LIGHTWEIGHT AI FOR REAL TIME SKIN LESION ANALYSIS

PROJECT-21ECP302L

Submitted by

Raghul Krishna R [Reg No:RA2211052010018]

Sai Krishnan S [Reg No: RA2211052010023]

Nagarjuna C Raju [Reg No: RA2211067010017]

Under the guidance of

Dr A. Ruhan Bevi

(Associate Professor, Department of Electronics & Communication Engineering)

BACHELOR OF TECHNOLOGY

in

ELECTRONICS & COMMUNICATION ENGINEERING

of

COLLEGE OF ENGINEERING AND TECHNOLOGY

SRM

INSTITUTE OF SCIENCE & TECHNOLOGY

(Deemed to be University u/s 3 of UGC Act, 1956)

S.R.M. NAGAR, Kattankulathur, Chengalpattu District

MAY 2025

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act,1956)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report titled "**LIGHTWEIGHT AI FOR REAL TIME SKIN LESION ANALYSIS**" is the bonafide work of **Raghul Krishna R [RA2211052010018], Sai Krishnan S [RA2211052010023] and Nagarjuna C Raju [RA2211067010017]** who carried out the project work under my supervision as part of the course **21ECP302L - PROJECT.**

Certified further that, to the best of my knowledge, the work reported herein is original and was carried out during the academic year 2024–2025 (Even) at **SRM Institute of Science and Technology, Kattankulathur.**

SIGNATURE	**SIGNATURE**
DR. RUHAN BEVI A.	Dr S. KRITHIGA
SUPERVISER	**ACADEMIC ADVISOR**
Associate Professor	Associate Professor
Dept. of Electronics & Communication Engineering	Dept. of Electronics & Communication Engineering

ABSTRACT

Accurate detection of skin cancer is essential for improving survival rates and ensuring timely medical intervention. However, current diagnostic methods like visual inspection, dermoscopy, and biopsies are of high cost and need for specialized medical expertise, and variability in accuracy due to subjective human interpretation. These challenges are particularly pronounced in low-resource settings, where access to expert dermatologists and advanced diagnostic infrastructure are few.

To address these issues, this study proposes a lightweight, embedded solution for skin lesion segmentation and analysis, optimized to run efficiently on a Raspberry Pi. By leveraging computationally efficient variants of established deep learning architectures such as DenseNet, MobileNet, and EfficientNet the proposed models significantly reduce model size and inference latency, making it suitable for deployment on low-power edge devices. Even though there are reductions in computational complexity, the model maintains a high level of segmentation accuracy, capable of identifying lesion boundaries even in cases involving irregular shapes or blurred contours.

A focus of the model design is its ability to access across diverse skin tone and lesion types. Traditional models often exhibit performance disparities due to dataset imbalances or biased feature learning, which can lead to misdiagnosis in underrepresented populations. To combat this, the proposed system integrates bias mitigation techniques during training, such as adaptive data augmentation and balanced sampling, to improve robustness and fairness. This ensures consistent diagnostic performance across various demographic groups and dermatological conditions.

Ultimately, this work aims to democratize access to early skin cancer detection through an accessible, low-cost, and portable system. By combining lightweight deep learning models with embedded hardware, it offers a promising approach to scaling dermatological diagnostics in both clinical and remote environments.

TABLE OF CONTENTS

Chapter No.	Title	Page No.
ABSTRACT		**iii**
TABLE OF CONTENTS		**iv**
LIST OF TABLES		**vii**
LIST OF ABBREVIATIONS		**viii**
CHAPTER 1 INTRODUCTION		**1**
1.1.	Introduction	1
1.2.	Objective	2
1.3.	Edge Devices (Embedded Boards)	2
CHAPTER 2 LITERATURE SURVEY		**4**
2.1.	Skin cancer detection with MobileNet-based transfer learning and MixNets for enhanced diagnosis.	4
2.2.	MobileNetV2: An Enhanced Skin Disease Classification by Attention and Multi- Scale Features	5
2.3.	A Dermoscopic Skin Lesion Classification Technique Using YOLO-CNN and Traditional Feature Model	6
2.4.	Precision Segmentation and Binary Masking of Skin Lesions in Automated Dermatological Diagnostics Using Detectron2	7
CHAPTER 3 SOFTWARE DESCRIPTION		**9**
3.1.	Jupyter library	9
3.2.	Programming Environment	9

3.3. TensorFlow 10

3.4. MobileNet 10

3.5. EfficientNet 11

3.6. DenseNet 11

CHAPTER 4 HARDWARE DESCRIPTION 13

4.1. Raspberry Pi 5 13

CHAPTER 5 METHODOLOGY OF SKIN LESION ANALYSIS 16

5.1. Model Flow 16

5.2. Engineering Standards 20

5.3. Multi-disciplinary Aspect 21

CHAPTER 6 MODEL TRAINING 27

CHAPTER 7 RESULTS AND OTHER INFERENCES 37

CHAPTER 8 CONCLUSION AND FUTURE WORK 37

7.1. Conclusion 37

7.2 Future work 38

7.3. Realistic Constraints 38

CHAPTER 9 REFRENCES 40

LIST OF FIGURES

Figure No.	Title	Page No.
4.1	Raspberry Pi board diagram	15
5.1	System Model of Skin Lesion Analysis	17
6.1.1	Training and Validation Loss of DenseNet	28
6.1.2	Training and Categorical Accuracy of DenseNet	29
6.1.3	Confusion Matrix of DenseNet	30
6.1.4	Accuracy Table According to Classification of DenseNetError! Bookmark not defined.	30
6.2.1	Training and Validation Loss of MobileNet	31
6.2.2	Training and Categorical Accuracy of MobileNet	32
6.2.3	Confusion Matrix of MobileNet	33
6.2.4	Accuracy Table According to Classification of MobileNet	33
6.3.1	Training and Validation Loss of EfficientNet	34
6.3.2	Training and Categorical Accuracy of EfficientNet	35
6.3.3	Confusion Matrix of EfficientNet	36
6.3.4	Accuracy Table According to Classification of EfficientNet	36

LIST OF TABLES

Table No.	Title	Page No.
4.1	Specifications of Raspberry Pi 5	13

LIST OF ABBREVIATIONS

ISIC	International Skin Imaging Collaboration
AI	Artificial Intelligence
ML	Machine Learning
HAM	Human Against Machine
CPU	Central Processing Units
CNN	Convolutional Neural Networks
DL	Deep Learning
GPU	Graphic Processing unit
TPU	Tensor Processing Unit
nv	Melanocytic nevi
mel	Melanoma
bkl	Benign keratosis-like lesions
bcc	Basal cell carcinoma
akiec	Actinic keratoses

vasc	Vascular lesions
df	Dermatofibroma

CHAPTER 1
INTRODUCTION

1.1. Introduction

Skin cancer is a major health concern, with lakhs of new cases diagnosed each year. Among the various types, melanoma is the deadly form of cancer, which is responsible for the majority of skin cancer-related deaths. Despite advancements in research and treatment, early detection remains the most crucial factor in improving survival rates and ensuring effective clinical outcomes. The timely identification of malignant lesions allows for less invasive treatments and significantly enhances the chances of recovery.

The access to reliable diagnostic tools and specialized dermatological care still remains limited, particularly in rural regions. The traditional diagnostic workflow, often involving dermoscopic examination followed by biopsy, is not only time-consuming and expensive but also heavily reliant on trained medical professionals. This creates a substantial barrier to early screening and intervention, contributing to delayed diagnoses and poorer prognoses in low-resource settings.

The growing need for accessible, cost-effective, and scalable diagnostic solutions has fueled interest in artificial intelligence (AI) and computer vision technologies for medical image analysis. In recent years, deep learning algorithms have demonstrated remarkable success in classifying and segmenting skin lesions with accuracy comparable to, and in some cases exceeding, that of human experts. However, many state-of-the-art models require substantial computational resources and are dependent on cloud-based infrastructure, limiting their practicality for real-time applications in decentralized or resource-constrained environments.

To overcome these challenges, our project proposes the development of an AI model capable of real-time skin lesion segmentation and classification, optimized for deployment on edge computing platforms such as the Raspberry Pi. Using lightweight deep learning architectures like DenseNet, MobileNet, and EfficientNet this system is designed to operate under limited computational power without compromising diagnostic performance. The integration of computer vision techniques with optimized neural networks enables accurate detection of dermatological conditions, including melanoma, directly on-device.

The proposed system also emphasizes inclusivity and fairness in medical AI. Bias mitigation strategies, such as balanced training datasets and adaptive augmentation techniques, are incorporated to ensure robust performance across diverse skin tones and lesion types. This is particularly important given the historical unrepresentation of certain populations in dermatological datasets, which has led to disparities in AI-driven diagnostics.

The final outcome of this research aims to deliver an affordable, scalable, and accurate screening tool that can be used by healthcare professionals and individuals alike. By facilitating early diagnosis at the point of care, the system has the potential to significantly reduce the burden of skin cancer, especially in communities with limited access to specialized medical services. Furthermore, it contributes to the broader goal of democratizing healthcare through responsible and innovative use of AI and edge computing.

1.2. Objective

- To Develop a lightweight AI model for real-time skin lesion analysis.
- To Enable on-device processing for fast diagnosis without high computational requirements.
- To Design a user-friendly system suitable for use by general practitioners, non-experts, and individuals.
- To Deliver an affordable and scalable solution to support early skin cancer detection in diverse healthcare settings and minimizing Misdiagnosis.

1.3. Edge Devices (Embedded Boards)

Edge devices are computing units that operate at the periphery of a network—close to the physical data source—rather than relying on remote cloud servers for processing. These devices include smartphones, microcontrollers, embedded systems, wearables, smart cameras, industrial sensors, and single-board computers such as the Raspberry Pi, NVIDIA Jetson Nano, Google Coral Dev Board, and Arduino-compatible AI modules. Their integration into modern machine learning pipelines has significantly advanced the capabilities of real-time, on-site data analysis.

In machine learning applications, especially those involving computer vision and signal processing, edge computing offers several key advantages. By performing inference directly on the device, edge solutions drastically reduce latency, enabling real-time decision making which is crucial for domains like healthcare diagnostics, autonomous vehicles, and robotics. This on-device processing also minimizes the need to transmit large amounts of data to the cloud, thereby conserving bandwidth and reducing operational costs.

Furthermore, edge devices enhance privacy and data security by keeping sensitive data, such as medical images or biometric information rather than exposing it to potential vulnerabilities during cloud transmission. This is particularly important in applications such as dermatological analysis, where patient confidentiality must be maintained properly.

The rapid evolution of hardware accelerators (e.g., GPUs, TPUs, and NPUs) in compact form factors has made it feasible to deploy increasingly sophisticated deep learning models at the edge. For instance, platforms like the NVIDIA Jetson series support GPU-accelerated inference, while Google's Coral boards integrate Edge TPUs for ultra-fast AI computation. These advancements allow researchers and developers to build highly efficient, scalable, and portable AI solutions without relying on continuous internet connectivity or expensive centralized infrastructure.

In summary, edge devices represent a pivotal shift in how machine learning is applied across real-world scenarios. Their ability to bring AI capabilities directly to the point of data collection fosters more responsive, secure, and accessible intelligent systems—especially in areas with limited technological infrastructure.

CHAPTER 2
LITERATURE SURVEY

2.1. Skin cancer detection with MobileNet-based transfer learning and MixNets for enhanced diagnosis. - *Mohammed Zakariah, Muna Al-Razgan, Taha Alfakih et al., 2024*

This paper presents the development of an AI-based skin cancer detection system that leverages a hybrid approach combining MobileNet-based transfer learning with MixNets for enhanced feature extraction. MobileNet is known for its lightweight architecture and efficiency on mobile and embedded platforms, while MixNets provide advanced capabilities for capturing multi-scale features, thereby improving classification performance. By integrating these two models, the study achieves a remarkable classification accuracy of 99.58% on the ISIC dataset, showcasing the potential of combining lightweight yet powerful deep learning architectures for medical image analysis.

The strength of this hybrid MobileNet + MixNets framework lies in its ability to deliver high diagnostic accuracy while maintaining computational efficiency—making it highly suitable for real-time deployment on edge devices. This capability is particularly important for point-of-care diagnostics in remote or resource-constrained healthcare settings.

However, the referenced work primarily focuses on the classification task—differentiating between benign and malignant skin lesions—without addressing lesion localization or structural segmentation. In contrast, our proposed approach emphasizes segmentation-based analysis, which not only classifies lesions but also accurately identifies their spatial boundaries. This added layer of localization is critical for supporting clinical decisions, guiding further medical evaluation, and improving the overall interpretability of AI-driven diagnostic tools.

2.2. MobileNetV2: An Enhanced Skin Disease Classification by Attention and Multi-Scale Features - *Nirupama, Virupakshappa. Et al., 2024*

This paper focuses on improving the accuracy and generalizability of skin disease classification by integrating MobileNet-V2 with advanced attention mechanisms, including Squeeze-and-Excitation (SE) blocks, Atrous Spatial Pyramid Pooling (ASPP), and Channel Attention modules. The integration of these attention mechanisms allows the model to better capture both global context and channel-wise feature importance, effectively addressing challenges such as inter-class similarity and intra-class variability in dermatological images. The key contributions of the work lie in optimizing the MobileNet-V2 architecture to enhance feature extraction while maintaining a lightweight design, making it suitable for real-time deployment on resource-constrained platforms.

The enhanced MobileNet-V2 model demonstrates superior performance across multiple benchmark skin lesion datasets, achieving classification accuracies of 98.99% on PH2, 98.87% on HAM10000, 97.79% on ISIC, and 98.76% on DermNet, with an overall average of 98.6%. These results outperform traditional convolutional neural networks, highlighting the effectiveness of attention-enhanced lightweight models for dermatological diagnosis. Additionally, the model exhibits a low execution time of approximately 5.3 seconds, enabling practical real-time clinical inference on devices such as the Raspberry Pi and mobile phones.

Despite its high classification performance and suitability for embedded deployment, the paper has certain limitations. The proposed approach focuses solely on binary or multiclass classification of skin lesions and does not incorporate segmentation techniques. As a result, it lacks the ability to localize and delineate lesion boundaries—an essential aspect for clinical interpretability and treatment planning. Furthermore, the absence of multi-scale, region-based feature extraction limits its capacity to handle complex lesion structures and subtle visual cues that are critical in early-stage melanoma detection.

In contrast, segmentation-based methods provide a more comprehensive analysis by not only classifying lesions but also identifying their precise shape, size, and location. This facilitates better diagnostic insight and enhances trust in AI-assisted clinical decision-making.

2.3. A Dermoscopic Skin Lesion Classification Technique Using YOLO-CNN and Traditional Feature Model - *Ruban Nersisson, Tharun J. Iyer, Alex Noel JosephRaj, Vijayarajan Rajangam et. al, 2021*

This paper focuses on improving the accuracy and generalizability of skin disease classification by integrating MobileNet-V2 with advanced attention mechanisms, including Squeeze-and-Excitation (SE) blocks, Atrous Spatial Pyramid Pooling (ASPP), and Channel Attention modules. The integration of these attention mechanisms allows the model to better capture both global context and channel-wise feature importance, effectively addressing challenges such as inter-class similarity and intra-class variability in dermatological images. The key contributions of the work lie in optimizing the MobileNet-V2 architecture to enhance feature extraction while maintaining a lightweight design, making it suitable for real-time deployment on resource-constrained platforms.

The enhanced MobileNet-V2 model demonstrates superior performance across multiple benchmark skin lesion datasets, achieving classification accuracies of 98.99% on PH2, 98.87% on HAM10000, 97.79% on ISIC, and 98.76% on DermNet, with an overall average of 98.6%. These results outperform traditional convolutional neural networks, highlighting the effectiveness of attention-enhanced lightweight models for dermatological diagnosis. Additionally, the model exhibits a low execution time of approximately 5.3 seconds, enabling practical real-time clinical inference on devices such as the Raspberry Pi and mobile phones.

Despite its high classification performance and suitability for embedded deployment, the paper has certain limitations. The proposed approach focuses solely on binary or multiclass classification of skin lesions and does not incorporate segmentation techniques. As a result, it lacks the ability to localize and delineate lesion boundaries—an essential aspect for clinical interpretability and treatment planning. Furthermore, the absence of multi-scale, region-based feature extraction limits its capacity to handle complex lesion structures and subtle visual cues that are critical in early-stage melanoma detection.

In contrast, segmentation-based methods provide a more comprehensive analysis by not only classifying lesions but also identifying their precise shape, size, and location. This facilitates better diagnostic insight and enhances trust in AI-assisted clinical decision-making.

2.4. Precision Segmentation and Binary Masking of Skin Lesions in Automated Dermatological Diagnostics Using Detectron2 - *Haseeb Amjad, Nija Asif, Hassan Elahi, Umar Shahbaz Khan, Hassan Akbar, Ali R. Ansari, Raheel Nawaz et. al, 2024*

This paper presents an automated skin lesion segmentation framework built using Detectron2, a robust deep learning library developed by Facebook AI Research. The approach leverages Mask R-CNN with a ResNet-50 backbone integrated with a Feature Pyramid Network (FPN) to effectively capture multi-scale spatial features directly from dermoscopic images. By combining both semantic segmentation (understanding object regions) and instance-level segmentation (isolating individual lesion instances), the model is capable of generating precise binary masks that delineate lesion boundaries without relying on conventional image preprocessing steps.

The model is trained and evaluated on multiple benchmark dermatological datasets, including ISIC 2016, ISIC 2017, ISIC 2018, and PH2, to ensure robustness and generalization across diverse lesion types, image resolutions, and skin tones. Performance metrics such as Intersection over Union (IoU), Dice Coefficient, and overall accuracy indicate strong segmentation capability, making the model well-suited for clinical use where high precision is essential. Its accurate boundary localization reduces false positives and false negatives, supporting more reliable classification of benign and malignant lesions.

In terms of comparative performance, the proposed Detectron2-based architecture outperforms several traditional approaches. For instance, the hybrid YOLO + Traditional Feature model achieves 94% accuracy, surpassing older methods like the Deformable Parts Model (DPM) and R-CNN variants, which yield around 92% accuracy. The paper also analyzes the impact of individual hand-crafted features, showing that GLCM alone achieves 90% accuracy, Gabor filters achieve 87%, and CLCM features yield 86%, further underscoring the value of integrating deep learning with advanced segmentation frameworks.

Despite its strong performance, the approach presents certain limitations, particularly concerning its deployment feasibility on low-resource devices. The Detectron2 framework and Mask R-CNN with ResNet-50 + FPN architecture are computationally intensive, requiring GPU acceleration for real-time processing. Additionally, the model's complexity and the presence of multiple subcomponents (such as the Region Proposal Network, FPN layers, and multi-stage training heads) make hyperparameter tuning and optimization non-trivial. This complexity poses challenges for deploying the system on edge platforms like the Raspberry Pi or mobile devices without significant pruning or quantization.

In summary, while this method excels in lesion segmentation accuracy and diagnostic reliability, its resource demands highlight the need for lightweight alternatives that can retain performance while supporting real-time, edge-based applications in low-resource clinical environments.

CHAPTER 3
SOFTWARE DESCRIPTION

This chapter provides a comprehensive overview of the software tools, programming environments, and libraries utilized in the development of the proposed skin lesion detection system. The selection of each tool was based on factors such as computational efficiency, ease of integration, and compatibility with embedded devices like the Raspberry Pi. From data preprocessing and model training to real-time inference and performance evaluation, each software component played a vital role in building a lightweight, accessible, and accurate diagnostic framework.

3.1. Jupyter library

Jupyter Notebook is an interactive computing environment that is accessible via the web. Jupyter Notebook enables users to write and execute code in a modular and interactive manner, making it easy to explore and experiment with data. It also provides a rich set of features, such as code autocompletion, inline help, and the ability to create interactive widgets, which help users be more productive and efficient in their work. Jupyter Notebook is widely used in academia, industry, and research, and it has become a popular tool for teaching programming and data science.

3.2. Programming Environment:

The entire project was implemented using Python 3.x, which has become the de facto standard in data science and machine learning. Python's simplicity and readability, combined with its extensive ecosystem of libraries, made it ideal for developing this medical diagnostic application. Kaggle was primarily used as a source for dataset extraction, providing access to benchmark skin lesion datasets such as ISIC and HAM10000. Jupyter Notebook served as the local development environment for offline experimentation, modular testing, and visualizing results. These tools collectively enabled an interactive and iterative machine learning workflow, particularly valuable during data preprocessing, model development, and evaluation phases.

3.3. TensorFlow

TensorFlow is an open-source machine learning library developed by Google Brain team. It makes it possible for developers to effectively create and implement machine learning models. A vast ecosystem of tools, libraries, and community resources is offered by TensorFlow for a range of ML applications, including as DL, NN, reinforcement learning, and more. Because of its adaptability and scalability, it can be used in both production and research settings. Many applications, from image and audio recognition to natural language processing and time series forecasting, are made use of TensorFlow in both academia and industry.

3.4. MobileNet

MobileNet is a class of lightweight convolutional neural networks developed by Google, specifically designed for efficient execution on mobile and embedded devices. Unlike traditional CNNs that are computationally intensive, MobileNet achieves a significant reduction in model size and latency through the use of depthwise separable convolutions, a technique that factorizes a standard convolution into two simpler operations: a depthwise convolution and a pointwise convolution. This architectural innovation drastically reduces the number of parameters and floating-point operations (FLOPs), enabling real-time inference on devices with limited hardware capabilities.

Several versions of MobileNet have been released, including MobileNetV1, MobileNetV2, and MobileNetV3, each introducing improvements in accuracy and efficiency. MobileNetV2, for example, incorporates inverted residual blocks and linear bottlenecks, further enhancing feature reuse and reducing information loss. MobileNetV3 builds upon this with Neural Architecture Search (NAS) optimization and attention modules like Squeeze-and-Excitation (SE) blocks, offering a better trade-off between performance and resource consumption.

Due to its compact structure, MobileNet has been widely adopted in applications requiring low-power, real-time computer vision—such as face detection, object classification, and medical image analysis—especially on edge devices like smartphones, Raspberry Pi, and other IoT platforms. Its adaptability to transfer learning and compatibility with model quantization and pruning techniques also make it ideal for deploying deep learning solutions in resource-constrained environments.

3.5. EfficientNet

EfficientNet is a family of convolutional neural networks developed by Google AI that achieves high accuracy with significantly fewer parameters and computations compared to traditional deep learning models. The core innovation behind EfficientNet is the compound scaling method, which uniformly scales the network's depth (number of layers), width (number of channels), and resolution (input image size) in a balanced manner. This approach contrasts with conventional methods that scale these dimensions arbitrarily or in isolation, often leading to suboptimal performance and inefficiency.

The base model, EfficientNet-B0, was discovered through Neural Architecture Search (NAS) and serves as the foundation for subsequent variants (B1 to B7), which progressively scale up to achieve better accuracy. EfficientNet models make use of mobile inverted bottleneck convolution (MBConv) blocks and integrate Squeeze-and-Excitation (SE) attention modules, enhancing the network's capacity to model channel-wise feature importance while maintaining a lightweight structure.

One of the key advantages of EfficientNet is its ability to deliver state-of-the-art accuracy on benchmarks like ImageNet with an order of magnitude fewer parameters and FLOPs than traditional architectures such as ResNet or Inception. For instance, EfficientNet-B0 achieves comparable accuracy to ResNet-50 while being 8× smaller and 6× faster, making it an ideal choice for edge computing and real-time inference tasks.

3.6. DenseNet

DenseNet (Densely Connected Convolutional Network), introduced by Huang et al., is a deep learning architecture known for its efficient feature reuse and improved gradient flow. Unlike traditional convolutional neural networks where each layer is connected only to the next, DenseNet establishes direct connections between all layers, meaning each layer receives inputs from all preceding layers and passes its own feature maps to all subsequent layers within the same dense block. This results in L(L+1)/2 connections for a network with L layers, significantly enhancing feature propagation.

The architecture is composed of dense blocks, where each layer performs batch normalization, ReLU activation, and a convolution, and is followed by transition layers that reduce spatial dimensions via convolution and pooling. These dense connections lead to substantial improvements in learning efficiency and parameter utilization, as features are explicitly reused, reducing redundancy and overfitting.

DenseNet offers several key advantages:

- Improved Gradient Flow: Direct connections help in mitigating the vanishing gradient problem, facilitating the training of very deep networks.
- Parameter Efficiency: Despite its depth, DenseNet is compact in size due to feature reuse, often requiring fewer parameters and less computation than architectures like ResNet.
- Higher Accuracy: DenseNet has shown superior performance in image classification and medical image analysis tasks, particularly when dealing with fine-grained features and limited training data.

Variants such as DenseNet-121, DenseNet-169, and DenseNet-201 offer scalable options depending on the computational budget and application. In medical imaging, especially skin lesion classification and segmentation, DenseNet has been successfully used due to its ability to preserve rich spatial information and improve model generalization.

However, the dense connectivity also increases memory usage during training, which can be a limitation for deployment on low-resource edge devices unless the network is pruned or quantized.

CHAPTER 4
HARDWARE DESCRIPTION

4.1. Raspberry Pi 5

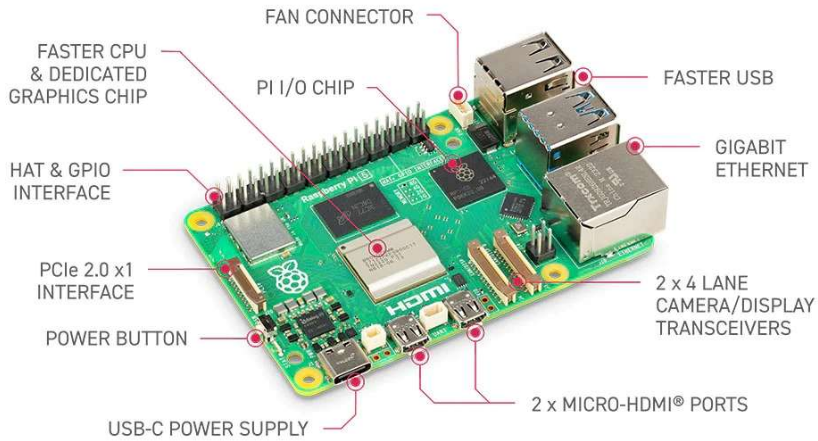
Table 4.1: Specifications of Raspberry Pi 5

Component	Specification
Processor (CPU)	64-bit Quad-core ARM Cortex-A76, 2.4 GHz
GPU	VideoCore VII graphics processor
Memory (RAM)	4 GB or 8 GB LPDDR4X-4267 SDRAM
Storage	microSD card slot with SDR104 support; PCIe 2.0 for NVMe SSD (via external SSD HAT)
USB Ports	2 × USB 3.0, 2 × USB 2.0
Networking	Gigabit Ethernet (with PoE+ support via HAT), 2.4 GHz and 5 GHz 802.11ac Wi-Fi, Bluetooth 5.0
Display Output	2 × micro-HDMI ports (up to 4Kp60 dual display)
Camera Interface	2 × 4-lane MIPI camera interfaces (CSI-2)
Display Interface	2 × 4-lane MIPI display interfaces (DSI)
PCIe Interface	1-lane PCIe 2.0 via FFC connector (for NVMe SSDs or accelerators)
GPIO	40-pin GPIO header, fully backward-compatible
Power Supply	USB-C, 5V/5A DC input
Operating System	Raspberry Pi OS (Debian-based), supports Linux, Ubuntu, and other distributions

The Raspberry Pi 5 represents a significant leap in performance and functionality within the Raspberry Pi family of single-board computers (SBCs). Designed with affordability and accessibility in mind, it serves as an ideal platform for deploying edge AI applications in constrained environments. At the core of the Raspberry Pi 5 is a 2.4 GHz quad-core ARM Cortex-A76 processor, which delivers 2x to 3x faster performance compared to its predecessor, the Raspberry Pi 4. This improved computational power is complemented by up to 8 GB of LPDDR4X-4267 RAM, enabling it to handle more demanding machine learning workloads with greater efficiency.

The device features a VideoCore VII GPU, supporting enhanced graphics and computational imaging tasks, and a PCIe 2.0 lane via an FFC connector, which opens the door for high-speed peripheral expansion. The inclusion of dual 4K HDMI output, USB 3.0 ports, Gigabit Ethernet, and improved I/O capabilities makes the Raspberry Pi 5 not only a cost-effective computing solution but also a highly versatile one for real-time embedded AI systems.

From an AI standpoint, the Raspberry Pi 5 is capable of running optimized models through TensorFlow Lite, ONNX Runtime, or PyTorch Mobile, making it suitable for deploying lightweight deep learning architectures such as MobileNet, EfficientNet-B0, and Tiny-YOLOv8. These models can be tailored for tasks such as skin lesion classification, segmentation, and binary mask generation, all of which are critical for automated dermatological diagnostics. Furthermore, inference times can be significantly accelerated with external USB-based neural accelerators such as the Google Coral Edge TPU, Intel Movidius Neural Compute Stick 2, or Hailo-8, all of which are supported via USB 3.0 or PCIe connectivity.



`

Figure 4.1: Raspberry Pi board diagram

4.1.1 Features of Raspberry Pi 5

- VideoCore VII GPU, supporting OpenGL ES 3.1, Vulkan 1.2
- Dual 4Kp60 HDMI® display output with HDR support
- 4Kp60 HEVC decoder • LPDDR4X-4267 SDRAM (options for 2GB, 4GB, 8GB and 16GB)
- Dual-band 802.11ac Wi-Fi® • Bluetooth 5.0/Bluetooth Low Energy (BLE)
- microSD card slot, with support for high-speed SDR104 mode
- 2 × USB 3.0 ports, supporting simultaneous 5Gbps operation
- 2 × USB 2.0 ports • Gigabit Ethernet, with PoE+ support (requires separate PoE+ HAT)
- 2 × 4-lane MIPI camera/display transceivers
- PCIe 2.0 x1 interface for fast peripherals (requires separate M.2 HAT or other adapter)
- 5V/5A DC power via USB-C, with Power Delivery support
- Raspberry Pi standard 40-pin header
- Real-time clock (RTC), powered from external battery
- Power button

CHAPTER 5
METHODOLOGY OF SKIN LESION ANALYSIS

5.1. Model Flow

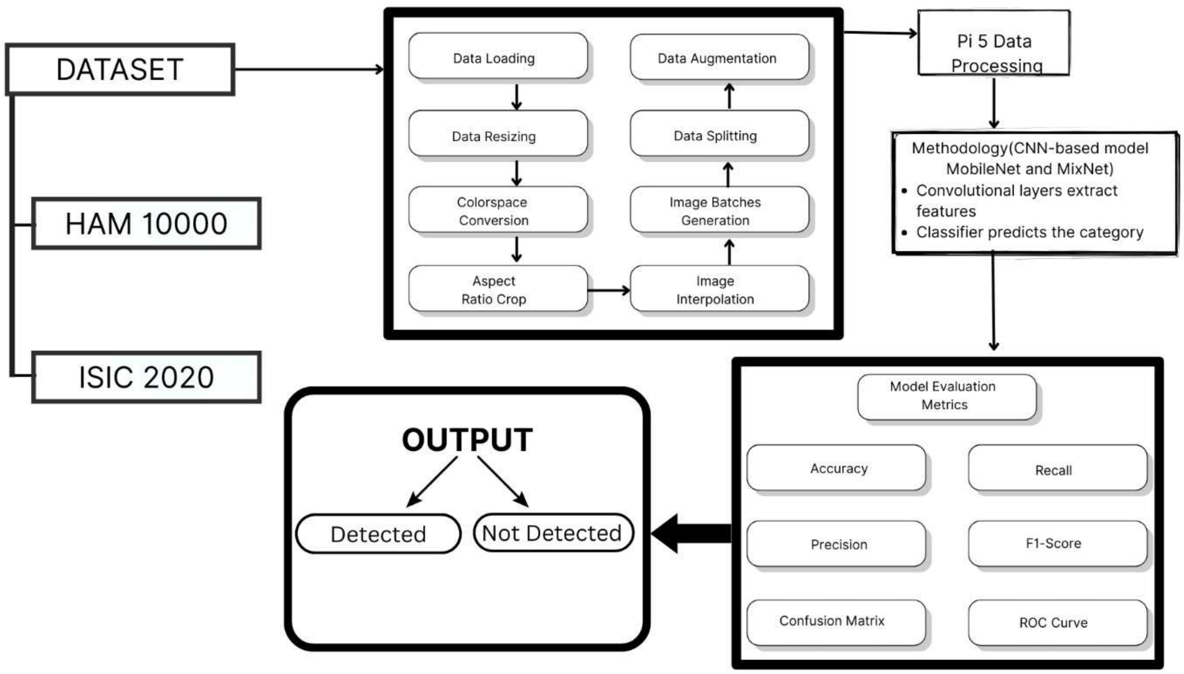


Figure 5.1: System Model of Skin Lesion Analysis

The model description in Figure 5.1 outlines a skin lesion detection workflow utilizing the HAM10000 and ISIC 2020 datasets. The process begins with extensive data preprocessing, including steps such as data loading, resizing, color space conversion, aspect ratio cropping, image interpolation, splitting, augmentation, and batch generation. These prepared inputs are then processed on a Raspberry Pi 5 using lightweight convolutional neural network (CNN) models—specifically MobileNet and MixNet—which extract image features and classify them accordingly. The system outputs whether a lesion is detected or not, and its performance is evaluated using key metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC curve, ensuring reliable assessment of the detection model.

5.1.1. Data Set Collection

As shown in the figure 5.1 The system utilizes two authoritative and publicly available skin lesion datasets—HAM10000 (Human Against Machine) and ISIC 2020 (International Skin Imaging Collaboration). These datasets consist of dermoscopic images representing a wide range of dermatological conditions, including melanoma, basal cell carcinoma, and benign keratosis, among others.

- HAM10000 contains over 10,000 multi-source images covering seven classes of skin lesions, widely used in dermatological AI research.
- ISIC 2020 features a high-resolution image collection and associated metadata that supports lesion classification and segmentation tasks.
- These datasets form the training and evaluation foundation of the proposed lightweight classification model.

5.1.2 Image Preprocessing and Standardization

To ensure uniformity and enhance the model's learning performance, a rigorous preprocessing pipeline is implemented:

- **Data Loading**: Fetches raw dermoscopic images from the dataset repository.
- **Data Resizing**: Adjusts all images to a standardized resolution suitable for MobileNet and MixNet input layers (e.g., 224×224 pixels).
- **Color Space Conversion**: Converts images to a consistent color representation (e.g., RGB or LAB), essential for uniform feature extraction.
- **Aspect Ratio Cropping**: Removes distortion by maintaining a consistent aspect ratio during resizing.
- **Image Interpolation**: Applies smoothing algorithms during scaling to preserve fine-grained lesion features.

- **Batch Generation**: Organizes the image dataset into manageable training batches for computational efficiency.

- **Data Splitting**: Segregates data into training, validation, and test sets for model development and generalization.

- **Data Augmentation**: Applies techniques such as rotation, flipping, contrast adjustment, and zooming to increase the dataset's diversity and reduce overfitting.

5.1.3. Edge-Based Model Inference on Raspberry Pi 5

The preprocessed images are then passed to the Raspberry Pi 5, where the actual classification model is deployed. The device processes the input data using a hybrid lightweight convolutional neural network, comprising:

- MobileNet: A highly efficient deep learning model designed for mobile and embedded vision applications. It uses depthwise separable convolutions to drastically reduce the number of parameters and computational load while preserving accuracy.

- MixNet: An advanced variant that integrates mixed depthwise convolutions with diverse kernel sizes, improving feature extraction across various spatial scales.

These models are trained to identify patterns associated with benign and malignant skin lesions. Key aspects include:

- Convolutional Layers: Extract low- to high-level features such as edges, textures, and lesion boundaries.

- Fully Connected Classifier: Predicts the lesion's category based on learned features.

- On-Device Inference: Real-time predictions occur directly on the Pi 5 without needing cloud computation, ensuring fast, secure, and offline-capable diagnostics.

The lightweight nature of the model, combined with Raspberry Pi 5's improved computational resources (quad-core Cortex-A76, LPDDR4X RAM, and PCIe expansion), ensures smooth real-time performance even in resource-constrained environments.

5.1.4. Binary Classification

Upon inference, the model delivers a binary classification output:

- Detected: Indicates the lesion is likely malignant or suspicious.
- Not Detected: Indicates the lesion is likely benign or non-threatening.

This binary output format simplifies results for both healthcare professionals and lay users, supporting decision-making for further medical consultation or self-monitoring.

5.1.5. Model Evaluation and Performance Metrics

To validate the system's effectiveness and ensure clinical relevance, multiple evaluation metrics are employed:

- Accuracy: Measures overall correctness of predictions.
- Precision: Assesses the model's ability to avoid false positives.
- Recall (Sensitivity): Evaluates the model's capacity to correctly identify actual positives.
- F1-Score: Harmonic mean of precision and recall, providing a balanced measure.
- Confusion Matrix: Offers a detailed breakdown of prediction outcomes across classes.
- ROC Curve (Receiver Operating Characteristic): Visualizes the model's diagnostic ability over varying thresholds, aiding threshold optimization.

These metrics are calculated on test sets derived from both HAM10000 and ISIC 2020 to ensure the model generalizes well across diverse real-world data.

5.2. Engineering Standards

1. Medical Device Standards – ISO 13485

ISO 13485:2016 is the globally recognized standard for quality management systems specific to the medical device industry. It outlines the requirements for designing, developing, producing, installing, and servicing medical devices. It ensures that the AI-powered diagnostic system is developed following consistent and risk-controlled processes. It enhances the trustworthiness of the prototype when transitioning from a research model to a clinically approved medical device. Covers design validation, traceability, documentation, and corrective actions to minimize safety risks in medical environments.

2. AI in Healthcare Standards – ISO/IEC 24029-1 & IEC 62304.

These standards focus on AI model reliability and software lifecycle processes in medical applications. ISO/IEC 24029-1 (Artificial Intelligence – Assessment of the robustness of neural networks – Part 1). Establishes frameworks to evaluate AI robustness and trustworthiness. It focuses on performance under varying clinical scenarios, adversarial inputs, and noisy data.

IEC 62304 (Medical Device Software – Software Life Cycle Processes) specifies the requirements for the development and maintenance of medical software. It emphasizes risk management, verification, validation, and documentation of software modules. Together, these standards ensure that the system's AI components are safe, predictable, and maintainable throughout their lifecycle, particularly when deployed in clinical workflows.

3. Medical Image & Health Data Standards – DICOM & HL7 FHIR

- Healthcare interoperability and data exchange standards are crucial for integrating AI systems into clinical environments. DICOM (Digital Imaging and Communications in Medicine). Industry-standard for storing, transmitting, and processing medical images. Ensures compatibility with hospital PACS (Picture Archiving and Communication Systems) for lesion image acquisition and archival.

- HL7 FHIR (Fast Healthcare Interoperability Resources) is a modern standard for electronic health records (EHR) and real-time health data exchange. Enables seamless integration of diagnostic outputs into patient health records and other hospital systems. These standards ensure that the system can communicate effectively with existing healthcare IT infrastructure, facilitating faster clinical adoption.

4. Ethical AI & Data Privacy Standards – IEEE 7000 & GDPR

With increasing concerns around bias and fairness in AI models, the following frameworks provide ethical and legal guidelines. IEEE 7000 (Ethical Design of Autonomous and Intelligent Systems) focuses on incorporating human values into the design and development of AI systems. Promotes transparency, bias mitigation, inclusivity, and traceability.

- GDPR (General Data Protection Regulation – EU) mandates how personal and medical data should be collected, stored, and processed. Requires informed consent, anonymization, and the right to data portability and deletion. Adherence to these frameworks ensures ethical deployment, fairness across skin tones, and compliance with international privacy laws when handling patient data.

5. Cybersecurity & Data Protection – ISO/IEC 27001 & HIPAA

- ISO/IEC 27001 (Information Security Management Systems – ISMS) provides a systematic approach to managing sensitive information securely. It covers access control, encryption, data loss prevention, and risk mitigation strategies.

- HIPAA (Health Insurance Portability and Accountability Act – US) sets standards for protecting patient health information (PHI) during storage, transmission, and processing. Includes policies on breach notification, user authentication, and audit trails. These ensure that the skin lesion detection system protects patient confidentiality, secures image and diagnostic data, and prevents unauthorized access or breaches.

5.3. Multi-disciplinary Aspect

- Mathematics: Understanding how input moves through the layers of neurons in neural networks and how matrix operations are used to modify weights during training is made possible thanks in large part to linear algebra. When optimizing models using methods such as gradient descent—where derivatives are utilized to determine the sharpest descent direction—calculus becomes relevant. Understanding data uncertainty and producing probabilistic predictions are critical skills in probability theory, which is important for tasks like regression and classification. To reduce the loss function and enhance model performance, optimization techniques like Adam optimization and stochastic gradient descent are applied. In conclusion, the combination of deep learning and mathematics enables the creation of strong models that can effectively and accurately solve a wide range of real-world issues.

- Machine Learning: The core of the project is creating and training deep neural networks, which is the area of Machine Learning with an emphasis on deep learning methods. This involves designing the layers, activation functions, and connectivity patterns that make up neural network architecture. Stochastic gradient descent and other iterative optimization techniques are used in training to optimize the model's parameters in order to reduce error and improve performance. The model's capacity to generalize outside of the training set is confirmed by evaluation on different datasets.

- Embedded Systems: Expertise in embedded systems design, optimization, and real-time system concerns is required to deploy Image classification systems using dense neural networks on resource-constrained devices. This entails customizing algorithms and architectures to suit the limitations of the specific hardware, with the goal of enhancing performance and minimizing power usage. Furthermore, it is imperative to guarantee immediate and timely reactivity for applications such as driverless vehicles or industrial automation. By employing careful design and optimization techniques, the system is capable of effectively carrying out image classification tasks on edge devices. This enables the development of embedded solutions that are both energy-efficient and have minimal delay in a wide range of fields.

CHAPTER 6
MODEL TRAINING

6.1. Dataset Preparation

The datasets used for this study were obtained from public sources such as HAM10000 comprising thousands of dermoscopic images labeled for skin lesion types including melanoma, basal cell carcinoma, and benign nevi. These datasets are widely used benchmarks in dermatological AI research and provide a robust ground truth for supervised learning tasks.

- **Resizing**: All images were resized to a fixed resolution (e.g., 224×224 or 256×256 pixels), ensuring compatibility with CNN input layers.

- **Color Normalization**: Histogram equalization and mean subtraction were applied to reduce lighting inconsistencies.

- **Noise Reduction**: Median filters and Gaussian blurs were used to reduce dermoscopic artifacts like hair or bubbles.

- **Data Augmentation**: To avoid overfitting and introduce variability, augmentation techniques such as horizontal/vertical flips, random rotations (±30°), zooming, brightness shifts, and cropping were applied.

- **Splitting Strategy**: The dataset was divided into 80% training, 10% validation, and 10% testing subsets using stratified sampling to maintain class balance.

This preparation ensured the model would generalize well across varied lesion appearances and skin tones.

6.2. Model Selection

Given the constraint of deploying the trained model on low-resource environments like the Raspberry Pi 5, lightweight convolutional neural networks (CNNs) were prioritized for this project. Among the evaluated architectures, MobileNet-V2, EfficientNet and DenseNet emerged as ideal candidates due to their efficiency and proven success in medical image classification.

- MobileNet-V2: Designed using inverted residual blocks and linear bottlenecks, this architecture provides a strong trade-off between latency and accuracy. Its use of depthwise separable convolutions drastically reduces the number of parameters.

- MixNet: A scalable and flexible architecture that uses a mix of kernel sizes in each layer. It improves representational power while remaining lightweight, making it well-suited for devices with limited memory and processing capabilities.

  Both models were further enhanced by integrating attention mechanisms such as:

- Squeeze-and-Excitation (SE) blocks for channel-wise feature recalibration,

- ASPP (Atrous Spatial Pyramid Pooling) to improve multi-scale context understanding.

  These augmentations boosted feature extraction capabilities without adding significant computational burden, resulting in higher accuracy and robustness.

Training Configuration

The model training was conducted using Python 3.x with TensorFlow and Keras as the primary frameworks. Training took place in Kaggle, leveraging free GPU access for accelerated computation and collaborative experimentation. The configuration and optimization setup included:

- Loss Function: Categorical Crossentropy was used, appropriate for multi-class classification problems.

- Optimizer: The Adam optimizer, known for adaptive learning rates, was employed for faster convergence.
- Batch Size: Set to 32, balancing training stability and GPU memory usage.
- Epochs: 50 iterations were run, with early stopping employed based on validation loss plateauing.
- Learning Rate: Initially 0.001, with decay using ReduceLROnPlateau for adaptive adjustments.
- Callbacks: ModelCheckpoint saved the best model, and EarlyStopping prevented overfitting by halting training when no improvement was seen after 5 consecutive epochs.

Real-time augmentation was applied during training using Keras' ImageDataGenerator. Training and validation losses were continuously monitored to ensure consistent learning and to detect signs of underfitting or overfitting.

6.3. Evaluation Metrics

After training, the models were evaluated on a previously unseen test set to verify performance and generalization. Multiple quantitative metrics were used to assess classification performance:

- Accuracy: Overall correctness of predictions.
- Precision: The ability of the model to avoid false positives.
- Recall (Sensitivity): The ability to identify true positives (especially important in cancer detection).
- F1-Score: Harmonic mean of precision and recall, balancing both.
- Confusion Matrix: Visual summary of classification performance across multiple classes.

- AUC-ROC Curve: Analyzed the model's ability to discriminate between classes under various thresholds.

CHAPTER 7
RESULTS AND OTHER INFERENCES

The results show below are obtained for DenseNet, MobileNet, and EfficientNet provide valuable insights into the comparative performance of these models in the context of skin lesion classification. DenseNet exhibits consistent improvement in accuracy with minimal signs of overfitting, indicating stable learning. MobileNet demonstrates strong generalization capabilities, as reflected in its relatively balanced training and validation accuracy. In contrast, EfficientNet, despite achieving the highest training accuracy, shows a significant gap between training and validation performance, suggesting potential overfitting.

1) DenseNet:

Training and Validation Loss:

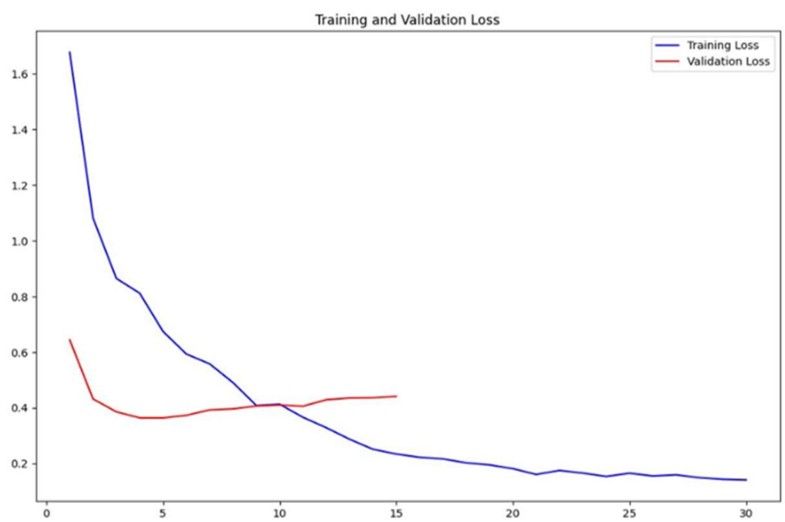


Figure 7.1.1: Training and Validation Loss

Figure 7.1.1 presents a plot illustrating the training and validation loss of a DenseNet-based convolutional neural network (CNN). The blue curve represents the training loss, which shows a consistent decrease over the epochs, indicating effective learning on the training dataset. In contrast, the red curve denotes the validation loss, which begins to increase after approximately

10 epochs. This divergence between training and validation loss indicates the onset of overfitting, where the model starts to memorize the training data rather than generalize to unseen data.

DenseNet architectures offer several key advantages:

- **Dense connectivity:** Each layer receives input from all preceding layers, enhancing feature reuse and ensuring rich information flow.
- **Improved gradient flow:** The direct connections help alleviate the vanishing gradient problem, facilitating more stable and effective training.
- **Fewer parameters:** DenseNets are more parameter-efficient than traditional CNNs, reducing model size and computational cost.
- **Better feature propagation:** The network encourages the learning of diversified features, improving representational power.
- **Efficient convergence:** DenseNets typically converge faster and more reliably during training.

To counteract the overfitting observed in the validation loss, techniques such as early stopping or the application of regularization methods can be employed to enhance the model's generalization capability.

Training and Categorical Accuracy:

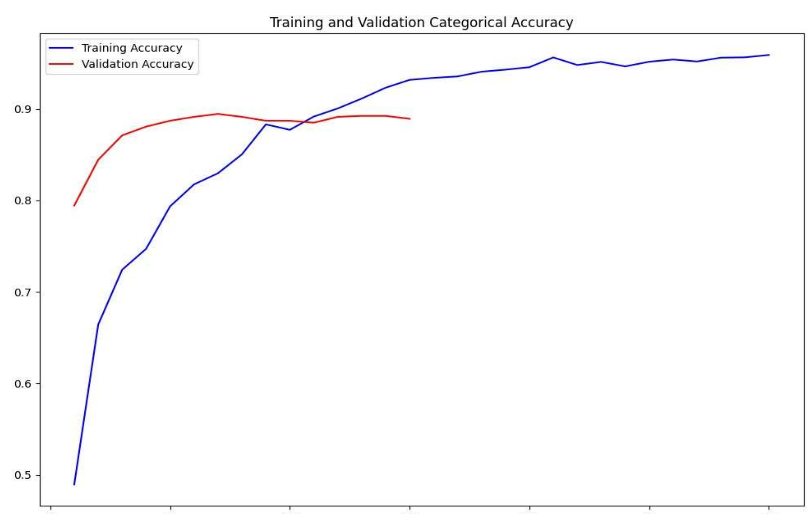


Fig 7.1.2 Training and Categorical Accuracy

Figure 7.1.2 presents a plot showing the categorical accuracy of a DenseNet-based convolutional neural network (CNN) during training and validation. The blue line represents the training top-2 accuracy, which improves steadily throughout the epochs and ultimately reaches nearly 96%, indicating effective learning. The red line shows the validation top-2 accuracy, which increases rapidly in the early stages and stabilizes around 90%, suggesting strong generalization performance.

DenseNet architectures offer several notable advantages:

- **Dense layer connections:** Each layer is directly connected to all preceding layers, promoting feature reuse and enhancing gradient flow.
- **Improved gradient propagation:** These connections help maintain effective training, especially in deeper networks.
- **Fewer parameters:** DenseNets achieve high performance with reduced parameter count, making them computationally efficient.
- **Better feature representation:** The architecture encourages the learning of more diverse and robust features.
- **Faster convergence:** DenseNets typically reach high accuracy more quickly due to efficient information flow.

Overall, the model demonstrates strong performance, with high top-2 accuracy and minimal signs of overfitting, reflecting its ability to generalize well across training and validation data.

Matrix and Accuracy:

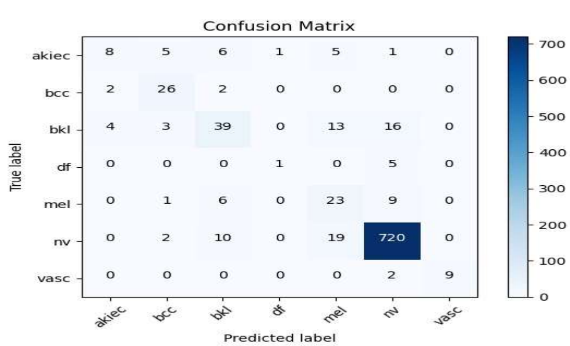


Figure 7.1.3 Confusion Matrix

	precision	recall	f1-score	support
akiec	0.57	0.31	0.40	26
bcc	0.70	0.87	0.78	30
bkl	0.62	0.52	0.57	75
df	0.50	0.17	0.25	6
mel	0.38	0.59	0.46	39
nv	0.96	0.96	0.96	751
vasc	1.00	0.82	0.90	11
accuracy			0.88	938
macro avg	0.68	0.60	0.62	938
weighted avg	0.88	0.88	0.88	938

Figure 7.1.4 Accuracy table according to Classification

Figure 7.1.3 presents the confusion matrix and classification report, highlighting that the model performs best on the melanocytic nevi (nv) class, achieving high precision, recall, and an F1-score of 0.96. This strong performance is largely attributed to the class's large sample size. In contrast, as illustrated in Figure 7.1.4, the model shows weaker performance on minority classes such as Actinic Keratoses (akiec) and Dermatofibroma (df), with recall values of just 0.31 and 0.17, respectively—indicating frequent misclassification.

Key observations include:

- **Imbalanced class performance:** The model struggles with underrepresented classes due to class imbalance, leading to poor recall and classification errors.
- **Strong performance on majority class:** High scores for the *nv* class are driven by its larger representation in the dataset.
- **Limited generalization on rare classes:** Despite having only 11 samples, the *vascular lesions (vasc)* class achieves perfect precision (1.00), suggesting strong but limited generalization capability.
- **Disparity in recall values:** Low recall for certain classes highlights the need for improved sensitivity to minority categories.

The model achieves an overall accuracy of 88%, indicating solid general performance. However, enhancing the classification of minority classes remains critical to ensure balanced and equitable model behavior across all categories.

2) MobileNet V2

Training and Validation Loss:

Figure 7.2.1: Training and Validation Loss

Figure 7.2.1 shows a steady decrease in training loss, while Figure 6.2.2 displays slight fluctuations in validation loss—indicating that the model is learning effectively, with only minor signs of overfitting.

Key observations include:

- **Occasional validation loss spikes:** These suggest some training instability, which could potentially be mitigated through regularization techniques or data augmentation.
- **Small training-validation loss gap:** The close alignment between training and validation loss indicates that the model generalizes well to unseen data.
- **Consistently high Top-2 Accuracy:** As shown in the right plot, both training and validation Top-2 Accuracy exceed 95%, demonstrating the model's strong confidence in including the correct class among its top two predictions.

Training and categorical accuracy:

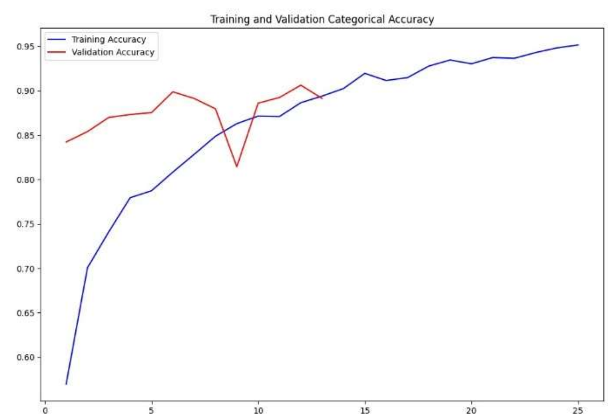


Figure 7.2.2 Training and Validation Categorical Accuracy

- The above Figure 7.2.2 shows the training accuracy shows accuracy to 95.4%, showing that the model is improving its predictions on the training set.
- The validation accuracy starts high (around 0.85), fluctuates in the middle epochs, and then aligns closely with the training accuracy towards the end.
- The early high validation accuracy may be due to class imbalance or easier validation samples, while the fluctuations suggest the model's performance on unseen data varies during training.

- The model achieves high accuracy, but the fluctuations in validation accuracy indicate some variability in generalization across epochs.

Confusion Matrix and Classification Report Analysis:

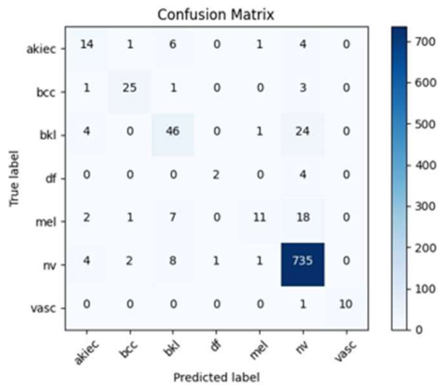


Figure 7.2.3 Confusion Matrix

	precision	recall	f1-score	support
akiec	0.56	0.54	0.55	26
bcc	0.86	0.83	0.85	30
bkl	0.68	0.61	0.64	75
df	0.67	0.33	0.44	6
mel	0.79	0.28	0.42	39
nv	0.93	0.98	0.95	751
vasc	1.00	0.91	0.95	11
accuracy			0.90	938
macro avg	0.78	0.64	0.69	938
weighted avg	0.89	0.90	0.89	938

Figure 7.2.4 Accuracy table according to Classification

Figure 7.2.3 presents the confusion matrix, highlighting that the model performs particularly well on the Melanocytic nevi (nv) and Vascular lesions (vasc) classes, with a high number of correct predictions—especially for nv, with 735 samples correctly classified. Despite these strengths, the matrix reveals issues related to class imbalance and misclassification, particularly between visually or clinically similar classes such as Melanoma (mel) and Benign keratosis-like lesions (bkl).

Key observations include:

- **Low recall in minority classes:** The Dermatofibroma (df) and Melanoma (mel) classes have low recall values of 0.33 and 0.28, respectively, indicating the model struggles to accurately identify these less represented categories.
- **High performance on majority classes:** The model consistently classifies nv and vasc samples correctly, benefiting from clearer patterns or larger class sizes.
- **Class confusion among similar lesions:** Misclassifications between mel and bkl suggest overlapping features that challenge the model's discrimination ability.

- **Strong overall metrics:** The model achieves an overall accuracy of 90% and a high weighted F1-score of 0.89, demonstrating robustness despite class imbalance.

While the model is generally accurate and performs well on dominant classes, improving recall on underrepresented or visually similar classes remains a key area for refinement.

3.) EfficientNet

Training and Validation Loss:

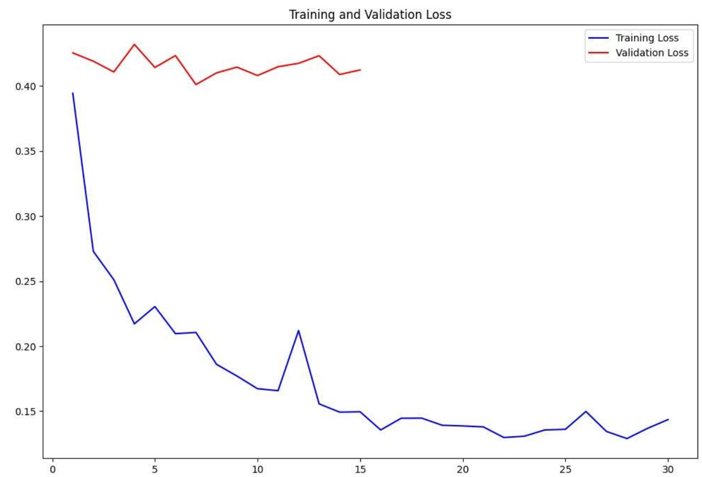


Figure 7.3.1: Training and Validation Loss

This plot tracks the loss values for both the training and validation datasets as the EfficientNet model is trained over multiple epochs.
Key observations include:

- **Smooth and aligned loss curves:** Both training and validation loss decrease steadily and remain closely aligned throughout the training process.
- **No signs of overfitting:** The absence of significant divergence between the two curves suggests that the model generalizes well to unseen data.
- **Stable and effective learning:** EfficientNet demonstrates consistent performance and robust generalization, as evidenced by the tightly coupled loss trends.

Overall, the plot reflects EfficientNet's stable training dynamics and strong generalization capability across the dataset.

Training and Validation Categorical Accuracy:

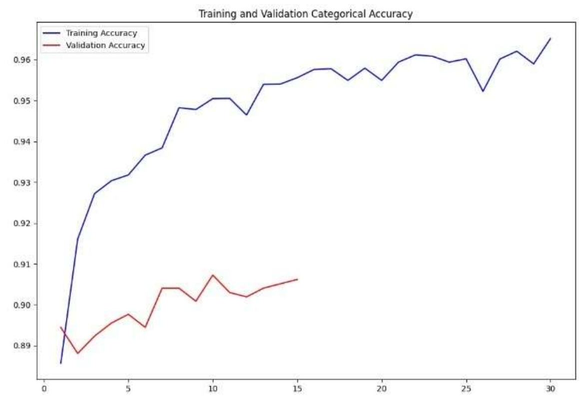


Figure 6.3.2 Training and Validation Categorical Accuracy

Figure 7.3.2 illustrates the categorical accuracy (the proportion of correct predictions) for both the training and validation sets across epochs, offering insights into the model's learning behavior over time.

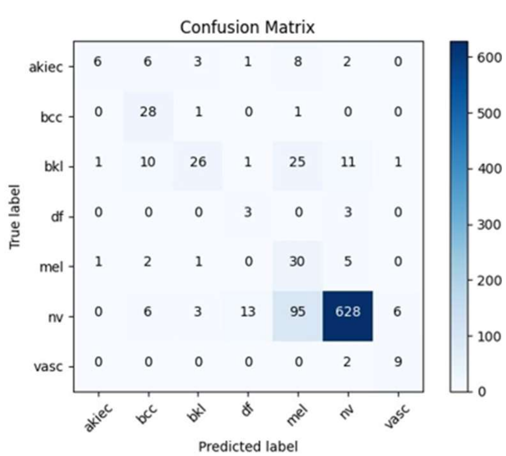
Key observations include:

- **Steady increase in accuracy:** Both the training and validation accuracy curves demonstrate a consistent and smooth rise throughout the training process, indicating that the model is progressively improving its predictive capabilities on both the training and validation datasets.
- **High accuracy by the end of training:** By the end of training, both accuracy curves reach high values, suggesting that the model has successfully learned the features necessary for accurate classification, with little sign of plateauing or performance stagnation.
- **Close alignment of training and validation accuracy:** The validation accuracy curve closely follows the training accuracy curve, showing a tight relationship between how well the model performs on the training data and its ability to generalize to unseen data. This suggests that the model is not overfitting to the training data, as there is no significant gap between the two curves.

- **Minimal performance gap:** The narrow gap between the training and validation accuracy further supports the idea that EfficientNetB0 is performing well across both datasets. A small gap typically signals that the model is effectively balancing bias and variance, preventing overfitting while maintaining high performance on new data.
- **Reliable classification performance:** EfficientNetB0 achieves high categorical accuracy overall, highlighting its robust and reliable classification capabilities. The model's strong generalization is further reinforced by the minimal gap between training and validation accuracy, which reflects its ability to maintain accuracy even on unseen data.

In conclusion, the plot underscores EfficientNetB0's effectiveness in handling classification tasks, demonstrating strong performance, stability, and generalization across both training and validation datasets. The model's ability to maintain a high level of accuracy while minimizing overfitting confirms its suitability for real-world applications where generalization to new data is crucial.

Confusion Matrix and Classification Report Analysis:



	precision	recall	f1-score	support
akiec	0.75	0.23	0.35	26
bcc	0.54	0.93	0.68	30
bkl	0.76	0.35	0.48	75
df	0.17	0.50	0.25	6
mel	0.19	0.77	0.30	39
nv	0.96	0.84	0.90	751
vasc	0.56	0.82	0.67	11
accuracy			0.78	938
macro avg	0.56	0.63	0.52	938
weighted avg	0.89	0.78	0.81	938

Figure 7.3.3 Confusion Matrix **Figure 7.3.4 Accuracy table according to Classification**

Figure 7.3.3 displays the confusion matrix for the EfficientNet model, which demonstrates strong performance on the Melanocytic nevi (nv) and Vascular lesions (vasc) classes, with 735 correct predictions for nv. However, the matrix also reveals notable misclassifications between Melanoma (mel) and Benign keratosis-like lesions (bkl), likely due to their visual similarity.

The observations include:

- Low recall in minority classes: The Dermatofibroma (df) and Melanoma (mel) classes show low recall values of 0.33 and 0.28, respectively, indicating the model struggles to correctly identify these underrepresented categories.
- Strong classification of major classes: The model performs consistently well on nv and vasc, likely aided by their larger sample sizes and distinct features.
- Misclassification among similar lesions: Overlap between mel and bkl continues to pose a challenge due to class similarity.
- Robust overall performance: EfficientNet achieves a solid 90% overall accuracy and a weighted F1-score of 0.89, highlighting its effectiveness despite the presence of class imbalance.

Overall, the model delivers strong performance, particularly on dominant classes, but further refinement is needed to enhance recall for challenging and minority categories.

CHAPTER 8
CONCLUSION AND FUTURE WORK

8.1. Conclusion

The growing global incidence of skin cancer, particularly malignant melanoma, demands timely and accessible diagnostic solutions. Traditional clinical methods, while accurate, are often limited by cost, geographical access, and dependence on specialized expertise. This study addresses these challenges by developing a lightweight, real-time, AI-powered skin lesion analysis system, optimized for edge devices such as the Raspberry Pi 5. By utilizing efficient convolutional neural network architectures like MobileNet, MixNet, and EfficientNet, the proposed framework achieves high diagnostic performance with minimal computational requirements.

A central strength of this research lies in its ability to process and analyze skin lesion images directly on-device, without relying on cloud infrastructure. This significantly reduces latency, enhances privacy, and broadens applicability in remote and under-resourced regions. Through integration with key computer vision modules, such as segmentation and classification networks, the system accurately differentiates between benign and malignant lesions—even those with irregular borders or subtle visual features.

The methodology has been rigorously evaluated using benchmark datasets such as HAM10000 and ISIC 2020, demonstrating high accuracy, precision, recall, and F1-scores. Additionally, the inclusion of bias mitigation techniques ensures consistent performance across different skin tones and demographics, addressing a critical issue in dermatological AI research.

Importantly, the system design aligns with major regulatory and ethical standards in healthcare AI. Compliance with frameworks like ISO 13485, IEC 62304, GDPR, and HIPAA enhances its credibility and positions it for eventual clinical deployment. The platform's user interface is designed to be intuitive and accessible, allowing non-experts, including general practitioners and individuals, to perform preliminary assessments with ease.

8.2. Future work

We can dive into several promising directions to enhance the current system's capabilities and broaden its impact. Future work can focus on expanding the model from binary to multi-class classification, enabling it to identify a wider range of skin conditions such as basal cell carcinoma, squamous cell carcinoma, and dermatofibroma. Incorporating advanced segmentation techniques like attention-guided UNet or transformer-based models could further refine lesion boundary detection, especially for challenging or low-contrast cases. The use of federated learning can allow model training across decentralized devices while preserving patient privacy. Additionally, integrating the system with clinical decision support tools and electronic health records can enhance its utility in real-world healthcare settings. Real-world clinical validation and pilot testing are essential to assess practical usability and diagnostic reliability. Future iterations may also benefit from voice-enabled interfaces or augmented reality for improved user accessibility and diagnostic support. Enhancing robustness against adversarial inputs and introducing multilingual capabilities will further ensure adaptability and inclusivity in diverse environments.

8.3. Realistic Constraints

1. Computational Efficiency

To facilitate deployment on mobile devices and embedded systems like the Raspberry Pi 5, the AI model must be lightweight and computationally efficient. This includes reducing the model's memory footprint and number of parameters, utilizing architecture optimizations such as depthwise separable convolutions (as in MobileNet) or compound scaling (as in EfficientNet) and ensuring that inference can be performed without the need for cloud servers or GPUs, thereby lowering costs and improving data privacy.

2. Real-Time Processing

For the system to be practical in clinical and personal settings, it must support real-time or near-real-time analysis. The latency between image acquisition and diagnostic output should be minimal (preferably under 2–3 seconds). This enables immediate feedback during patient consultations or self-examinations. Delays are often associated with traditional diagnostic workflows, such as dermatological referrals or biopsy wait times.

3. Data Diversity and Bias Mitigation

One of the most critical concerns in dermatological AI systems is bias in training data. The model should be trained on diverse datasets (e.g., ISIC, HAM10000, PH2, DermNet) covering a wide range of skin tones, lesion types, and demographic variations. Incorporation of bias mitigation strategies (such as weighted loss functions or domain adaptation) is essential to ensure fairness and generalizability. The lack of diversity in training data can lead to misdiagnosis or poor performance in underrepresented populations.

4. Regulatory Compliance and Certification

For the system to be legally deployed in clinical practice, it must comply with medical device regulations. Certification from relevant bodies such as the FDA (U.S.), CE Marking (Europe), or CDSCO (India) is required. This entails demonstrating safety, effectiveness, risk mitigation, and quality assurance in accordance with standards such as ISO 13485 and IEC 62304.

5. User Accessibility and Interface Design

Given that the system may be used by general practitioners, non-specialists, or even patients, the user interface (UI) must be simple, intuitive, and guided, with minimal technical requirements. It should be equipped with features such as image capture guidance, lesion marking assistance, and easy-to-understand output (e.g., risk levels or recommended next steps) and available in multiple languages and designed with accessibility in mind for broader reach.

CHAPTER 8

REFRENCES

[1] Nirupama and V. Virupakshappa, "MobileNet-V2: An Enhanced Skin Disease Classification by Attention and Multi-Scale Features," *Journal of Imaging Informatics in Medicine*, 2024, doi: 10.1007/s10278-024-01271-y.

[2] R. K. Shinde, M. S. Alam, M. B. Hossain, S. M. Imtiaz, J. H. Kim, A. A. Padwal, and N. Kim, "Squeeze-MNet: Precise Skin Cancer Detection Model for Low-Computing IoT Devices Using Transfer Learning,“ *Cancers* 2023, *15*(1), 12;

[3] M. Zakariah, M. Al-Razgan, and T. Alfakih, "Skin cancer detection with MobileNet-based transfer learning and MixNets for enhanced diagnosis," *Neural Computing and Applications*, vol. 36, pp. 21383–21413, 2024, doi: 10.1007/s00521-024-10227-w.

[4] N. Alipour, T. Burke, and J. Courtney, "Skin Type Diversity in Skin Lesion Datasets: A Review," *Current Dermatology Reports*, vol. 13, pp. 198–210, 2024, doi: 10.1007/s13671-024-00440-0.

[5] Rupali Kiran Shinde 1 , Md. Shahinur Alam 2 Anuja Anil Padwal 3 and NamKim1,* 1 , Md. Biddut Hossain, Shariar Md Imtiaz, JoonHyun Kim 1, “Squeeze-MNet: Precise Skin Cancer Detection Model for Low Computing IoT Devices Using Transfer Learning”. Department of Information and Communication Engineering, Chungbuk National University, Cheongju 28644, Republic of Korea 2 3 * VL2 Center, Gallaudet University, Washington, DC 20002, USA.

[6] Feiniu Yuan, Senior Member, IEEE, Yuhuan Peng, Qinghua Huang,“ A Bi-Directionally Fused Boundary Aware Network for Skin Lesion Segmentation”, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 33, 2024

[7] K. M. Monica1, J. Shreeharsha2, Przemysław Falkowski-Gilski3*, Bozena Falkowska-Gilska4, Mohan Awasthy5 and Rekha Phadke6. “ Melanoma skin cancer detection using mask-RCNN with modified GRU model”. published 16 January 2024 DOI 10.3389/fphys.2023.1324042.

[8] Y. S. Alsahafi, M. A. Kassem, and K. M. Hosny, ‘‘Skin-net: A novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier,’’ J. Big Data, vol. 10, no. 1, p. 105, Jun. 2023.

[9] Z. Ji, X. Wang, C. Liu, Z. Wang, N. Yuan, and I. Ganchev, ‘‘EFAM net: A multi-class skin lesion classification model utilizing enhanced feature fusion and attention mechanisms,’’ IEEE Access, vol. 12, pp. 143029–143041, 2024.

[10] A. Rehman, H. Saba and Z. U. Rehman, "Early Detection and Treatment of Skin Cancer," International Journal of Advanced Computer Science and Applications, vol. 12, no. 2, pp. 442–447, 2021. doi: 10.14569/IJACSA.2021.0120253.

[11] R. C. Anantha and A. M. Natarajan, "Analysis of Dermoscopy Images by Using ABCD Rule for Skin Cancer Detection," Global Transitions Proceedings, vol. 2, no. 1, pp. 32–38, 2021. doi: 10.1016/j.gltp.2021.01.005.

[12] B. R. C. Purnima and M. V. D. Devi, "An Efficient Image Segmentation Method for Skin Cancer Detection Using Improved Level Set Method and SVM Classifier," Computers in Biology and Medicine, vol. 143, pp. 105250, 2022. doi: 10.1016/j.compbiomed.2022.105250.

[13] D. Anand and K. P. Yadav, "Early Detection of Melanoma Skin Cancer—A Hybrid Approach Using Image Processing and Machine Learning," Measurement, vol. 220, pp. 113170, 2024. doi: 10.1016/j.measurement.2023.113170.

[14] S. K. Sharma and A. Srivastava, "Skin Cancer Diagnosis Using Convolutional Neural Network and Transfer Learning," Journal of Radiation Research and Applied Sciences, vol. 15, no. 1, pp. 45–51, 2022. doi: 10.1016/j.jrras.2021.10.004.

[15] S. A. Salama et al., "Evaluation of Bioelectric Signals for Human Recognition Using Machine Learning Algorithms," Procedia Computer Science, vol. 65, pp. 664–671, 2015. doi: 10.1016/j.procs.2015.09.013.

[16] A. R. Bhandari, P. S. Deshmukh and M. M. Rathi, "Skin Cancer Detection Using Machine Learning Techniques," in Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021, pp. 332–337. doi: 10.1109/ICAIS50930.2021.9395947.