[1].axis('off')
plt.tight_layout()
plt.show()

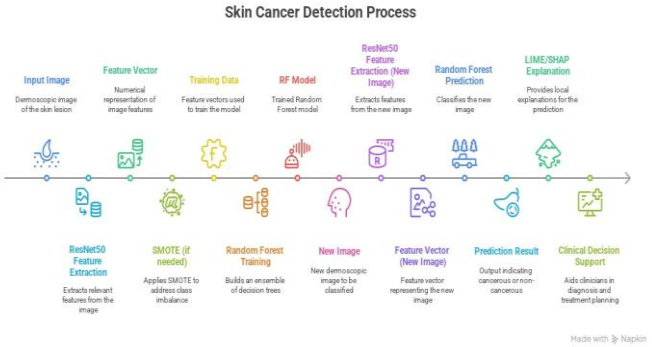
**Step 12: SHAP (SHapley Additive exPlanations)**

**Purpose**: Provide quantitative feature importance explanations based on game theory.

**Step 13: LIME (Local Interpretable Model-Agnostic Explanations)**

**Purpose**: Generate human-interpretable explanations for individual predictions.
**Code:** fig = explanation.as_pyplot_figure()

```
plt.title("LIME Feature Importance")
plt.tight_layout()
plt.show()
```




## 6. Results

**1) ResNet50 Deep Features + Metadata → 32,771 Features: -**

- used a pre-trained ResNet50 model (a **deep convolutional neural network**) to extract deep image features from each skin lesion image.
- These features are high-level, abstract representations learned by the network (e.g., texture, edges, lesion shape).

**2) PCA Reduced to 50 Principal Components: -**

- Having 32,771 features is computationally expensive and can lead to **overfitting**.
- So applied **Principal Component Analysis (PCA)** — a dimensionality reduction technique.
- Selected the **top 50 principal components**, reducing dimensionality while retaining significant information.

**3) PCA Top Variance Contribution: -**

- [0.084, 0.060, 0.054, 0.046, 0.030, …]
- This list shows how much **variance (information)** each principal component captures.

For example:

- ❖ PC1 explains **8.4%** of the total data variance,
- ❖ PC2 explains **6.0%**, and so on.

**4) Applied SMOTE to Fix Class Imbalance: -**

- The dataset likely had **unequal class distribution** (some lesion types had fewer samples).
- Used SMOTE (Synthetic Minority Oversampling Technique)
- After applying SMOTE, each of the 7 classes had **exactly 1,337 samples**, creating a **perfectly balanced dataset**.

**1) Random Forest:-**

- Random Forest Classification Report:

| Lesion Category | Precision | Recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 1.00 | 1.00 | 1341 |
| 1 | 0.98 | 1.00 | 0.99 | 1341 |
| 2 | 0.95 | 0.98 | 0.96 | 1341 |
| 3 | 1.00 | 1.00 | 1.00 | 1341 |
| 4 | 0.95 | 0.98 | 0.97 | 1341 |
| 5 | 0.97 | 0.89 | 0.93 | 1341 |
| 6 | 1.00 | 1.00 | 1.00 | 1341 |

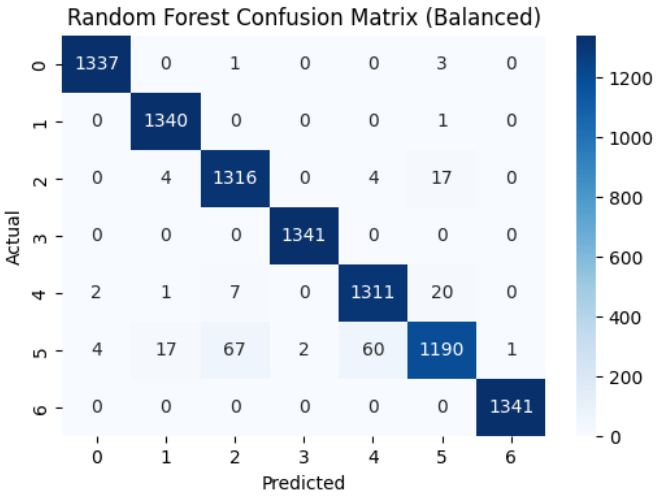
## Evaluation Metrics

- Precision: TP / (TP + FP)
- Recall(Sensitivity): TP / (TP + FN)
- F1-Score: Harmonic mean of Precision and Recall

| | Precision | Recall | f1-score | support |
|---|---|---|---|---|
| accuracy | - | - | 0.98 | 9387 |
| macro avg. | 0.98 | 0.98 | 0.98 | 9387 |
| weighted avg. | 0.98 | 0.98 | 0.98 | 9387 |

For Confusion Matrix: -

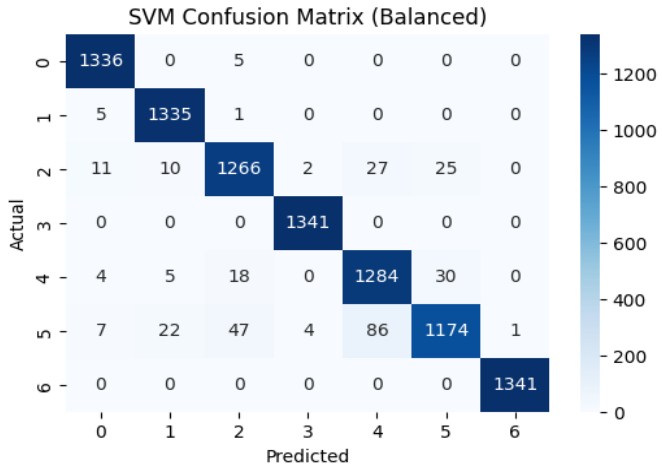
- Each **row** = actual (ground truth) class
- Each **column** = predicted class
- Diagonal values = correct predictions
- Off-diagonal values = misclassifications




## 2) Support Vector Machine (SVM):-

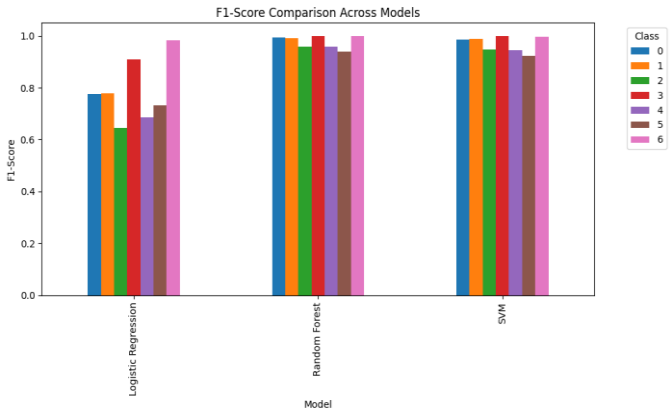
SVM Classification Report:

| Lesion Category | Precision | Recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.98 | 1.00 | 0.99 | 1341 |
| 1 | 0.97 | 1.00 | 0.98 | 1341 |
| 2 | 0.95 | 0.94 | 0.95 | 1341 |
| 3 | 1.00 | 1.00 | 1.00 | 1341 |
| 4 | 0.92 | 0.96 | 0.94 | 1341 |
| 5 | 0.96 | 0.88 | 0.91 | 1341 |
| 6 | 1.00 | 1.00 | 1.00 | 1341 |



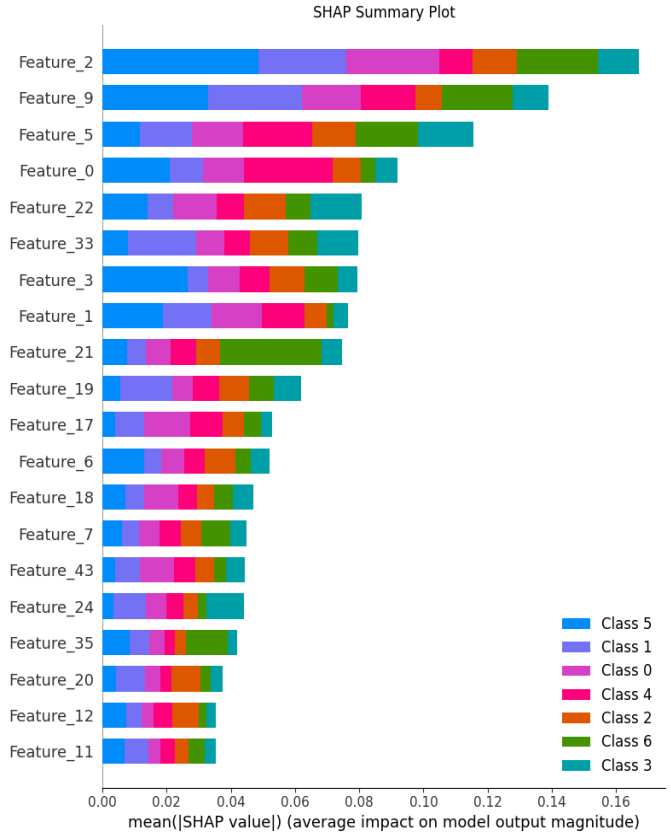

For Confusion Matrix: -

- Each **row** = actual (ground truth) class
- Each **column** = predicted class
- Diagonal values = correct predictions
- Off-diagonal values = misclassifications



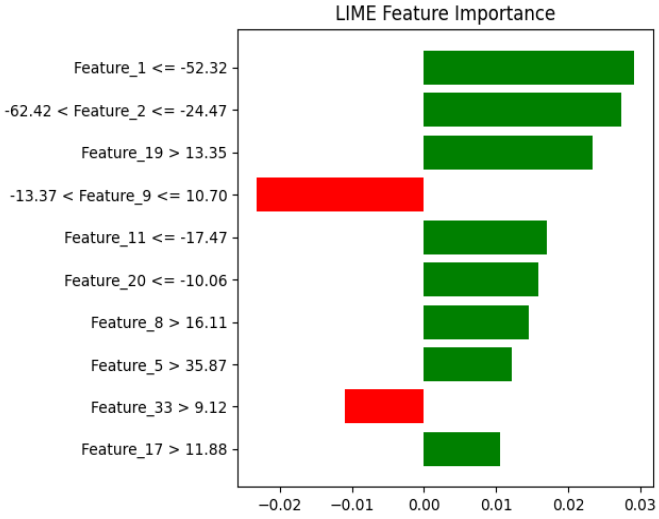

**SHAP Summary Plot (Global Explainability): -**

- All top 20 features shown in the image are **PCA components** of **ResNet50 deep features**, meaning your model primarily relying on **visual lesion characteristics** (e.g., shape, texture, border) learned by CNN.
- Top Influential Features: Feature_2, Feature_5, Feature_9
- SHAP value shows contribution **per class**
- Used for model transparency and fairness




**LIME Local Feature Importance: -**

- Just like with SHAP, all LIME-highlighted features are from **PCA-reduced deep features** (from ResNet50).
- Explains a **single prediction**
- Positive contribution (green) and negative (red)
- Top features: Feature_1, Feature_2, Feature_19




## 7. Challenges and Learnings

**1) Class Imbalance**

- **Issue:** Datasets like HAM10000 or ISIC often have an uneven distribution of lesion types (e.g., many benign, few melanoma).
- **Impact:** Models tend to be biased toward majority classes.
- **Solution:** Applied techniques like SMOTE, class weighting, and oversampling.

**2) Intra-Class Variability and Inter-Class Similarity**

- **Issue:** Some benign lesions look very similar to malignant ones and vice versa.
- **Impact:** Misclassification, lower precision/recall.

- **Solution:** Use of deep CNN features and dimensionality reduction (PCA) to capture fine details.

**3) Low Quality and Artifacts in Images**

- **Issue:** Dermoscopic images may contain hair, shadows, ruler markings, or poor lighting.
- **Impact:** Misleading features in training.
- **Solution:** Preprocessing: hair removal, contrast enhancement, resizing.

4) **Limited Metadata Use**

- **Issue**: Important features like age, gender, lesion location are often underutilized.
- **Impact**: Reduces the model's real-world diagnostic relevance.
- **Solution**: Integrating metadata with image features to improve accuracy.

5) **Overfitting Due to Small Data**

- **Issue**: DL models tend to overfit when data is insufficient.
- **Impact**: Poor generalization on unseen data.
- **Solution**: Applied data augmentation, transfer learning (e.g., ResNet50), and cross-validation.

6) **Explainability & Trust**

- **Issue**: ML/DL models are often "black boxes."
- **Impact**: Hard for clinicians to trust predictions.
- **Solution**: Implemented Grad-CAM, SHAP, LIME for interpretability.

**Key Learnings from the Project**

- Importance of Data Quality: High-resolution, well-annotated images are crucial, Preprocessing steps directly impact downstream performance.
- Hybrid Approaches Perform Better: Combining deep learning (for feature extraction) with traditional ML (like SVM, RF) gave better results than using either alone.
- Metadata Boosts Model Intelligence: Including clinical information (e.g., lesion location or patient age) improved recall and reduced false negatives.
- Explainability Is Not Optional: Visual explanations from Grad-CAM or SHAP increased trust and revealed important feature correlations.
- Evaluation Beyond Accuracy: Metrics like Recall, F1-Score, and ROC-AUC are more meaningful, especially for critical classes like melanoma.
- External Validation Is Essential: Testing on a separate dataset (e.g., train on HAM10000, test on $PH^2$) ensures the model is generalizable.

## 8.Conclusion

- The early and accurate detection of skin cancer remains a critical challenge in the field of medical diagnostics. Through this project, a machine learning-based approach was developed to classify skin lesions using dermoscopic images and patient metadata. Leveraging datasets such as HAM10000 and ISIC, a comprehensive pipeline was implemented involving image preprocessing, feature extraction, class balancing, model training, evaluation, and explainability.

- In conclusion, the developed skin cancer detection system demonstrates that machine learning, when applied thoughtfully and supported by clinical data and explainability techniques, can serve as a valuable decision-support tool for dermatologists. With further validation, real-time deployment, and integration into clinical workflows, such systems hold the potential to improve early detection rates and ultimately save lives.

- Looking ahead, future work can explore the development of a real-time mobile or web-based diagnostic tool to aid dermatologists and general practitioners in clinical and rural settings. Further improvement can be achieved through larger and more diverse datasets, multimodal inputs (e.g., combining clinical photos, pathology, and dermoscopy), and continuous learning frameworks. With regulatory validation and clinician

collaboration, this skin cancer detection model has the potential to be translated into a practical, life-saving solution in dermatological care.

## 9. Future Work

While this project demonstrates the feasibility and effectiveness of machine learning for skin cancer detection, there remains significant scope for future enhancement and real-world applicability. The following directions can further improve the model's performance, robustness, and clinical relevance:

**1) Integration of Multimodal Data**

- Future research can focus on incorporating multimodal inputs such as clinical images, patient history, genetic information, and histopathological data alongside dermoscopic images. This could lead to a more holistic analysis, improving the model's diagnostic accuracy and clinical utility.

**2) Improved Lesion Segmentation**

- Advanced segmentation models such as U-Net or Mask R-CNN can be employed to isolate lesions more accurately, removing surrounding noise and enhancing feature extraction. Segmentation-aware classification may improve predictions, especially for small or irregular lesions.

**3) Real-Time and Mobile Deployment**

- The model can be packaged into a mobile or web-based application for use by general physicians or patients. Real-time classification on smartphones, combined with cloud-based inference, could facilitate early self-screening and remote diagnosis in rural or underserved areas.

**4) Clinical Validation and Collaboration**

- Collaboration with dermatologists and clinical institutions is essential to evaluate the model's effectiveness in real-world diagnostic workflows. Conducting prospective studies or clinical trials can validate model performance, identify limitations, and guide regulatory approval.

## 10. References

Useful Website: -

- Cleveland Clinic,
- IBM
- NIH National Library of Medicine, U.S Gov.

References: -


1) A Systematic Review of Artificial Intelligence Techniques in Cancer Prediction and Diagnosis Yogesh Kumar, Surbhi Gupta, Ruchi Singla & Yu-Chen Hu

2) Machine learning applications in cancer prognosis and prediction Author links open overlay panelKonstantina Kourou [a], Themis P. Exarchos [a b], Konstantinos P. Exarchos [a], Michalis V. Karamouzis [c], Dimitrios I. Fotiadis [a b]

3) Enhancing cancer detection and prevention mechanisms using advanced machine learning approaches Author links open overlay panelKamta Nath Mishra [a], Alok Mishra [b], Soumya Ray [a], Anjali Ku mari [a], Saad Misbah Waris [a]

4) Skin cancer detection using deep machine learning techniques Author panelOlusoji Akinrinade, Chunglin Du

5) Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis by Mohammed Amine Naji, Sanaa El Filali, Kawtar Aarika, EL Habib Benlahmar, Rachida Ait Abdelouhahid, Olivier Debauche

6) Polytomous Logistic Regression Based Random Forest Classifier for Diagnosing Cancer Disease Suganthi Jeyasingh1* and Malathy Veluchamy2

7) Giaquinto AN, Sung H, Miller K, et al. Breast cancer statistics, 2022. CA Cancer J Clin. 2022;72(6):524-541. [DOI] [PubMed] [Google Scholar]

8) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries by Freddie Bray BSc, MSc, PhD, Mathieu Laversanne MSc, Hyuna Sung PhD, Jacques Ferlay ME, Rebecca L. Siegel MPH, Isabelle Soerjomataram MD, MSc, PhD, Ahmedin Jemal DVM, PhD

9) Detection with SVM, RF, kNN

- *"Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers"* (J Med Syst, 2019).

10) Hybrid CNN + Metadata + ML

*"Enhanced skin cancer diagnosis: deep feature extraction framework"* (Frontiers in Medicine, 2024).

Hybrid models combining CNN with SVM, RF, LR, and metadata from patient info achieved ≥5% performance boost

11) Meta-learning & transfer learning advances

*"Advancements in skin cancer classification: review of ML techniques"* (2024).

Highlights multi-stage feature fusion, pre-trained networks + SVM ensembles achieving >96% accuracy on ISIC