# SKIN CANCER – MELANOMA PREDICTION AND SKIN LESION CLASSIFICATION USING DEEP LEARNING MODELS

PHASE I REPORT

*Submitted by*


| | |
|---|---|
| **ADITHAN K** | **21011101009** |
| **AGILAN MA** | **21011101011** |
| **C GANESH RAM** | **21011101034** |
| **DHEV MUGUNDDHAN A** | **21011101040** |


*in partial fulfilment for the award of the degree of*

**BACHELOR OF TECHNOLOGY**
**IN**
**ARTIFICIAL INTELLIGENCE**
**&**
**DATA SCIENCE**



**DEPARTMENT OF COMPUTER SCIENCE**
**AND ENGINEERING**
**SCHOOL OF ENGINEERING**
**SHIV NADAR UNIVERSITY CHENNAI**


**NOVEMBER 2024**

# SHIV NADAR UNIVERSITY CHENNAI

## BONAFIDE CERTIFICATE

Certified that this report titled "**SKIN CANCER – MELANOMA PREDICTION AND SKIN LESION CLASSIFICATION USING DEEP LEARNING MODELS**" is the bonafide work of **ADITHAN K** (Reg. No: **21011101009**), **AGILAN MA** (Reg. No: **21011101011**), **C GANESH RAM** (Reg. No: **21011101034**), **DHEV MUGUNDDHAN A** (Reg. No: **21011101040**) who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.


**SIGNATURE**

Dr. T. Nagarajan

**Professor**

**Head of the Department**

Department of Computer Science & Engineering

School of Engineering

Shiv Nadar University Chennai,

Kalavakkam -603110.

**SIGNATURE**

Mrs. Dhivya S

**Assistant Professor**

**Supervisor**

Department of Computer Science & Engineering

School of Engineering

Shiv Nadar University Chennai,

Kalavakkam -603110.

# ABSTRACT

The purpose of our study is to perform skin lesion categorization on dermoscopic pictures. We worked on the ISIC segmentation and classification challenge datasets. we are trying to build an efficient mode with lesser parameters and run time complexity during inference to perform diagnosis efficiently. The technique of identifying skin lesions typically involves four stages: pre-processing, segmentation, feature extraction, and classification. Base model: ResNet50 and ResNet151 transfer learning.

Other Methods and models: Resizing and contrast enhancement in preprocessing, UNet segmentation for ROI, Orb feature extraction and xgBoost model. DCNN feature extraction then feature fusion with ORB features using Hadamard Product and Attention based feature fusion techniques. We find DCNN feature extraction and feature fused with Hadamard Product proves to be a less computationally expensive as well as provide better results than the other alternatives experimented.

**Keywords** - Skin Lesion Classification, Melanoma Prediction, CNN, ResNet, Attention net, Feature Fusion, ISIC Challenge, ORB features

# ACKNOWLEDGEMENT

we would like to express our sincere gratitude to all those who have provided invaluable guidance, support, and encouragement throughout the course of this project.

First and foremost, We are deeply grateful to **Dr. T. Nagarajan**, Professor and Head, Department of Computer Science and Engineering, School of Engineering, Shiv Nadar University Chennai, for his steadfast guidance and unwavering encouragement. His insights into the field of artificial intelligence and cognitive science have been instrumental in shaping the direction of this research.

We would like to extend our heartfelt thanks to our supervisor, **Mrs. Dhivya S**, Assistant Professor, Department of Computer Science and Engineering, School of Engineering, Shiv Nadar University Chennai, for her mentorship and commitment to the success of this project. Her expertise in the domain of Artificial Intelligence and Medical Imaging has been invaluable, providing us with the clarity and focus needed to tackle complex challenges and refine the methodology of this work.

We also acknowledge with gratitude the support of our family, whose understanding and encouragement have been crucial to the implementation of this project. Their unwavering belief in us has been a constant source of motivation and strength.

**ADITHAN K**
**(21011101009)**

**AGILAN MA**
**(21011101011)**

**C GANESH RAM**
**(21011101034)**

**DHEV MUGUNDDHAN A**
**(21011101040)**

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF SYMBOLS, ABBREVIATIONS

| | |
|---|---|
| $\hat{y}_{i,c}$ | The predicted probability for the $i$-th sample and $c$-th class |
| $C$ | The total number of classes |
| $L$ | The loss function (e.g., cross-entropy loss) |
| $N$ | The total number of data samples |
| $y_{i,c}$ | The ground truth (true label) for the $i$-th sample and $c$-th class |

# CHAPTER 1: INTRODUCTION

## 1.1 BACKGROUND

### 1.1.1 Overview

Melanoma, a highly aggressive form of skin cancer, accounts for around 4% of all skin cancer cases but is responsible for 75% of skin cancer-related deaths. Its danger lies in its tendency to spread rapidly to other organs if not detected and treated early. In recent years, the incidence of melanoma has increased due to increased exposure to ultraviolet (UV) radiation, often driven by lifestyle changes such as more frequent sun-seeking behaviors and inadequate protection from harmful sun rays. These factors have contributed to an alarming increase in melanoma cases worldwide. Early detection is essential for survival; studies show a five-year survival rate of approximately 94% when melanoma is identified early.[1]

To address the need for a timely diagnosis, machine learning-based computer-aided design (CAD) systems are being investigated as reliable tools to improve the accuracy of melanoma detection. Machine learning models, especially deep learning (DL) techniques, have shown potential for consistent, efficient, and automated diagnosis. Therefore, this project explores the application of deep learning to detect melanoma by analyzing dermoscopic images, which are high-resolution skin images that provide detailed views of lesions. By automating lesion classification, this approach aims to make melanoma detection faster and more accurate, potentially saving lives and optimizing medical resources.

### 1.1.2 Problem Formulation

The core problem addressed by this project is the need for an accurate, efficient, and accessible method of detecting melanoma early on. Traditional diagnostic techniques, such as visual inspection by dermatologists and biopsy procedures, require specialized expertise, are

subjective, and can be slow. These limitations are compounded by a global shortage of dermatologists, particularly in remote or underserved areas. Consequently, many patients may experience delays in diagnosis or misdiagnosis, reducing the chance of effective treatment.

To solve this problem, the project proposes a deep learning-based CAD system that can analyze dermoscopic images and classify skin lesions to detect potential melanoma cases. This approach leverages recent advancements in deep learning, particularly convolutional neural networks (CNNs), which have shown exceptional capabilities in image classification tasks. CNNs, along with techniques such as transfer learning, data augmentation, and ensemble learning, are used to develop a robust, scalable, and high-performance system that can differentiate melanoma from benign skin lesions with high accuracy. This automated system could greatly reduce diagnostic wait times, provide more reliable initial evaluations, and offer an accessible diagnostic tool for those in regions lacking specialist resources.

### 1.1.3 Objectives

The main objectives of this project are outlined below, focusing on both the technical development and the broader implications for melanoma detection:

1. Develop an Automated Skin Lesion Classification Model: The primary goal is to build a model that can classify dermoscopic images of skin lesions to detect melanoma. Deep learning, particularly convolutional neural networks (CNNs), forms the backbone of this system, given its superior image processing abilities for complex patterns and textures typical of skin lesions.[1]

2. Implement Data Augmentation Techniques to Address Class Imbalance: In medical datasets, positive cases (e.g., melanoma) are often much less common than negative cases. This class imbalance can lead to biased models. Techniques such as synthetic image generation using Generative Adversarial Networks (GANs) and dynamic augmentation strategies will be applied to ensure the model learns from a balanced dataset, which is crucial for improving accuracy and robustness.

3. Explore Transfer Learning and Ensemble Models to Enhance Performance: Transfer learning will leverage pre-trained models (e.g., ResNet50, InceptionV3, and EfficientNet) that are already highly effective in image recognition tasks. Using these models as a starting point enhances performance on limited medical datasets. Furthermore, ensemble learning, which combines the predictions of multiple models, will provide better accuracy and robustness by leveraging the strengths of different models.

4. Integrate Novel Architectures for Improved Feature Extraction and Segmentation: Advanced techniques, such as Mask R-CNN for precise lesion segmentation, will be used to automatically identify and focus on melanoma-prone regions within dermoscopic images. Additionally, Speeded-Up Robust Features (SURF) will help extract essential keypoints from images, capturing the most relevant details for improved classification.

5. Evaluate Model Performance with Robust Metrics: The effectiveness of the system will be measured using the accuracy, precision, recall and AUC-ROC scores (area under the receiver operating characteristic curve). This multi-metric evaluation approach is essential to ensure that the system is clinically viable and accurately distinguishes between melanoma and benign lesions.

### 1.1.4 Motivation

The motivation for developing an automated melanoma detection system is driven by the need to improve diagnostic accuracy and accessibility, addressing gaps in current diagnostic methods. Conventional approaches such as visual inspection and biopsy have limitations. Visual inspection by dermatologists can be subjective, leading to variability in diagnosis based on individual experience. Biopsies, while more reliable, are invasive, time-consuming, and costly. These methods can create barriers to timely and accurate diagnosis, particularly in areas where dermatologists are not readily available.

Given melanoma's high fatality rate when left undetected, developing an automated solution can have a significant impact on public health. An AI-driven CAD system could function

as a quick, reliable tool for initial screenings, helping detect potential melanoma cases early on, even in primary care settings or through self-screening applications. Such a system could reduce reliance on specialized expertise, making accurate diagnostics more widely available, improving survival rates, and potentially lowering healthcare costs associated with advanced melanoma treatments.

In summary, this project aims to provide a sophisticated and accessible solution that bridges the gap between patient needs and diagnostic resources. By creating an AI-powered tool for early melanoma detection, this project contributes to both technological innovation in healthcare and improved patient outcomes.

# CHAPTER 2: LITERATURE SURVEY

The surveyed literature demonstrates that deep learning models, particularly CNNs and their ensembles, have shown significant promise in the classification of skin lesions and the detection of melanoma. Techniques such as transfer learning, data augmentation, and hybrid approaches further enhance model performance, especially in scenarios with limited data. The combination of these methods with innovative training techniques, such as dynamic augmentation, is pushing the boundaries of what is possible in automated melanoma detection. Future research should focus on improving data availability, exploring novel architectures, and refining augmentation techniques to make these models more robust and widely applicable in clinical settings.[3]

Table 2.1: Sample Table

| Model Name | Parameter count | Accuracy |
|---|---|---|
| Basic CNN | 1M-5M | 70%-80% |
| ResNet | 25M (ResNet-50), 45M (ResNet-101) | 85%-92% |
| Efficient Net | 5M (B0) to 65M (B7) | 88%-94% |
| SENet | Similar to ResNet ( 25M-50M) | 90% |

## 2.1 BASIC CNN

Convolutional Neural Networks (CNNs) were among the earliest deep learning models applied to skin cancer detection[6]. These models demonstrated the potential of automated image analysis for classifying skin lesions. Basic CNN architectures typically consisted of a few convolutional layers followed by pooling and fully connected layers. While they showed promise, their accuracy was limited due to several factors:

- Early CNN models achieved classification accuracies in the range of 70-80% for distinguishing between benign and malignant lesions, depending on the dataset used

- Lack of depth in architecture limited their ability to capture complex features in dermoscopic images.
- Overfitting was common due to small datasets, as skin lesion datasets were not as extensive or diverse at the time.
- Inability to generalize well across different imaging conditions or patient demographics.

Despite these limitations, basic CNNs laid the groundwork for more advanced architectures that addressed these challenges.

## 2.2 ADVANCED ARCHITECTURES

### 2.2.1 ResNet (Residual Neural Network)

ResNet introduced residual connections, which allowed networks to be much deeper without suffering from vanishing gradient problems. This innovation significantly improved performance in skin cancer detection tasks.[4]

- ResNet-based models achieved classification accuracies ranging from 85% to 92% on benchmark datasets like ISIC (International Skin Imaging Collaboration).
- For melanoma detection specifically, ResNet variants like ResNet-50 and ResNet-101 demonstrated sensitivity rates exceeding 90% when trained on large-scale datasets.
- High computational cost: Training deep ResNet models required significant computational resources, making them less accessible for smaller research groups or clinical settings with limited infrastructure.
- Sensitivity to data imbalance: Like other deep learning models, ResNet struggled with imbalanced datasets where malignant cases were underrepresented compared to benign ones

### 2.2.2 EfficientNet

EfficientNet introduced a novel scaling method that balanced network depth, width, and resolution systematically. This approach enabled EfficientNet models to achieve state-of-the-art performance while being computationally efficient.[12]

- EfficientNet-B0 through B7 variants achieved classification accuracies between 88% and 94%, depending on the specific variant and dataset used.
- On ISIC datasets, EfficientNet-B4 achieved an AUC (Area Under Curve) score of approximately 94%, outperforming many traditional architectures.
- Computational efficiency comes at the cost of increased complexity in model design and hyperparameter optimization.

### 2.2.3 SENet (Squeeze-and-Excitation Network)

SENet introduced a mechanism called "squeeze-and-excitation," which adaptively re-calibrates channel-wise feature responses by modeling interdependencies between channels. This enhancement made SENet particularly effective at focusing on critical lesion features in dermoscopic images.[13]

- SENet-based models achieved accuracies comparable to or slightly better than ResNet and EfficientNet in some studies, often exceeding 90%.
- For melanoma detection tasks specifically, SENet demonstrated improved sensitivity rates due to its ability to emphasize important features like irregular borders or color variations in lesions.
- Increased model complexity: The squeeze-and-excitation mechanism adds additional computation overhead during both training and inference phases.

# CHAPTER 3: METHODOLOGY

This section details the various steps undertaken in our research, including data preprocessing, model selection, training protocols, and evaluation metrics. Each phase is meticulously documented to ensure the reproducibility and transparency of our findings. By outlining our systematic approach, we aim to contribute to the ongoing advancements in AI-driven medical imaging and provide a robust framework for future studies in melanoma prediction.

## 3.1 Dataset

We have used the ISIC 2018 challenge datasets for Task 1 (Segmentation) and Task 3 (Classification).

**Task 1 Dataset:**

Train data: The dataset includes 2594 images and 12970 corresponding ground truth response masks (5 for each image).

Test Data: The dataset includes 1000 images and 1000 corresponding ground truth response masks.

**Task 3 Dataset:**

Train data: Train dataset contains 10015 images and 1 ground truth response CSV file (containing 1 header row and 10015 corresponding response rows). The labels are seven categories: melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis/Bowen's disease, benign keratosis, dermatofibroma, and vascular lesion.

Test Data: Test dataset contains 1512 images. and 1 ground truth response CSV file (containing 1 header row and 1512 corresponding response rows).

The classification pipeline in this study incorporates several stages: pre-processing, segmentation, feature extraction, and classification, each tailored to address the unique challenges posed by dermoscopic images.[2]

## 3.2 Overall Process in Skin Lesion Classification

The process of skin lesion classification typically involves four main stages:

1. **Pre-processing** prepares the images for further analysis.

2. **Segmentation** isolates the lesion from surrounding background.

3. **Feature extraction** captures critical information from the lesion.

4. **Classification** assigns the lesion to one of several diagnostic categories.

This structured pipeline not only enhances the model's focus on lesion-specific features but also ensures a consistent input format and quality, which are essential for reliable classification performance.

## 3.3 Base Model: ResNet with Transfer Learning

For the core of our model, we selected ResNet-50 and ResNet-151 as base architectures. These models, known as residual networks, utilize skip connections that enable the training of very deep layers by mitigating issues like the vanishing gradient problem. This residual structure allows ResNet models to effectively learn complex hierarchical representations of images, making them well-suited for medical image analysis where subtle differences between lesion types are critical for accurate classification.[5]

We utilize transfer learning by initializing these ResNet models with pre-trained weights from large-scale image datasets, enhancing the model's generalization on medical data despite limited dermoscopic images. Transfer learning allows us to capitalize on previously learned feature representations, significantly speeding up training and often improving per-

formance, which is particularly valuable in tasks like skin lesion classification where available labeled data can be limited.[8]

## 3.4 Preprocessing Techniques

Preprocessing plays a fundamental role in standardizing and enhancing dermoscopic images, enabling more effective learning by the model. The following preprocessing steps are employed:

Resizing: We resize each image to a fixed resolution of 256x256 pixels, which maintains consistency across the dataset and fits the input requirements of the ResNet models. Bilinear interpolation is used for resizing, as it helps preserve essential image details by averaging nearby pixel values. Consistent image sizing is essential for CNN-based models, as it ensures that the model's filters and learned parameters apply uniformly across all images.

Contrast Enhancement: Dermoscopic images often suffer from variable lighting and contrast, which can obscure critical lesion features. By enhancing contrast, we ensure that essential patterns and borders in the lesions are more visible. This step enhances the model's ability to detect differences in color and texture—key indicators for distinguishing between different lesion types. For example, contrast enhancement makes irregular lesion boundaries or color changes within a lesion more distinguishable, which are often diagnostic markers for melanoma.

## 3.5 UNet Segmentation for Region of Interest (ROI)

Dermoscopic images typically contain irrelevant background features that may interfere with accurate lesion classification. To focus on the lesion itself, we employ a UNet segmentation model to generate masks that isolate the region of interest (ROI). UNet, a widely used architecture in biomedical image segmentation, is designed with an encoder-decoder structure that captures both high-level context and fine details, making it highly effective for

precise segmentation tasks.

By applying UNet-based segmentation, we generate masks that clearly outline the lesion, allowing us to eliminate non-essential background elements from the images. This step directs the model's attention solely to the lesion region, ensuring that features relevant to classification are highlighted. Segmenting out irrelevant background details not only improves classification accuracy but also reduces noise, as the model can concentrate on identifying subtle patterns specific to each lesion class.

## 3.6 Feature Extraction Techniques

Beyond standard CNN feature extraction, we incorporate additional feature extraction techniques to enhance the feature space available for classification. This combined approach ensures that both global and local patterns within lesions are captured.

ORB (Oriented FAST and Rotated BRIEF) Feature Extraction: ORB is a computationally efficient feature extraction method known for its robustness to rotational changes and lighting variations, both of which are common in dermoscopic images. ORB identifies and describes local keypoints, which can capture intricate texture and structural patterns within lesions. These ORB features provide complementary information to CNN-based features, allowing for a richer and more detailed feature representation.

DCNN-Based Feature Extraction: Deep convolutional neural networks (DCNNs) extract complex patterns and high-level image features by learning from large amounts of data. DCNNs, such as ResNet, are particularly adept at capturing hierarchical features, which are essential for identifying subtle patterns across different lesion types. These features provide the model with high-level insights into lesion structure, shape, and texture.

PCA and XGBoost for ORB Feature Classification: To reduce the dimensionality of ORB features, we apply Principal Component Analysis (PCA), which minimizes redundant information and focuses on the most informative components. This step is crucial for han-

dling high-dimensional data, making it computationally manageable while preserving critical lesion details. The dimensionality-reduced ORB features are then fed into an XGBoost classifier, a gradient boosting method known for its effectiveness in handling structured data. XGBoost serves as a secondary classifier to further validate and improve the results from the primary CNN model.

## 3.7 Feature Fusion with Hadamard Product and Attention Mechanisms

Feature fusion is applied to combine the distinct information captured by CNN-based and ORB features, enhancing the model's representation of each lesion. We achieve feature fusion using the Hadamard product[9], an element-wise multiplication technique that emphasizes overlapping and complementary information between feature sets. By combining CNN-derived global features and ORB-derived local features, the Hadamard product creates a unified feature representation that captures both broad patterns and fine-grained details.[7]

**Attention-Based Feature Fusion**: To further refine feature fusion, we integrate attention mechanisms that allow the model to focus on the most informative parts of the feature space. Attention layers dynamically adjust the importance of different features, highlighting those most relevant to the classification task. This approach helps the model prioritize lesion-specific attributes, such as irregular shapes, unique textures, or distinct color patterns, which may be crucial for distinguishing between similar lesion classes. By employing attention-based fusion, our model gains additional sensitivity to subtle, yet significant, visual cues in the lesion images.[10]

**DCNN Features Extraction**: The SkinCancerCNN (assumed to be a deep convolutional neural network) extracts 1024 features from the input image. The extracted dcnn_features have a shape of (batch_size, 1024).

**Attention on DCNN Features**: The attention mechanism for DCNN features involves

a sequence of fully connected layers: Linear layer reducing from 1024 to 512 dimensions. ReLU activation introduces non-linearity. Linear layer increasing dimensions back to 1024. Sigmoid activation normalizes the attention weights. The resulting dcnn_attention_weights (shape (batch_size, 1024)) are element-wise multiplied with the original dcnn_features, producing dcnn_features_attended. This operation acts as a soft selection mechanism, enhancing or suppressing different elements of the feature vector.

**Attention on ORB Features**: ORB features are assumed to have 300 dimensions (e.g., obtained using ORB feature extraction). The ORB attention mechanism similarly uses a fully connected layer sequence: Linear layer reducing from 300 to 150 dimensions. ReLU activation. Linear layer increasing dimensions back to 300. Sigmoid activation normalizes the attention weights. The resulting orb_attention_weights (shape (batch_size, 300)) are element-wise multiplied with orb_features, producing orb_features_attended.

**Feature Fusion**: The attended DCNN and ORB features are concatenated along the feature dimension, resulting in a fused feature vector of shape (batch_size, 1324) (1024 + 300). The fusion_layer then reduces this vector to 512 dimensions using a fully connected layer.

**Loss Function**: In melanoma skin classification tasks, such as those using the ISIC dataset, cross-entropy loss is a widely used objective function for training deep learning models. This is primarily because melanoma classification is a binary or multi-class classification problem, where the goal is to predict the probability distribution over classes (e.g., melanoma vs. non-melanoma). Cross-entropy loss measures the dissimilarity between the predicted probability distribution output by the model and the true class labels. It penalizes incorrect predictions more heavily when they are confident but wrong, making it an effective choice for imbalanced datasets like ISIC, where only 2% of samples may belong to the positive class.

Categorical Cross-Entropy Loss For multi-class classification problems (e.g., distinguishing between multiple types of skin lesions), categorical cross-entropy loss generalizes binary

cross-entropy and is defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c})$$

These probabilities are typically obtained using a softmax activation function applied to the model's raw output logits. Categorical cross-entropy ensures that higher penalties are assigned when predicted probabilities diverge significantly from one-hot encoded true labels.[11]

# CHAPTER 4: EXPERIMENTS

In this study, a range of deep learning and machine learning models were employed to tackle the challenges of automated skin lesion classification. The goal was to develop accurate and efficient systems capable of identifying melanoma from dermoscopic images. The models implemented include ResNet-50 and ResNet-152, which leverage transfer learning to utilize pre-trained weights for robust feature extraction from medical images. Additionally, UNet-based segmentation was integrated to isolate regions of interest, reducing noise from irrelevant background features.

To enhance the richness of the feature space, local features were extracted using ORB (Oriented FAST and Rotated BRIEF), which were subsequently classified with the XGBoost algorithm. Advanced models incorporated deep convolutional neural networks (DCNNs) combined with feature fusion techniques, including the Hadamard product and attention mechanisms. These fusion strategies allowed for a seamless integration of high-level CNN-derived features and local ORB features, enabling the models to capture both global and fine-grained lesion characteristics.

Each model was carefully tuned and evaluated for its performance in terms of accuracy and computational efficiency, offering insights into the potential and limitations of deep learning in automated melanoma detection.

## 4.1 RESNET MODELS

The ResNet-50 and ResNet-152 models were employed as base architectures for skin lesion classification using transfer learning. These models leveraged pre-trained weights to capitalize on feature representations from large-scale image datasets, enabling effective learning even with limited medical data. Images were resized to 224x224 pixels and un-

derwent contrast enhancement as preprocessing steps. ResNet-50, with 23M trainable parameters, achieved 90% training accuracy but only 50% on test data, indicating overfitting. Similarly, ResNet-152, with 60M trainable parameters, achieved 90% training accuracy and 60% test accuracy, showing a marginal improvement.

## 4.2 UNET WITH XGBOOST

The UNet model was implemented for segmentation to isolate regions of interest, enhancing focus on lesion areas while discarding irrelevant background features. Segmented regions were processed using ORB (Oriented FAST and Rotated BRIEF) for local feature extraction. These features were then classified using XGBoost, a robust gradient-boosting algorithm. Despite the improved focus achieved by segmentation, this pipeline achieved a test accuracy of 60.25%, indicating the need for more robust feature integration.

## 4.3 DCNN WITH HADAMARD PRODUCT FUSION

A DCNN-based model incorporated feature fusion using the Hadamard product to combine high-level features from CNNs with local ORB-derived features. This method emphasized overlapping and complementary information, resulting in a richer feature representation. The model, trained on 256x256 preprocessed images with 29M trainable parameters, achieved a training accuracy of 95.66%. However, test accuracy remained at 57.01%, suggesting challenges in generalization.

## 4.4 DCNN WITH ATTENTION-BASED FUSION

To refine feature fusion further, an attention mechanism was introduced, enabling the model to dynamically prioritize lesion-specific features. This approach combined attended CNN and ORB features, enhancing the focus on critical patterns. The attention-based fusion model, with 31M trainable parameters, processed 256x256 input images and achieved

66.78% training accuracy and 60% test accuracy. While offering competitive performance, it highlighted the complexity of accurately modeling skin lesion characteristics.

This structured approach to model development demonstrates the exploration of different architectures and techniques to optimize melanoma detection. However, achieving consistent generalization remains a challenge, necessitating further exploration of advanced techniques or dataset augmentation.

| **Model** | **Trainable Parameters** | **Preprocessing Techniques** | **Input Image Size** | **Segmentation** | **Accuracy (Train/Test)** |
|---|---|---|---|---|---|
| Resnet 50 | 23M | Resize, Contrast | 224x224 | No | 90%/50% |
| Resnet 152 | 60M | Resize, Contrast | 224x224 | No | 90%/60% |
| XGBoost | - | Resize, Contrast | 256x256 | Yes | NIL/60.25% |
| DCNN with Feature Fusion (Hadamard product) | 29M | Resize, Contrast | 256x256 | Yes | 95.66%/57.01% |
| DCNN with Feature Fusion (Attention Net) | 31M | Resize, Contrast | 256x256 | Yes | 66.78%/60% |

Table 4.1: Models Info

# CHAPTER 5: CONCLUSION

This study presents a deep learning-based approach for automated skin lesion classification, aiming to enhance melanoma detection accuracy and efficiency. By using the ISIC datasets, we developed a structured pipeline that includes pre-processing, UNet segmentation, feature extraction, and classification to address the complexities of dermoscopic image analysis.

Our model uses ResNet architectures with transfer learning as a foundation, supported by a series of preprocessing steps and UNet segmentation to isolate the region of interest effectively. A significant aspect of our approach is the integration of CNN-based and ORB-derived features. By combining high-level patterns captured by DCNNs with local details extracted via ORB, we created a comprehensive feature set. Dimensionality reduction and XGBoost further refined ORB features, contributing to a robust classification foundation.

Through feature fusion using the Hadamard product, we merged CNN and ORB features, enhancing the model's ability to distinguish between lesion types. Attention-based fusion added further refinement by focusing on essential lesion characteristics. Our findings reveal that DCNN feature extraction with Hadamard Product fusion provides a more computationally efficient and effective solution compared to other methods, making it a suitable option for real-time clinical application.

In conclusion, this study demonstrates the potential of combining deep learning and traditional machine learning techniques for skin lesion classification, providing a foundation for future improvements in automated melanoma detection. Further exploration into advanced fusion techniques and attention mechanisms could continue to enhance this model's accuracy and applicability in clinical settings.

# REFERENCES


[1] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018; https://arxiv.org/abs/1902.03368

[2] Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi:10.1038/sdata.2018.161 (2018).

[3] Azeem, M.; Kiani, K.; Mansouri, T.; Topping, N. SkinLesNet: Classification of Skin Lesions and Detection of Melanoma Cancer Using a Novel Multi-Layer Deep Convolutional Neural Network. Cancers 2024, 16, 108. https://doi.org/10.3390/cancers16010108

[4] M. A. Al-masni, M. A. Al-antari, H. M. Park, N. H. Park and T. -S. Kim, "A Deep Learning Model Integrating FrCN and Residual Convolutional Networks for Skin Lesion Segmentation and Classification," 2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), Okinawa, Japan, 2019, pp. 95-98, doi: 10.1109/ECBIOS.2019.8807441.

[5] C. K, P. C. Siddalingaswamy, S. Pathan and N. D'souza, "A Multiclass Skin Lesion classification approach using Transfer learning based convolutional Neural Network," 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 2021, pp. 1-6, doi: 10.1109/ICBSII51839.2021.9445175.

[6] A. Kumari and N. Sharma, "A Review on Convolutional Neural Networks for Skin Lesion Classification," 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), Jalandhar, India, 2021, pp. 186-191, doi: 10.1109/IC-

SCCC51823.2021.9478151.

[7] D. Sudha and M. Ramakrishna, ”Comparative Study of Features Fusion Techniques,” 2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT), Bangalore, India, 2017, pp. 235-239, doi: 10.1109/ICRAECT.2017.39.

[8] He, Kaiming, et al. ”Deep residual learning for image recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[9] Zhang, Xiaoyan, Mengmeng He, and Hongan Li. ”DAU-Net: A medical image segmentation network combining the Hadamard product and dual scale attention gate.” Mathematical Biosciences and Engineering 21.2 (2024): 2753-2767.

[10] Y. Zhang, Z. Chen, H. Yu, X. Yao and H. Li, ”Feature Fusion for Segmentation and Classification of Skin Lesions,” 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), Kolkata, India, 2022, pp. 1-5, doi: 10.1109/ISBI52829.2022.9761474.

[11] A. Demirkaya, J. Chen and S. Oymak, ”Exploring the Role of Loss Functions in Multiclass Classification,” 2020 54th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 2020, pp. 1-5, doi: 10.1109/CISS48834.2020.1570627167.

[12] S. Papiththira and T. Kokul, ”Melanoma Skin Cancer Detection Using EfficientNet and Channel Attention Module,” 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS), Kandy, Sri Lanka, 2021, pp. 227-232, doi: 10.1109/ICIIS53135.2021.9660759.

[13] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. ”Squeeze-and-Excitation Networks” arXiv preprint, arXiv:1709.01507, 2019. URL: https://arxiv.org/abs/1709.01507.

# APPENDIX A: Appendix

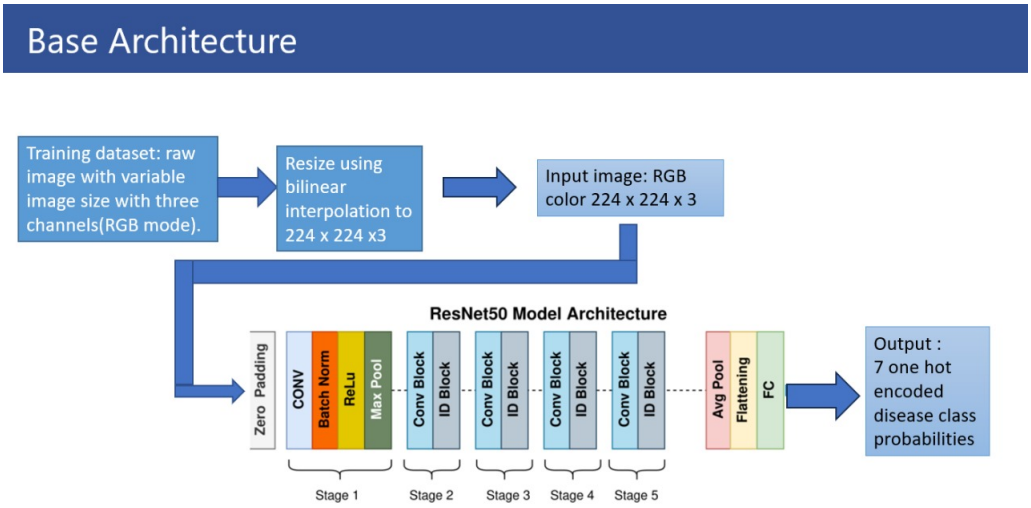


Figure A.1: Base Architecture

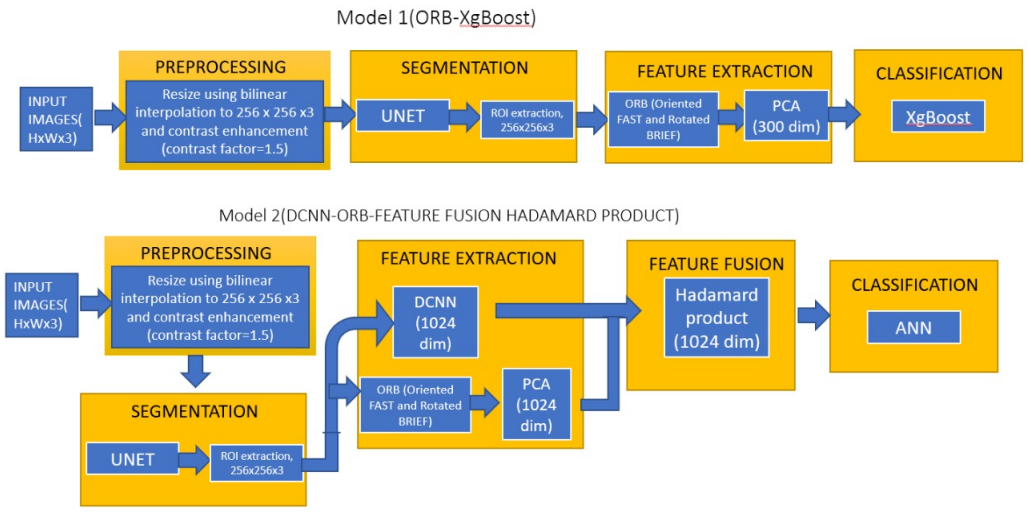


Figure A.2: Model 1(ORB-XgBoost) and Model 2(DCNN-ORB-Feature Hadamard Product)

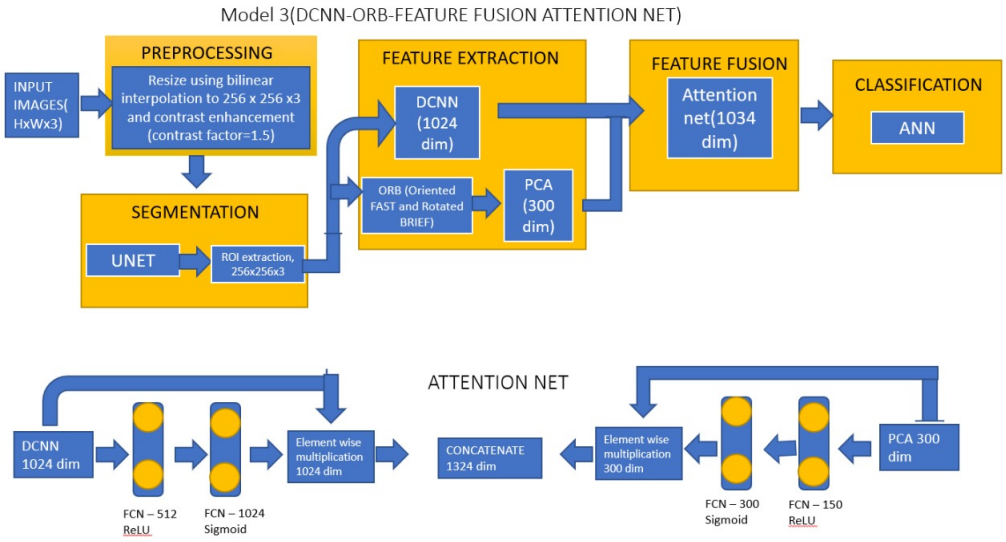


Figure A.3: Model 3(DCNN-ORB-Feature Fusion Attention Net)

UNET SEGMENTATION ARCHITECTURE

| Stage | Operation | Output Size | Features |
|---|---|---|---|
| Input | Preprocessing | 256×256×3 | RGB Image |
| Encoder Block 1 | Conv + ReLU + Conv + ReLU | 256×256×64 | Features |
| Max Pooling | Downsample | 128×128×64 | Reduced Size |
| Encoder Block 2 | Conv + ReLU + Conv + ReLU | 128×128×128 | Features |
| Max Pooling | Downsample | 64×64×128 | Reduced Size |
| Encoder Block 3 | Conv + ReLU + Conv + ReLU | 64×64×256 | Features |
| Max Pooling | Downsample | 32×32×256 | Reduced Size |
| Encoder Block 4 | Conv + ReLU + Conv + ReLU | 32×32×512 | Features |
| Max Pooling | Downsample | 16×16×512 | Reduced Size |
| Bridge | Conv + ReLU + Conv + ReLU | 16×16×1024 | Bottleneck |
| Upconv Block (4) | Upsample | 32×32×512 | Upsampled |
| Concatenate | Skip Connection | 32×32×(512+512) | Combined Features |
| Decoder Block (4) | Conv + ReLU + Conv + ReLU | 32×32×512 | Refined Features |
| Upconv Block (3) | Upsample | 64×64×256 | Upsampled |
| Concatenate | Skip Connection | 64×64×(256+256) | Combined Features |
| Decoder Block (3) | Conv + ReLU + Conv + ReLU | 64×64×256 | Refined Features |
| Upconv Block (2) | Upsample | 128×128×128 | Upsampled |
| Concatenate | Skip Connection | 128×128×(128+128) | Combined Features |
| Decoder Block (2) | Conv + ReLU + Conv + ReLU | 128×128×128 | Refined Features |
| Upconv Block (1) | Upsample | 256×256×64 | Upsampled |
| Concatenate | Skip Connection | 256×256×(64+64) | Combined Features |
| Decoder Block (1) | Conv + ReLU + Conv + ReLU | 256×256×64 | Refined Features |
| Final Convolution Layer | Convolution | 256×256×1 | Segmentation Output |

Figure A.4: UNET Segmentation Architecture