A Thesis Submitted to the Sylhet Engineering College for the Degree of
Bachelor of Science in Electrical and Electronic Engineering

LesionNet: A Custom CNN Model for Accurate Multiclass Classification of Dermoscopic Skin Lesions

By

Ashraf Ahmed Chowdhury

Md Fujael Ahmed

&

Rashedul Albab

Supervised by,
Salman Fazle Rabby
Assistant Professor
Department of Electrical and Electronic Engineering
Sylhet Engineering College

JUNE, 2025
Sylhet Engineering College, Sylhet
Affiliated with
Shahjalal University of Science & Technology

The thesis titled **“LesionNet: A Custom CNN Model for Accurate Multiclass Classification of Dermoscopic Skin Lesions”** submitted by **Ashraf Ahmed Chowdhury, Md Fujael Ahmed, Rashedul Albab**; Student ID: **2019338502**, **2019338526** and **2019338548**; Session 2019-2020, to the Department of Electrical and Electronic Engineering, Sylhet Engineering College, has been accepted as satisfactory in partial fulfillment of the requirement for the Degree of Bachelor of Science in Electrical and Electronic Engineering and approved as to its style and contents.

BOARD OF EXAMINERS

______________________ Chairman

Md. Shahid Iqbal
Assistant Professor and Head
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

______________________ Member and Supervisor

Salman Fazle Rabby
Assistant Professor
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

______________________ Member

Apurbo Biswas
Assistant Professor
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

______________________ Member

Md. Ashraful Alam
Lecturer
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

______________________ Member

Mahedi Kamal Ahmed
Lecturer
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

______________________ Member (External)

Arif Ahmmad
Assistant Professor
Department of Electrical and Electronic Engineering
Shahjalal University Of Science and Tecnology, Sylhet.

Acknowledgements

Firstly, we glorify the Almighty Allah, the creator and the sustainer of this world and belongs to the supreme power. He has given us the energy and spirit to fulfill all the requirements for the Degree of Bachelor of Science in Electrical and Electronic Engineering (EEE) including this thesis work. Next, it is one of the greatest pleasures to express our profound sense of obligation to our honorable supervisor **Salman Fazle Rabby**, Assistant Professor, Department of EEE, Sylhet Engineering College of his constructive academic advice, creative guidance, constant encouragement, valuable suggestions and all the other supports throughout the study. We are really benefitted from his excellent supervision. Moreover, it is an honor to express our gratitude to our head of dept. **Md. Shahid Iqbal** and faculty member **Md. Ashraful Alam and Mahedi Kamal Ahmed** for their nice help and cooperation during this journey. We would also like to thank **Md. Janibul Alam Soeb,** Assistant Professor, Dept. of FPM (Farm Power and Machinery), Faculty of Agricultural and Engineering Technology, Sylhet Agricultural University. His valuable discussions and other support during this thesis have made us grateful to him.

We have a tremendous debt to our parents and siblings for everything they've done to provide for us and invest in our lives through their prayers, support, and hard work and also to our classmates and friends in the EEE department at Sylhet Engineering College.

Abstract

Skin cancer is one of the most prevalent and potentially fatal forms of cancer worldwide. The ability to accurately classify different types of skin lesions at an early stage is critical for ensuring timely treatment and improving patient outcomes. This study proposes a deep learning-based approach to automate the classification of dermoscopic images into seven categories of skin lesions, using a custom-designed Convolutional Neural Network (CNN) architecture named LesionNet. The model was trained and evaluated on the HAM10000 dataset, which includes over 10,015 RGB dermoscopic images representing a variety of dermatological conditions. To address the dataset's inherent class imbalance, Random Oversampling was applied only to the training data to avoid biasing the test set. Comprehensive image preprocessing, including resizing, normalization, and augmentation (such as rotation, flipping, and zoom), was performed to improve the model's generalizability. In addition, Exploratory Data Analysis (EDA) was conducted on associated metadata to observe trends in class distribution, patient age, gender, and lesion location. The LesionNet model was trained using categorical cross-entropy loss and optimized with the Adam optimizer. Evaluation was carried out using standard multiclass performance metrics, including accuracy, precision, recall, and F1-score. The proposed model, LesionNet, achieved a test accuracy of 97.12%, showing robust classification performance across all seven lesion types. These results demonstrate that with careful preprocessing and targeted model design, a lightweight customized CNN like LesionNet can serve as an effective tool for automated skin lesion classification, aiding in the development of reliable diagnostic support systems.

***Keywords**: Skin Lesions, Multiclass Classification, LesionNet, Convolutional Neural Network (CNN), Deep learning, Medical Image Analysis, Dermoscopy, Image Augmentation.*

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Overview

Skin cancer is one of the most common and potentially deadly cancers worldwide. The World Health Organization reports 2 to 3 million non-melanoma and over 130,000 melanoma cases annually [1]. Rising incidence is linked to factors such as increased ultraviolet (UV) exposure, ozone depletion, aging populations, and better diagnosis awareness [2]. Melanoma, though less common, is the most aggressive and deadly form but is highly treatable if detected early [3]. Early detection often depends on identifying skin lesions—abnormal skin areas like discolorations, bumps, or growths. These can be benign (e.g., melanocytic nevi) or malignant (e.g., melanoma, basal cell carcinoma) [4]. Proper classification guides treatment, ranging from observation to surgery.

Diagnosing lesions manually is challenging due to their variability in shape, size, texture, pigmentation, and borders, causing frequent diagnostic errors even among experts [5]. There is also high inter-observer variability linked to clinician experience and training [6]. Dermoscopy, a non-invasive technique using a handheld device with magnified, polarized light, reveals subsurface skin features like pigment networks and vascular patterns, improving diagnosis [7]. However, it still requires expert interpretation, which is time-consuming and limited by specialist availability, especially in rural areas [8].

With a global shortage of dermatologists, especially in developing countries, automated, reliable diagnostic tools are urgently needed [9]. Artificial intelligence (AI) and computer-aided diagnosis systems offer promising solutions by objectively analyzing dermoscopic images to assist early detection and reduce healthcare burdens [10]. To be clinically useful, these automated systems must overcome challenges such as: high lesion variability across types, class imbalance with more benign than malignant samples, inconsistent image quality and need for efficient processing on low-resource devices. These issues highlight the necessity of developing advanced deep learning models tailored for real-world clinical data.

1.2 Types of Skin Lesions

Skin lesions are abnormal growths or appearances on the skin and can range from harmless conditions to serious forms of skin cancer. The differentiation between benign and malignant lesions is critical, as it influences the treatment plan and prognosis. In the context of automated image classification, understanding the distinct characteristics of each lesion type is essential to designing an effective and clinically relevant model. This research focuses on the classification of dermoscopic images into seven distinct categories as defined in the HAM10000 dataset. These classes include both benign and malignant lesions, each with unique visual patterns, textures, and clinical implications:

1.2.1 Melanocytic Nevi

These are commonly known as moles and are typically benign. They present as symmetrical, well-defined lesions with uniform pigmentation. Although most nevi are harmless, some may exhibit atypical features that resemble melanoma, making their accurate classification important in clinical practice.

1.2.2 Melanoma

Melanoma is a dangerous and potentially fatal form of skin cancer that arises from the uncontrolled growth of melanocytes. Dermoscopically, melanomas often appear as asymmetrical lesions with irregular borders, variegated pigmentation, and structural disorganization. Early detection is crucial, as melanoma can metastasize quickly if left untreated.

1.2.3 Benign Keratosis-like Lesions

This category includes seborrheic keratoses, solar lentigines, and lichen-planus-like keratoses. These lesions are benign and often appear in older individuals. Their texture may be rough or wart-like, and they can sometimes mimic malignant lesions in appearance, leading to diagnostic confusion.

1.2.4 Basal Cell Carcinoma

BCC is the most common type of skin cancer, arising from basal cells in the epidermis. While it rarely metastasizes, it can cause significant local tissue destruction if not treated. Dermoscopically, BCCs often show arborizing blood vessels, ulceration, and a translucent appearance.

1.2.5 Actinic Keratoses

These are pre-cancerous or early-stage cancerous lesions typically caused by prolonged sun exposure. They present as scaly, red patches or plaques and are at risk of evolving into squamous cell carcinoma. Early identification and treatment are essential to prevent progression.

1.2.6 Vascular Lesions

This class includes angiomas and hemorrhages. Vascular lesions are typically benign but can be confused with melanomas due to their reddish-purple pigmentation. Dermoscopic features often include red lacunae or homogeneous red areas.

1.2.7 Dermatofibroma

Dermatofibromas are benign fibrous nodules commonly found on the lower legs. They are firm to the touch and may show a central white scar-like area surrounded by a pigmented rim. While benign, their appearance can overlap with other pigmented lesions, necessitating careful examination.

Each of these lesion types presents specific challenges for automated classification due to intra-class variability and inter-class similarity. The inclusion of all seven classes in the model design enhances its clinical applicability and reflects the complexity of real-world dermatological diagnosis.

1.3 Deep Learning in Dermatological Diagnosis

Deep learning, a subset of machine learning based on artificial neural networks, has revolutionized the field of medical image analysis, particularly in dermatology. Among deep learning methods, Convolutional Neural Networks (CNNs) have emerged as the most powerful technique for visual pattern recognition tasks, including the classification of skin lesions from dermoscopic images. CNNs are designed to automatically and adaptively learn spatial hierarchies of features through convolutional layers, pooling layers, and fully connected layers. Unlike traditional machine learning models that rely heavily on handcrafted features, CNNs learn relevant features directly from raw pixel data, capturing complex patterns related to texture, color, shape, and edges, which are crucial for distinguishing between benign and malignant lesions. A popular approach to CNN-based diagnosis is transfer learning, where a model pre-trained on large-scale datasets like ImageNet is fine-tuned on medical images. Common architectures used in transfer learning include VGG16, ResNet, DenseNet, and Inception. These models provide a good starting point because they have already learned generic visual features that can be adapted to the specific characteristics of dermoscopic images.

However, despite their success, these pre-trained models have certain limitations in dermatological applications:

- **Domain mismatch:** Pre-trained CNNs are usually trained on natural images containing everyday objects, which differ significantly in texture, color distribution, and structure from medical images like dermoscopy. This mismatch can limit their ability to learn subtle lesion-specific features.

- **Data size constraints:** Medical datasets, especially dermoscopic datasets, tend to be smaller and more imbalanced compared to datasets like ImageNet. Large pre-trained models with millions of parameters can easily overfit when fine-tuned on small datasets, reducing generalization performance.

- **Lack of customization:** Pre-trained models come with fixed architectures and hyperparameters optimized for general image classification tasks. This limits flexibility in tailoring the model complexity and receptive fields to the unique characteristics of skin lesions.

To address these challenges, customized CNN architectures are often designed specifically for dermatological diagnosis. These models offer several advantages:

- **Architecture tailored to dermoscopy:** Custom CNNs can be designed with convolutional layers and filter sizes optimized to capture the distinct visual cues present in dermoscopic images, such as pigment networks, globules, and vascular structures.

- **Control over model complexity:** By adjusting the number of layers, filters, and neurons, customized CNNs can avoid overfitting by maintaining a model size appropriate for the available data volume.

- **Hyperparameter tuning:** Researchers can experiment with different activation functions, learning rates, batch sizes, dropout rates, and optimization algorithms to maximize performance on skin lesion classification tasks.

- **Integration of domain knowledge:** Custom architectures allow the incorporation of domain-specific pre-processing, attention mechanisms, or multi-scale feature extraction tailored to skin lesion morphology.

Recent studies have demonstrated that carefully designed CNNs outperform transfer learning models on dermoscopic image datasets by learning more discriminative features specific to skin lesions. Moreover, customized CNNs can be more computationally efficient, enabling deployment in real-time or resource-limited environments, such as mobile devices used in remote or underserved areas. In summary, while transfer learning has accelerated progress in automated skin cancer diagnosis, the development of customized CNN models is crucial for achieving higher accuracy, robustness, and clinical applicability in dermatology. These models offer the flexibility and precision needed to handle the diverse, complex visual patterns present in skin lesions, ultimately improving early detection and patient outcomes.

1.4 Motivation for the Work

Accurate classification of skin lesions is essential for early detection of skin cancer, but it remains a complex task due to visual similarities across different lesion types and significant variation within the same class. These challenges often lead to misdiagnoses, even among experienced dermatologists. Furthermore, most available datasets suffer from class imbalance, with benign lesions being far more prevalent than malignant ones. This imbalance can skew machine learning models toward the majority class, reducing their effectiveness in identifying critical malignant cases.

Another key concern is the limited availability of dermatologists, especially in rural and underserved regions. Many individuals lack access to timely expert evaluation, which can delay diagnosis and treatment of life-threatening conditions like melanoma. To address these issues, there is a strong need for an automated, reliable, and accessible diagnostic tool that can assist healthcare providers in identifying potentially dangerous lesions at an early stage. This work is motivated by that need.

The primary goal of this thesis is to develop a lightweight and efficient CNN architecture, specifically designed for multiclass skin lesion classification. Unlike large pre-trained models, the proposed model focuses on being resource-efficient and adaptable, making it suitable for deployment on low-power or mobile devices in real-world clinical environments.

1.5 Objectives of the Study

- To design and develop a lightweight CNN architecture, LesionNet, for accurate classification of seven skin lesion types, optimized for low-computation environments like mobile or edge devices, while addressing challenges such as class imbalance and low image resolution through preprocessing and augmentation techniques.
- To benchmark LesionNet against established CNN models (VGG16, DenseNet) and evaluate its performance in terms of accuracy, efficiency, and real-world diagnostic applicability, aiming to establish it as a practical tool for early and reliable skin lesion diagnosis.

Chapter 2: Literature Review

The growing interest in applying artificial intelligence to healthcare has led to a surge in research focused on using deep learning models for skin lesion classification. With skin cancer rates increasing globally, researchers have explored a range of computational methods to assist dermatologists in early and accurate diagnosis. From early techniques based on handcrafted features to the recent explosion of deep neural networks, the evolution of machine learning in dermatology reflects a broader trend toward automation and precision in medical diagnostics.

This chapter reviews existing work on skin lesion analysis, highlighting contributions from both classical machine learning algorithms and modern convolutional neural networks (CNNs). We also examine widely used datasets such as HAM10000, discuss the strengths and limitations of existing models, and identify research gaps that motivate the proposed study.

Shetty et al. [11] proposed a comprehensive classification framework for detecting seven types of pigmented skin lesions using a combination of machine learning and convolutional neural networks (CNN). Their study utilized the HAM10000 dataset and applied techniques such as image preprocessing, data augmentation, and global feature extraction. The authors evaluated various machine learning algorithms including Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Logistic Regression (LR), alongside a customized CNN model. Their results demonstrated that the CNN model achieved the highest accuracy of 95.18%, outperforming all other methods. The study highlights the effectiveness of CNN-based models for early and accurate diagnosis of skin diseases, contributing significantly to computer-aided dermatological diagnostics.

Adegun and Viriri [12] developed a two-stage deep learning framework combining Fully Convolutional Networks (FCNs) with DenseNet for automated lesion segmentation and classification. The FCN encoder–decoder network, enhanced with skip and residual connections, provided pixel-level segmentation refined by a Conditional Random Field (CRF). In the second stage, DenseNet classified the segmented regions, with hyperparameter tuning to optimize performance. Tested on the HAM10000 dataset, the model achieved around 98% accuracy, 98.5% recall, and nearly 99% AUC, demonstrating clinical potential through its integration of segmentation and classification.

Shelatkar et al. [13] proposed a lightweight deep learning model for brain tumor detection using MRI data. Although focused on a different medical domain, their work is notable for employing

YOLOv5 with transfer learning, adapted to the RSNA-MICCAI BraTS 2021 dataset. Their compact architecture achieved 88% precision while remaining computationally efficient, making it suitable for portable diagnostic devices. This study highlights how fine-tuning lightweight architectures can balance accuracy with low resource consumption, a concept relevant to skin lesion classification as well.

Khan et al. [14] introduced an intelligent two-stage framework named *Pixels to Classes*, which unified lesion segmentation and multiclass classification. They employed a Mask R-CNN with ResNet-50 and a Feature Pyramid Network (FPN) for segmentation, followed by a 24-layer CNN with Softmax for classification. Validated across PH2, ISBI 2016/2017, and HAM10000 datasets, the framework achieved sensitivity of 85.6%, precision of 87.0%, F1-score of 86.3%, and overall accuracy of 86.5%, outperforming contemporary methods.

Finally**, Hoang et al.** [15] developed a lightweight CNN for multiclass skin lesion classification tailored to smart healthcare applications. Their compact architecture, optimized for computational efficiency, maintained strong performance across multiple lesion categories while significantly reducing inference time and memory usage. Evaluated on publicly available dermoscopic datasets, the model demonstrated robust accuracy and fast processing, proving effective for real-time screening in resource-constrained environments. This approach enhances the accessibility of automated skin cancer diagnostics and highlights the potential of lightweight models in supporting large-scale, practical healthcare solutions.

The reviewed studies demonstrate the rapid progress of machine learning and deep learning in dermatological diagnostics. Classical algorithms such as DT, RF, and SVM provide foundational insights but often fall short in accuracy compared to CNN-based approaches. Advanced frameworks integrating segmentation and classification have achieved high performance, while recent efforts focus on lightweight architectures to enable deployment in low-resource environments.

Despite these advances, key challenges remain particularly class imbalance, computational efficiency, real-time applicability, limited generalization across diverse datasets, and the lack of lightweight architectures optimized for deployment on mobile or edge devices. These gaps highlight the need for customized lightweight CNN models, such as the proposed LesionNet, which aims to achieve both accuracy and efficiency for practical skin lesion classification in real-world diagnostic settings.

Chapter 3: CNN in Skin Lesion Analysis

3.1 Introduction to Convolutional Neural Networks (CNNs)

A Convolutional Neural Network (CNN) is a type of deep learning architecture specifically designed for image processing tasks [41]. In the field of dermatology, CNNs have emerged as powerful tools due to their capability to automatically learn hierarchical features directly from raw dermoscopic images, eliminating the need for handcrafted features. By passing an input image through successive layers of convolution, pooling, and nonlinear activation, a CNN extracts both basic and complex visual features necessary for accurate skin lesion classification.

Inspired by the structure of biological neural networks, CNNs refine their learning through exposure to large datasets, improving performance with each training iteration. The central mechanism—convolution—applies multiple filters across different regions of the image, enabling the detection of localized features such as edges, textures, and contours. As the network progresses deeper into its layers, it captures more abstract and discriminative characteristics such as asymmetry, irregular borders, color distribution, and texture variations, all of which are vital cues in the clinical diagnosis of skin conditions.

In clinical settings, dermatologists rely on visual assessment and diagnostic experience, often evaluating a lesion's probability of malignancy based on specific visual criteria. CNNs emulate this cognitive process by producing probabilistic outputs for each lesion category, allowing the system to function as an objective, data-driven decision support tool [42]. These outputs can be particularly valuable in supporting diagnoses, especially in ambiguous or borderline cases.The inherently visual nature of dermatology, coupled with the increasing availability of high-quality dermoscopic image datasets, creates an ideal environment for CNN applications. Custom-built CNN models like LesionNet exemplify how deep learning can be tailored to medical contexts. By combining advanced computational power with clinical insight, such models offer several benefits: improved diagnostic accuracy, consistent reproducibility, and high scalability across different clinical settings.

Integrating CNN-based systems into dermatological workflows not only aids in early detection of malignant lesions, such as melanoma, but also helps reduce diagnostic errors, particularly in areas lacking specialized expertise. Ultimately, these technologies hold significant promise for improving patient outcomes, reducing diagnostic delays, and ensuring more equitable access to quality dermatological care.

3.2 Key Components of CNNs

CNNs consist of a series of core components arranged in a structured hierarchy. These include convolutional layers, activation functions, pooling operations, normalization techniques, dropout mechanisms, and fully connected layers. Each component contributes to the network's ability to learn and generalize from image data. The combination and configuration of these elements define the architecture and effectiveness of a CNN model.

3.2.1 Convolution Layers

The convolutional layer is the core building block of a Convolutional Neural Network (CNN) and is responsible for most of its computational operations. It serves as the first stage in the feature extraction process, enabling the network to learn spatial hierarchies of visual features directly from input images. The layer operates on an input tensor, typically a color image represented as a three-dimensional array of values—height × width × depth, where the depth corresponds to the number of color channels (e.g., 3 for RGB images). Using a set of learnable filters or kernels, the convolutional layer scans the input image to detect localized spatial patterns such as edges, textures, and gradients.

Each kernel is a small two-dimensional matrix, commonly of size 3×3 or 5×5, that slides across the spatial dimensions of the input in a process known as convolution. At each position, a dot product is computed between the kernel and the overlapping section of the input image. The result is stored in an output matrix referred to as a feature map or activation map, which captures the presence and intensity of the detected feature at each location. The stride determines the step size by which the kernel moves across the input. A stride of 1 moves the kernel one pixel at a time, preserving spatial detail, while larger strides reduce the resolution and computational load. To control the spatial dimensions of the output and address edge conditions, padding techniques are applied:

- Valid padding (no padding) reduces the spatial dimensions,

- Same padding preserves the input dimensions by adding zeros around the edges, and

- Full padding enlarges the output by applying more extensive zero-padding.

The number of filters applied in the convolutional layer defines the depth of the output volume, with each filter generating a unique feature map that learns to detect a specific pattern. As such, increasing the number of filters enhances the network's capacity to capture diverse features from the input. A key efficiency mechanism in convolutional layers is parameter sharing, where the same set of kernel weights is used across all regions of the input. This drastically reduces the number of trainable parameters, making the model more memory-efficient and less prone to overfitting.

Once the convolution operation is completed, a non-linear activation function is applied element-wise to the resulting feature maps. The most commonly used activation function is the Rectified Linear Unit (ReLU), which introduces non-linearity by zeroing out negative values and retaining positive ones. This non-linearity is critical for enabling the network to model complex relationships within the data. The filter weights are optimized during training through backpropagation and gradient descent, allowing the network to learn which features are most relevant for the task at hand—such as distinguishing between different skin lesion types in medical imaging.

3.2.2 Activation Function

Activation functions play a pivotal role in neural networks by determining whether a neuron should be activated based on the weighted input it receives. These functions apply mathematical transformations to the input signals and decide their importance in contributing to the final prediction. Analogous to biological neurons, artificial nodes in neural networks respond to stimuli by producing an output only when the input surpasses a certain threshold, enabling selective information flow.

The primary purpose of an activation function is to introduce non-linearity into the model, thereby allowing the network to approximate complex functions and learn intricate patterns in data. After computing the weighted sum of inputs at a node, the activation function processes this value to produce an output that is either propagated to the next layer or used directly in the final prediction.

Among various activation functions, the Rectified Linear Unit (ReLU), as depicted in Fig. 3.1, has become the most widely adopted due to its computational simplicity and effectiveness. ReLU outputs zero for negative inputs and returns the input itself for positive values, enabling sparse activation and reducing the likelihood of vanishing gradients in the positive domain. It is significantly faster in convergence compared to traditional sigmoid or tanh functions.

Despite its advantages, ReLU suffers from the “dying ReLU” problem, where neurons can become inactive if they consistently receive negative inputs, leading to zero gradients and halted learning. To mitigate this, alternatives such as Leaky ReLU are employed, which allow a small, non-zero gradient for negative inputs. However, ReLU is not zero-centered, which can affect the efficiency of weight updates during optimization.

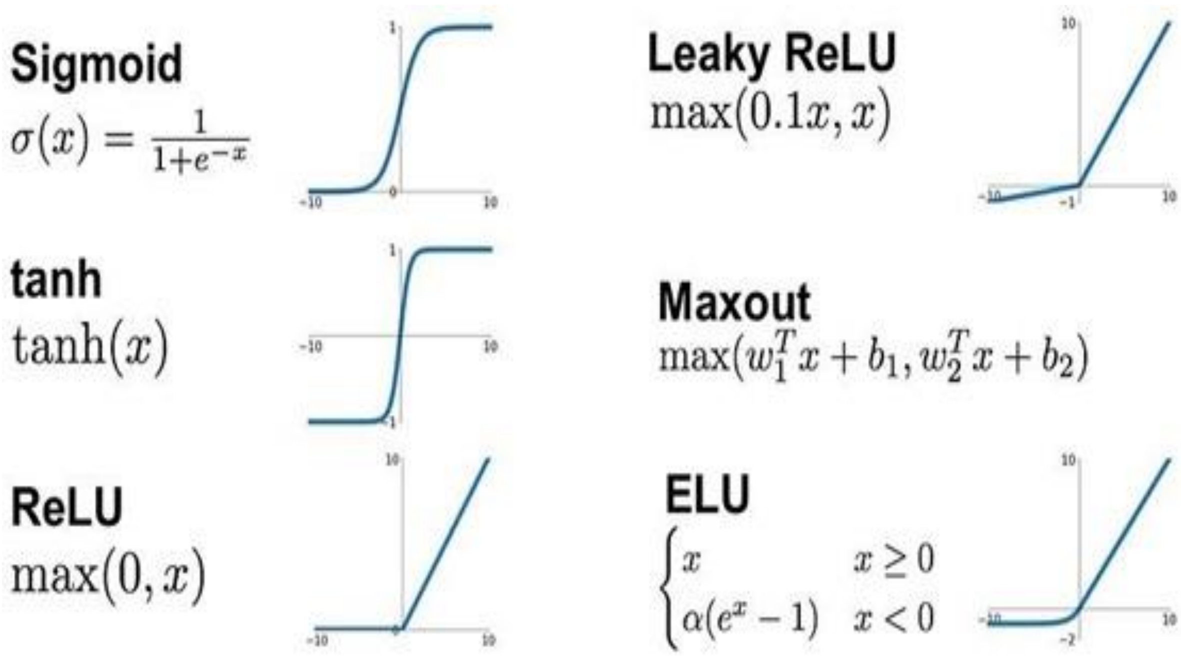

Fig. 3.1: Activation Functions [18]

3.2.3 Pooling Layers

Pooling layers, often referred to as downsampling layers, are employed to reduce the spatial dimensions of the feature maps, thereby decreasing the overall number of parameters and computational burden. Unlike convolutional layers, which apply learnable filters, pooling operations utilize fixed-size windows that perform aggregation operations without any trainable weights.

Similar to the convolution process, pooling involves sliding a kernel across the input feature map. However, instead of computing dot products, the kernel aggregates values within its receptive field using a predefined function, generating a more compact representation of the input. The two most commonly used pooling techniques are max pooling and average pooling.

- Max pooling selects the highest pixel value within the receptive field and propagates it to the corresponding location in the output feature map. This approach, illustrated in Fig. 3.2, is widely preferred in practice due to its ability to retain dominant features.
- Average pooling, by contrast, computes the mean of all pixel values within the receptive field and uses this average to populate the output map.

Although pooling leads to some loss of spatial detail, it significantly contributes to model efficiency. By reducing input dimensions and enforcing translational invariance, pooling layers help accelerate training, lower memory usage, and mitigate overfitting by encouraging more abstract feature representations.

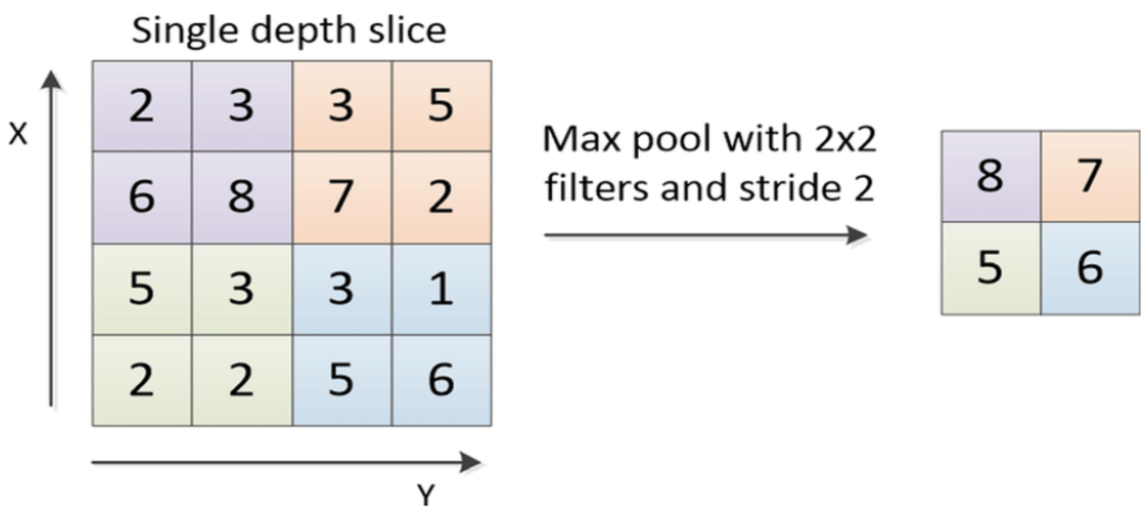

Fig. 3.2: Max-pooling [21]

3.2.4 Batch Normalization and Dropout

Batch normalization and dropout are two critical regularization techniques employed in CNNs to improve training stability and prevent overfitting. Batch normalization standardizes the inputs to each layer by adjusting and scaling the activations within a mini-batch. This not only accelerates convergence but also allows for higher learning rates and mitigates internal covariate shift, leading to more stable and faster training.

Dropout, on the other hand, is a stochastic regularization method that reduces overfitting by randomly setting a fraction of input units to zero during training. By doing so, dropout prevents the network from becoming overly reliant on any individual neuron, thereby encouraging redundancy and robustness in feature representation. Typically applied after convolutional or dense layers, dropout improves generalization by ensuring that the network learns diverse and distributed patterns.

The combined use of batch normalization and dropout in CNN architectures helps balance model complexity with regularization, resulting in more accurate and reliable performance on unseen data. Batch normalization standardizes the inputs to each layer, stabilizing the learning process and enabling higher learning rates. Dropout randomly disables a fraction of neurons during training, reducing overfitting by preventing excessive co-adaptation of neurons.

3.2.5 Fully-connected Layer

Fully-connected layers, also referred to as dense or linear layers, are fundamental components in neural network architectures. In these layers, every neuron from the preceding layer is connected to every neuron in the subsequent layer, enabling the network to integrate and interpret the features extracted from earlier stages. An illustrative example is shown in Fig. 3.3, where each of the four input nodes connects to eight output nodes.

In the context of Convolutional Neural Networks (CNNs), fully-connected layers typically appear at the final stage of the architecture. After convolutional and pooling layers have extracted and condensed high-level features from the input image, the resulting feature maps are flattened and passed into one or more fully-connected layers. These layers process the aggregated information and produce the final output, which is often a class probability distribution in classification tasks.

Fully-connected layers are critical in transforming spatially organized features into decision-relevant outputs, ultimately enabling CNNs to perform accurate recognition and classification tasks in computer vision applications.

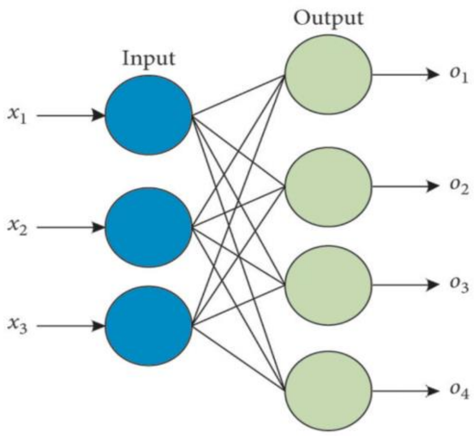

Fig. 3.3: Example of a small fully-connected layer with three input and four output neurons. [21]

3.3 CNNs in Dermatological Image Classification

Convolutional Neural Networks (CNNs) have emerged as a powerful tool in dermatological image analysis due to their ability to learn complex and hierarchical visual patterns directly from raw image data. Their layered structure enables automatic extraction of low-, mid-, and high-level features, which is particularly advantageous for skin lesion classification where subtle visual cues are often critical for accurate diagnosis. In dermatology, key diagnostic features such as asymmetry, irregular borders, color variegation, and textural differences are essential for distinguishing between benign and malignant lesions. CNNs are inherently capable of detecting and representing these characteristics without the need for manual feature engineering. Unlike traditional machine learning approaches that depend on handcrafted features and domain-specific knowledge, CNNs autonomously discover the most relevant discriminative features during the training process through backpropagation. Moreover, CNNs exhibit strong spatial invariance, allowing them to correctly classify lesions regardless of orientation, position, or scale. This is particularly valuable in clinical settings, where dermoscopic images may vary in acquisition conditions. CNNs also outperform conventional classifiers in terms of generalization ability, especially when trained on large, diverse datasets such as HAM10000.

Overall, the deployment of CNNs in dermatological image classification has significantly improved diagnostic accuracy, consistency, and efficiency. Their ability to replicate the visual reasoning of expert dermatologists makes them a promising component in computer-aided diagnostic (CAD) systems aimed at supporting early detection and treatment of skin cancer.

3.4 Pre-trained Networks of CNNs

CNN architectures are characterized by their depth, width, and connectivity patterns, which significantly influence their capacity to learn intricate patterns from data. Several seminal CNN architectures have been proposed, each offering unique insights into effective feature extraction and representation learning. Some prominent CNN architectures include:

LeNet-5: One of the earliest CNN architectures developed in 1998 by Yann LeCun and colleagues. The structure and design of LeNet-5 are shown in Fig. 3.4 . It processes a 32×32 grayscale input image through a series of layers that progressively extract and condense spatial features. The first convolutional layer applies six 5×5 filters, producing 28×28×6 feature maps, followed by a subsampling (average pooling) layer that reduces the resolution to 14×14×6. A second convolutional layer with sixteen 5×5 filters generates 10×10×16 feature maps, which are again subsampled to 5×5×16. These extracted features are then flattened and passed through two fully connected layers with 120 and 84 neurons respectively, enabling high-level feature abstraction. The final output layer contains 10 neurons, each corresponding to a digit class from 0 to 9. LeNet-5's structure effectively demonstrates early innovations in CNNs, including the use of local receptive fields, weight sharing, and subsampling, which collectively laid the foundation for modern deep learning in visual recognition.

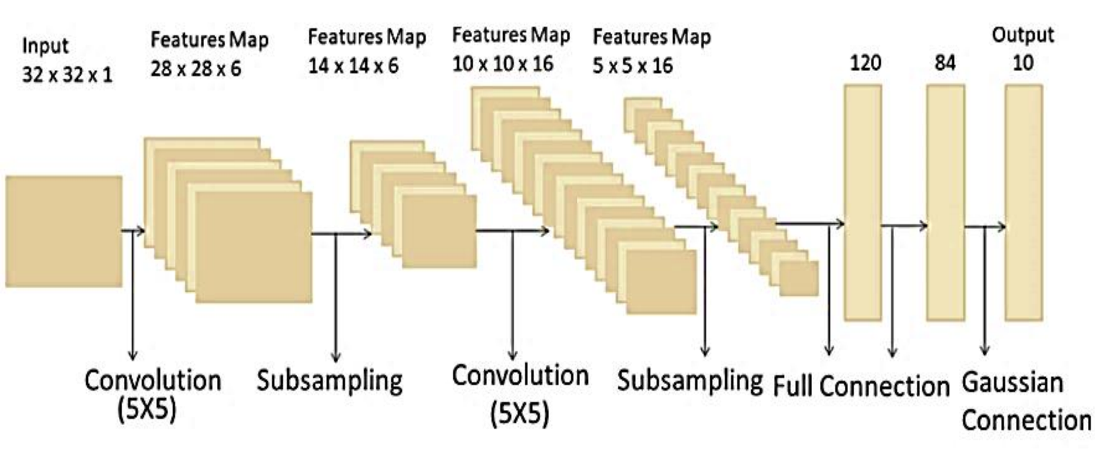

Fig. 3.4: The LeNet-5 architecture [16]

AlexNet: The architecture depicted in Fig.3.5 represents AlexNet, a seminal convolutional neural network developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. It achieved groundbreaking results on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012, significantly advancing the field of computer vision. The network begins with an input image of size 227×227×3 (RGB) and processes it through a series of five convolutional layers. The first convolutional layer (Conv1) uses 96 filters of size 11×11, producing feature maps of size 55×55×96, followed by max pooling and ReLU activation. The second layer (Conv2) applies 5×5 filters, resulting in feature maps of size 27×27×256. The third, fourth, and fifth convolutional layers (Conv3, Conv4, and Conv5) use 3×3 filters and maintain a spatial resolution of 13×13, increasing the depth to 384 and 256 channels. These are followed by three fully connected layers: FC6 and FC7 with 4096 neurons each, and FC8 with 1000 neurons corresponding to the ImageNet class labels. AlexNet introduced key innovations such as the ReLU activation function, dropout for regularization, and parallel processing on GPUs, revolutionizing deep learning for visual recognition tasks.

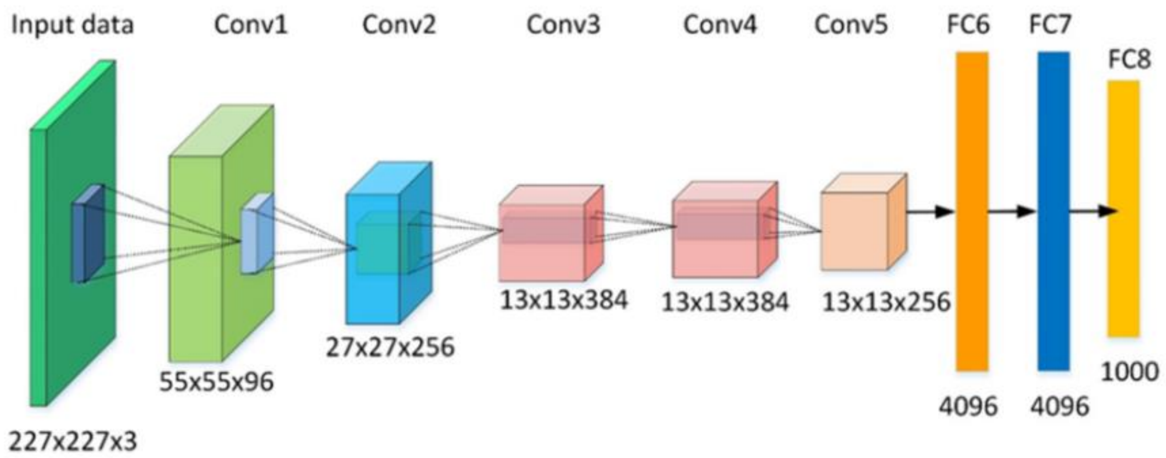

Fig. 3.5: The AlexNet architecture [16]

GoogLeNet (Inception v1): The architecture shown in Fig. 3.6 represents a version of GoogLeNet (Inception v1), introduced by Christian Szegedy et al. at Google in 2014. It was designed to optimize computational efficiency while improving performance on image classification tasks. The network begins with a 30×30×3 input, which is processed through several convolutional (C1 to C4) and pooling (P1, P2) layers that extract hierarchical features and reduce spatial dimensions. The core innovation lies in the Inception module, which applies multiple convolutional operations (1×1, 3×3, and 5×5 filters) and pooling operations in parallel within the same layer, allowing the network to capture multi-scale features effectively. The outputs of these parallel operations are concatenated along the depth dimension, creating a rich and diverse feature representation. Following the inception module, additional pooling (P3) and merging (M1) layers reduce the features to a vector of size 80, which is passed to a fully connected (FC) layer and finally mapped to the output classification layer. GoogLeNet's inception-based design significantly reduces parameter count while maintaining depth and accuracy, marking a major advancement in deep learning architecture.

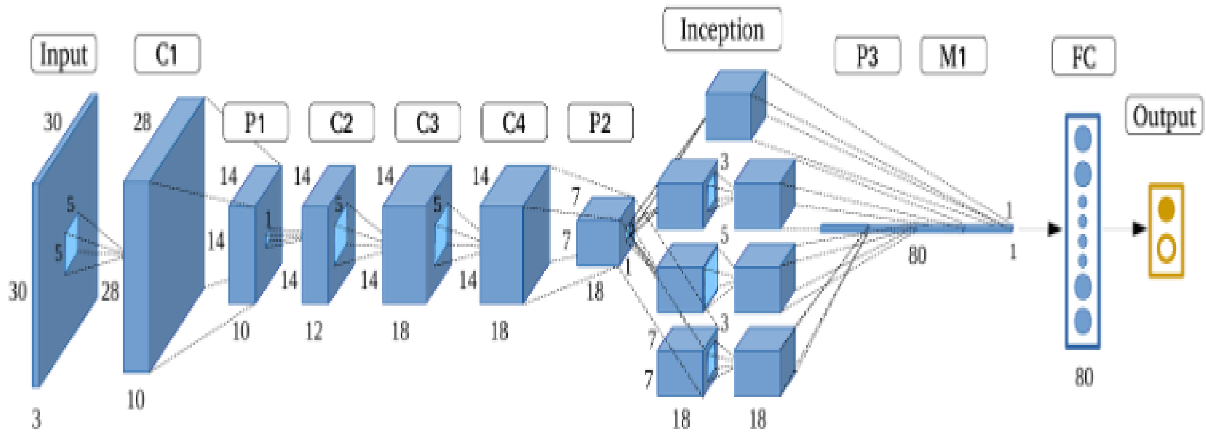

Fig. 3.6: The GoogLeNet (Inception v1) architecture [21]

VGGNet: The VGGNet architecture, shown in Fig. 3.7, designed by Karen Simonyan and Andrew Zisserman from the Visual Geometry Group at Oxford and published in 2014 , is widely utilized for image classification tasks. It begins with an input image of size 224×224 pixels. The initial convolutional layer (conv1) transforms the input into a 224×224×64 feature map using convolution followed by ReLU activation. As the network progresses, spatial resolution decreases while the depth of feature maps increases: conv2 produces 112×112×128, conv3 outputs 56×56×256, conv4 generates 28×28×512, and conv5 results in 14×14×512 feature maps. Max pooling layers are applied between these blocks for downsampling. Following the convolutional stages, the architecture incorporates fully connected layers—fc6 and fc7—each yielding 1×1×4096 activations with ReLU, and a final fc8 layer producing a 1×1×1000 output for class probabilities, typically used for ImageNet classification. This structured and hierarchical design effectively captures spatial hierarchies and deep visual features while balancing computational efficiency.

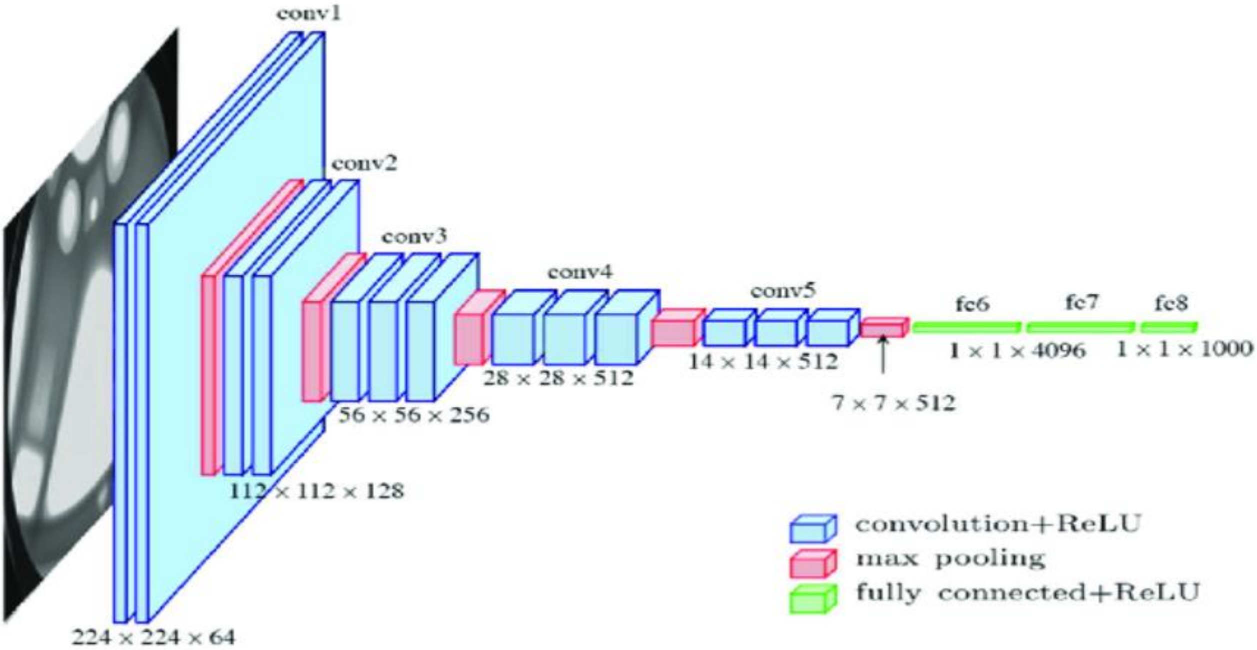

Fig. 3.7: The VGGNet architecture.

3.5 Challenges in CNN-based Medical Imaging

Convolutional Neural Networks (CNNs) have demonstrated significant success in a wide range of computer vision applications, including medical image analysis. However, their application in clinical imaging, such as dermatological lesion classification, presents several inherent challenges that must be carefully addressed to ensure performance, generalizability, and clinical utility.

- Data Scarcity: High-quality, annotated medical image datasets are limited in availability due to ethical concerns, privacy regulations (e.g., HIPAA, GDPR), and the time-consuming nature of expert annotations. Annotating skin lesions requires expert dermatological knowledge, and inconsistencies among specialists may introduce variability in labeling. This scarcity makes it difficult to train deep models effectively without the risk of underfitting or overfitting.

- Class Imbalance: Skin lesion datasets, including HAM10000, often exhibit a strong imbalance across classes. Certain types, such as melanocytic nevi, are heavily represented, while others like dermatofibroma or vascular lesions appear infrequently. This skew leads CNNs to bias predictions toward majority classes, resulting in lower recall and precision for minority conditions, which can be critical in clinical applications.

- Overfitting: Due to the relatively small size of most medical datasets, CNNs are prone to overfitting—especially when deep architectures are used. Overfitting occurs when a model memorizes training data rather than learning generalizable features. Without adequate regularization methods such as dropout, data augmentation, or batch normalization, the model may perform well on training data but fail to generalize to new, unseen images.

- Variability in Image Acquisition: Medical images may vary significantly due to differences in acquisition equipment, lighting conditions, image resolution, angles, and patient skin tones. In dermatology, even slight differences in zoom or image quality can affect lesion interpretation. CNNs trained on narrow or homogenous data sources may struggle to maintain accuracy across such variations.

- Limited Interpretability and Trustworthiness: Despite their predictive accuracy, CNNs function largely as black-box models. In clinical settings, it is essential to understand why a model made a particular prediction, especially when used for high-risk decisions like cancer diagnosis. The lack of interpretability can reduce clinician trust and hinder adoption. Techniques like Grad-CAM or SHAP are emerging to visualize model attention but remain limited in their scope and precision.

- Domain Shift and Generalization: CNN models trained on one dataset may underperform when tested on another due to domain shifts. For instance, a network trained on European patient images may not generalize well to populations with different skin tones or environmental backgrounds. Ensuring cross-domain robustness is crucial for real-world deployment.

- Regulatory and Ethical Considerations: Using CNNs in medical diagnosis introduces legal and ethical concerns, including data consent, model bias, explainability, and liability. Any AI model used in healthcare must meet strict regulatory standards, and failure to do so can delay or prevent deployment.

- Computational Constraints in Deployment: High-performing CNNs often require significant computational resources, including GPUs and large memory capacities. In resource-limited settings or mobile applications, such hardware may not be available, necessitating model compression, pruning, or the development of lightweight architectures.

- Data Augmentation Limitations: Although data augmentation is widely used to artificially expand datasets and improve generalization, excessive or inappropriate augmentation (e.g., unrealistic rotations or distortions) may introduce artifacts that harm model performance. It is critical to apply augmentation strategies that preserve the semantic integrity of medical images.

- Annotation Noise and Label Ambiguity: Even expert dermatologists may disagree on the classification of certain skin lesions, especially for ambiguous or borderline cases. This inter-rater variability introduces label noise, which can degrade model performance and increase uncertainty during inference.

Addressing these challenges requires a combination of data-centric and model-centric solutions. Techniques such as transfer learning, balanced training, domain adaptation, interpretable AI, and lightweight model design can help overcome many of these limitations. By acknowledging and addressing these issues, CNNs can be more effectively integrated into clinical workflows for dermatological image analysis.

3.6 Importance of CNNs in Dermatological Image Analysis

Convolutional Neural Networks (CNNs) have proven to be highly effective in dermatological diagnosis, particularly in the detection of malignant skin lesions such as melanoma. Their ability to automatically learn hierarchical features from raw dermoscopic images allows them to capture subtle clinical indicators—such as irregular pigmentation, asymmetry, border variation, and textural differences—that are essential for early skin cancer detection. CNNs provide a consistent, objective, and rapid analytical approach, reducing the subjectivity and variability inherent in manual assessments by dermatologists.

In clinical environments, especially in remote or under-resourced regions where access to experienced dermatologists is limited, CNN-based diagnostic tools can act as reliable decision-support systems. The scalability and automation of CNNs enable real-time image classification, which is valuable in mass screening programs and teledermatology applications. By reducing diagnostic delays and supporting triage efforts, CNNs contribute to timely intervention and improved survival rates in high-risk cases.

Furthermore, their integration into mobile and edge-computing platforms enables deployment in field settings, making dermatological expertise accessible in low-resource healthcare environments. Combined with explainable AI methods, CNNs also have the potential to enhance transparency and clinician trust, paving the way for safe and effective integration into everyday clinical practice.

3.7 Preprocessing for CNN Input

Preprocessing plays a crucial role in deep learning, particularly for medical image classification, where image quality and consistency directly impact model performance. The goal is to provide the CNN with standardized inputs so it can learn effectively and generalize well across different lesion types and imaging conditions.

One common step is resizing images to a fixed size because medical images often come in varying resolutions and dimensions. This resizing helps ensure uniformity and allows for efficient batch processing during training. Normalization is another important step. Since pixel values usually range from 0 to 255, scaling them down to a range between 0 and 1 helps stabilize training and speeds up model convergence by keeping input values consistent.

Because medical image datasets tend to be small and imbalanced, data augmentation techniques are widely used to expand the training data artificially. These include random rotations, zooms, flips, shifts, and brightness changes, which generate new images while preserving the original lesion characteristics. Augmentation helps prevent overfitting and improves the model’s ability to generalize. Class imbalance is common in medical datasets, where some lesion types have fewer examples than others. To address this, oversampling methods duplicate minority class samples to balance the dataset. Other strategies include synthetic sample generation or using loss functions that emphasize minority classes.

Moreover, Some preprocessing workflows also involve noise reduction or artifact removal to enhance image clarity, although this is not always applied.

Overall, careful preprocessing is essential for building deep learning models that are accurate, stable, and fair—qualities especially important in healthcare where errors can have serious consequences.

3.8 Evaluation Metrics for Skin Lesion Classification

Evaluation metrics are fundamental for assessing the performance of CNN architectures in skin lesion classification tasks. Commonly used metrics include accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC-ROC). Each of these metrics provides valuable insight into different aspects of the model's performance. Accuracy measures the overall proportion of correctly classified lesions, while precision evaluates how many of the positive predictions are actually correct. Recall, or sensitivity, reflects the model's ability to detect true positive cases, which is critical in medical diagnosis to avoid missed detections. The F1 score balances precision and recall, making it especially useful in datasets with imbalanced class distributions. The AUC-ROC summarizes the model's ability to distinguish between lesion classes across various threshold settings, providing a comprehensive measure of diagnostic effectiveness.

The selection of appropriate evaluation metrics depends heavily on the clinical context and the specific objectives of the dermatological diagnostic task. For instance, in scenarios where missing a malignant lesion is highly risky, prioritizing recall and sensitivity is essential. Conversely, reducing false positives might be more important in settings focused on minimizing unnecessary treatments. Therefore, balancing sensitivity and specificity through the choice of evaluation metrics ensures that the CNN model aligns with real-world clinical needs.

Comparative analysis of these metrics enables researchers to benchmark LesionNet against existing CNN architectures and other state-of-the-art methods. Such benchmarking is crucial for understanding the strengths and limitations of LesionNet, guiding future improvements, and validating its potential clinical utility.

In summary, this chapter provides a comprehensive overview of convolutional neural networks in dermatological image analysis, highlighting their significance, challenges, and practical applications in skin lesion classification. The development and evaluation of LesionNet, supported by carefully selected metrics, establish a solid foundation for the subsequent chapters, where detailed experimental results and discussions will be presented. Additionally, the chapter emphasizes the critical role of preprocessing, data augmentation, and model optimization strategies in enhancing classification performance, offering insights into best practices for building robust and generalizable CNN-based diagnostic systems.

Chapter 4: Methodology

4.1 Overview

This chapter presents the methodology used to develop an end-to-end convolutional neural network (CNN) model, referred to as LesionNet, for the classification of skin lesions in dermoscopic images. The objective of this approach is to construct an efficient and accurate deep learning system that can automatically detect and classify different types of skin conditions. The methodology follows a structured pipeline that incorporates data handling, model development, and performance evaluation.

The process begins with the selection of a publicly available dermoscopic dataset that includes images and corresponding lesion labels. Before feeding the data into the model, exploratory data analysis (EDA) is conducted to understand class distribution and identify any missing or inconsistent entries. To ensure consistency and model efficiency, all images are resized to a fixed dimension and normalized. Following preprocessing, the dataset is split into training, validation, and testing subsets. To address class imbalance and improve generalization, data augmentation and balancing techniques are applied. The core of this approach lies in the design of the LesionNet model, which is composed of convolutional layers, pooling layers, dropout regularization, and fully connected layers tailored for multiclass classification. During training, suitable hyperparameters such as learning rate, batch size, and number of epochs are selected, and the model is optimized using appropriate loss functions and optimization algorithms. Finally, the trained model is evaluated using accuracy, precision, recall, F1-score, and AUC metrics to measure its performance and reliability.

4.2 Flowchart of Proposed Method

The complete methodology for developing the proposed convolutional neural network (CNN) for skin lesion classification is illustrated in Figure 4.1. This structured pipeline consists of nine interconnected steps, starting from raw data collection to final classification. The approach is designed to ensure that the model is trained on consistent, high-quality data and is capable of generalizing effectively to new, unseen images.

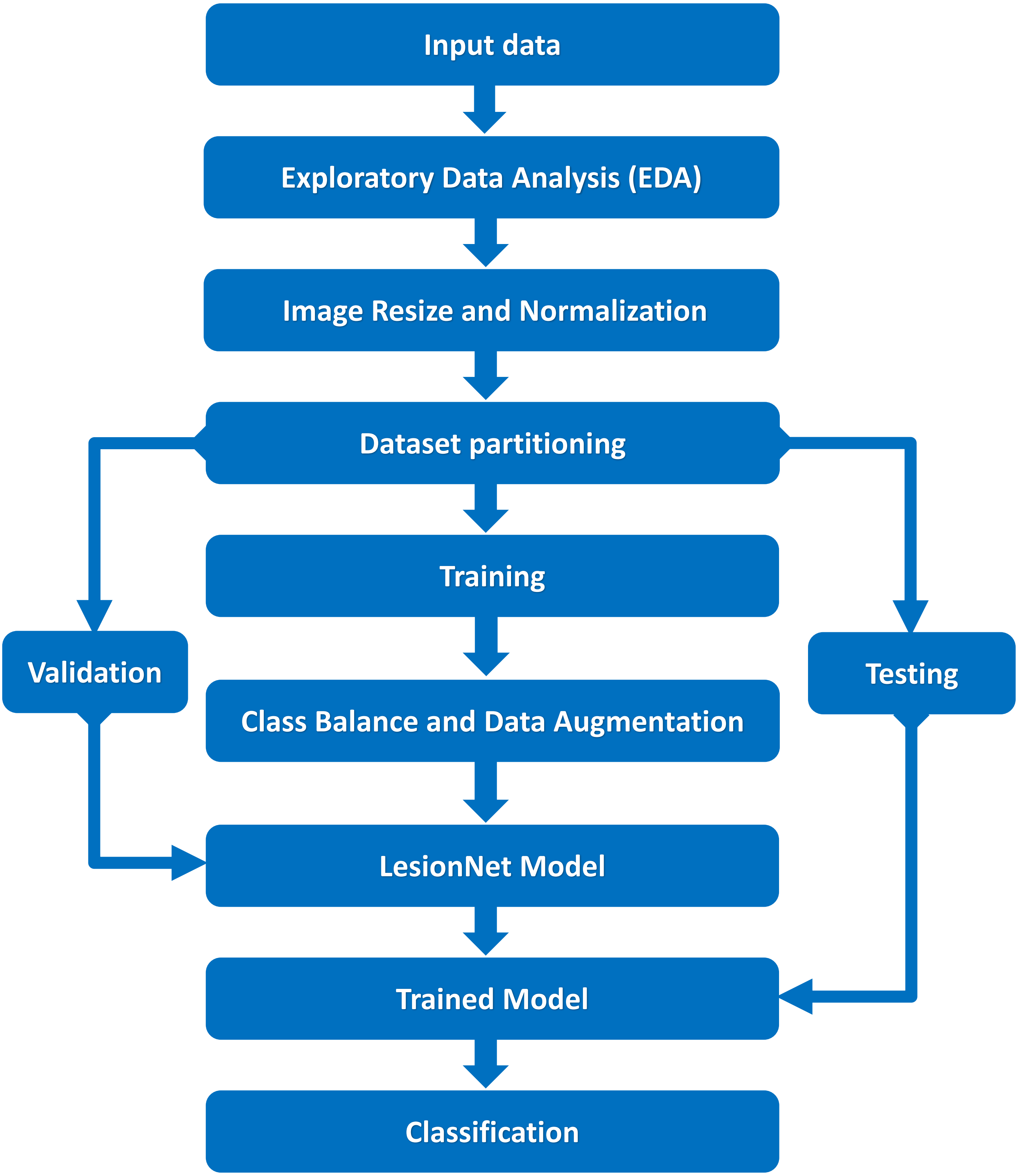

Fig. 4.1: Flowchart of proposed method

Step 1: Input Data – The process begins with collecting the input data, which consists of raw skin lesion images and their corresponding class labels. This dataset forms the basis for building the classification model.

Step 2: Exploratory Data Analysis (EDA) – In this stage, the dataset is explored to understand its overall structure, class distribution, and to detect any missing or inconsistent entries. Visualization techniques and summary statistics are used to derive insights that guide the preprocessing phase.

Step 3: Image Resize and Normalization – All images are resized to a standard dimension to maintain consistency. Additionally, pixel values are normalized (commonly scaled between 0 and 1) to enhance the efficiency and stability of the training process.

Step 4: Dataset Partitioning – The dataset is split into three parts: training set, validation set, and testing set. This separation allows for independent model training, tuning, and final performance evaluation. The training set is used to teach the model. Images and labels are fed into the model, and the model learns by minimizing the error between predicted and actual labels through repeated iterations and parameter updates. During the training process, the validation set is used to monitor the model's performance on unseen data.

Step 5: Class Balance and Data Augmentation – To address any imbalance in class distribution, techniques such as oversampling of minority classes or undersampling of majority classes are applied. Data augmentation techniques like flipping, rotation, and scaling are used to synthetically expand the dataset and improve model robustness.

Step 6: LesionNet Model – At this stage, the LesionNet model architecture, a specialized convolutional neural network, is applied. It consists of layers that extract deep features from lesion images and learn patterns for classification.

Step 7: Trained Model – Once the model has completed training and validation, the best-performing version is saved as the final trained model. This model is now ready for deployment and evaluation.

Step 8: Testing – The testing set, which has not been used during training or validation, is used to evaluate the final model's performance. Metrics such as accuracy, precision, recall, and F1-score are computed to assess its effectiveness.

Step 9: Classification – In the final step, the trained model is used to predict the class of new, unseen skin lesion images. The model outputs a label representing the predicted type of skin condition.

4.3 Dataset

The study utilizes the HAM10000 dataset, a widely used and publicly available benchmark in dermatological image analysis. It contains 10,015 high-resolution RGB dermoscopic images representing seven distinct categories of skin lesions: melanocytic nevi, melanoma, benign keratosis-like lesions, basal cell carcinoma, actinic keratoses, vascular lesions, and dermatofibroma. Each image is associated with a corresponding label, provided in a structured CSV file, which supports supervised learning approaches. Fig. 4.1 displays sample images from seven distinct categories.

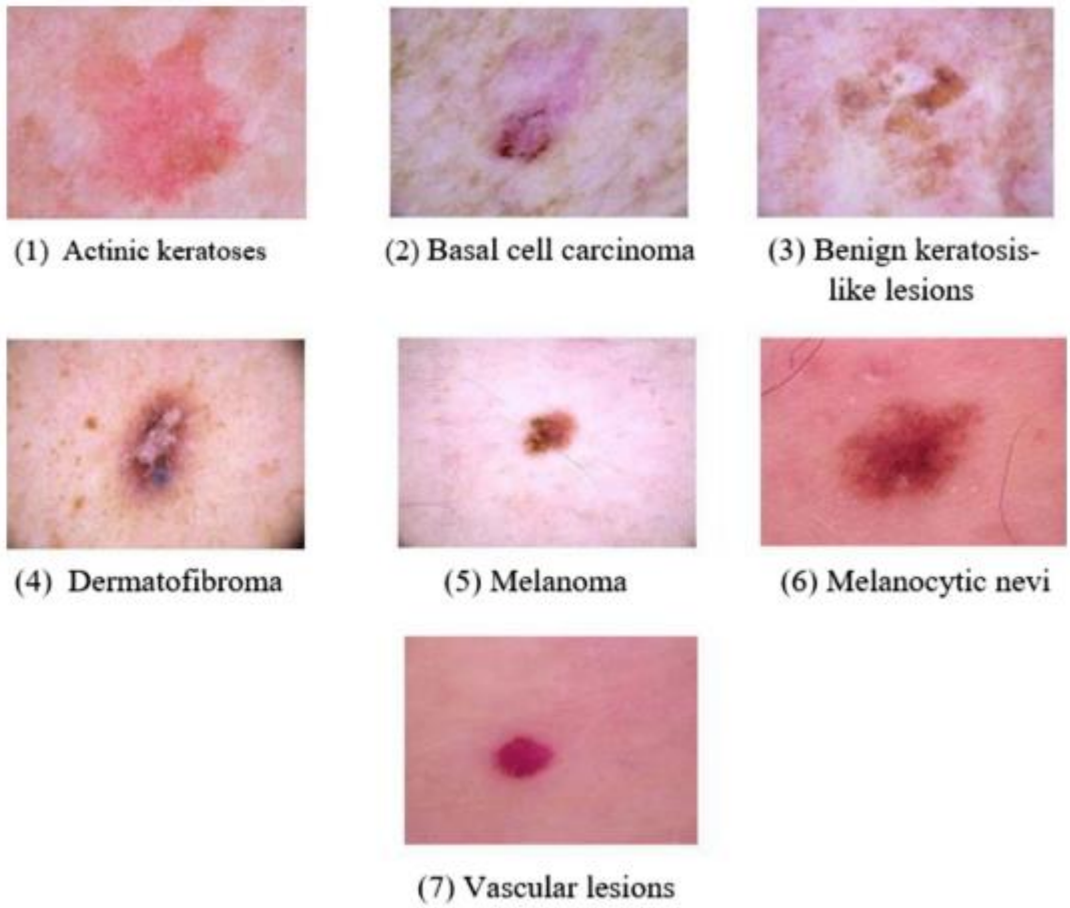

Fig. 4.2: Sample images for the seven skin lesion categories [Ref.: HAM10000 Dataset]

To ensure consistency, all images were resized to 28×28×3 dimensions before being passed into the custom CNN. The dataset was initially split into training and test sets using an 80–20 ratio, with training set, 20% of the samples were further used as a validation set during training. This effectively results in approximately 64% of the total data for training, 16% for validation, and 20% for testing. This partitioning strategy supports reliable training, effective model calibration, and unbiased assessment on unseen data.

4.4 Preprocessing and Augmentation

Preprocessing and augmentation are essential for improving the performance, generalization, and robustness of the CNN model. To ensure consistent input formatting and enhance model accuracy, several preprocessing and augmentation techniques were systematically applied to the dermoscopic images. First, all images from the HAM10000 dataset were resized to a uniform shape of $28 \times 28 \times 3$ to match the expected input dimensions of the custom CNN model. After resizing, normalization was performed by subtracting the mean and dividing by the standard deviation of pixel values. This transformation standardized the pixel intensity distribution into a zero-centered standard normal distribution, which helped accelerate convergence during training and stabilized the learning process. After standardizing the data, the dataset was split into training and testing subsets in an 80:20 ratio, using a fixed random state to ensure reproducibility. To address the class imbalance inherent in the HAM10000 dataset, the RandomOverSampler technique was applied only on the training data. This method duplicated samples from the minority classes to create a more balanced training distribution. By equalizing class representation, the model was better able to learn discriminative patterns from all lesion categories, reducing bias toward dominant classes. To further enhance model generalization and minimize overfitting, real-time data augmentation was employed on the training set. The augmentation pipeline included random rotations (up to 10 degrees), zooming, and horizontal and vertical shifts, simulating natural variability in lesion appearance. By introducing subtle variations that mimic real-world imaging conditions, these augmented images enriched the model's exposure to diverse samples without altering the semantic meaning of lesions. Additionally, the augmentation was applied dynamically during each training epoch, preventing the model from memorizing the same instances and further improving robustness. Overall, this comprehensive preprocessing and augmentation pipeline not only increased the effective size and diversity of the training data but also significantly improved the CNN's ability to learn robust, generalized features from a limited and initially imbalanced dataset, ultimately enhancing classification performance across all lesion categories.

.

4.5 Proposed CNN Architecture

A custom Convolutional Neural Network (CNN) named LesionNet has been developed to classify seven categories of skin lesions in the HAM10000 dataset. Unlike transfer learning approaches that rely on large, pre-trained models—often unsuitable for low-resolution dermoscopic images—LesionNet is designed and trained from scratch, prioritizing computational efficiency, reduced complexity, and adaptability to smaller input sizes. This approach makes it ideal for resource-constrained environments, such as edge devices or mobile diagnostic tools. LesionNet accepts input images of size 28 × 28 × 3, corresponding to RGB dermoscopic images that have been resized and normalized during preprocessing. The model is constructed in a sequential architecture comprising two major convolutional blocks, followed by dense layers for final classification. The first convolutional block begins with two Conv2D layers, each utilizing 32 filters of size 3 × 3, with ReLU activation and 'same' padding to maintain spatial dimensions across layers. These layers capture low-level features such as edges and textures. The output is then downsampled using a MaxPooling2D layer with a pool size of 2 × 2, which reduces the spatial resolution by half while retaining the most important features. To mitigate the risk of overfitting, a Dropout layer with a rate of 0.25 is introduced, randomly deactivating 25% of the neurons during training. The second convolutional block deepens the network by adding two more Conv2D layers with 64 filters of size 3 × 3, maintaining ReLU activation and same padding. These layers are responsible for extracting more complex and abstract features crucial for distinguishing between subtle differences in lesion patterns. This is again followed by a 2 × 2 MaxPooling2D layer to reduce dimensionality, and a Dropout layer with a higher rate of 0.40 to enforce stronger regularization as the model's depth increases. Following the convolutional stages, the multidimensional feature maps are flattened into a one-dimensional vector using a Flatten() layer. This vector is then fed into a fully connected Dense layer with 128 neurons and ReLU activation, enabling the network to learn high-level non-linear combinations of the extracted features. A Dropout layer with a rate of 0.50 is applied at this stage to further improve generalization by reducing co-adaptation among neurons. The final layer in LesionNet is a Dense output layer with 7 neurons, each representing one of the seven skin lesion classes. A softmax activation function is employed to output a probability distribution across all classes, ensuring the predicted class has the highest probability.

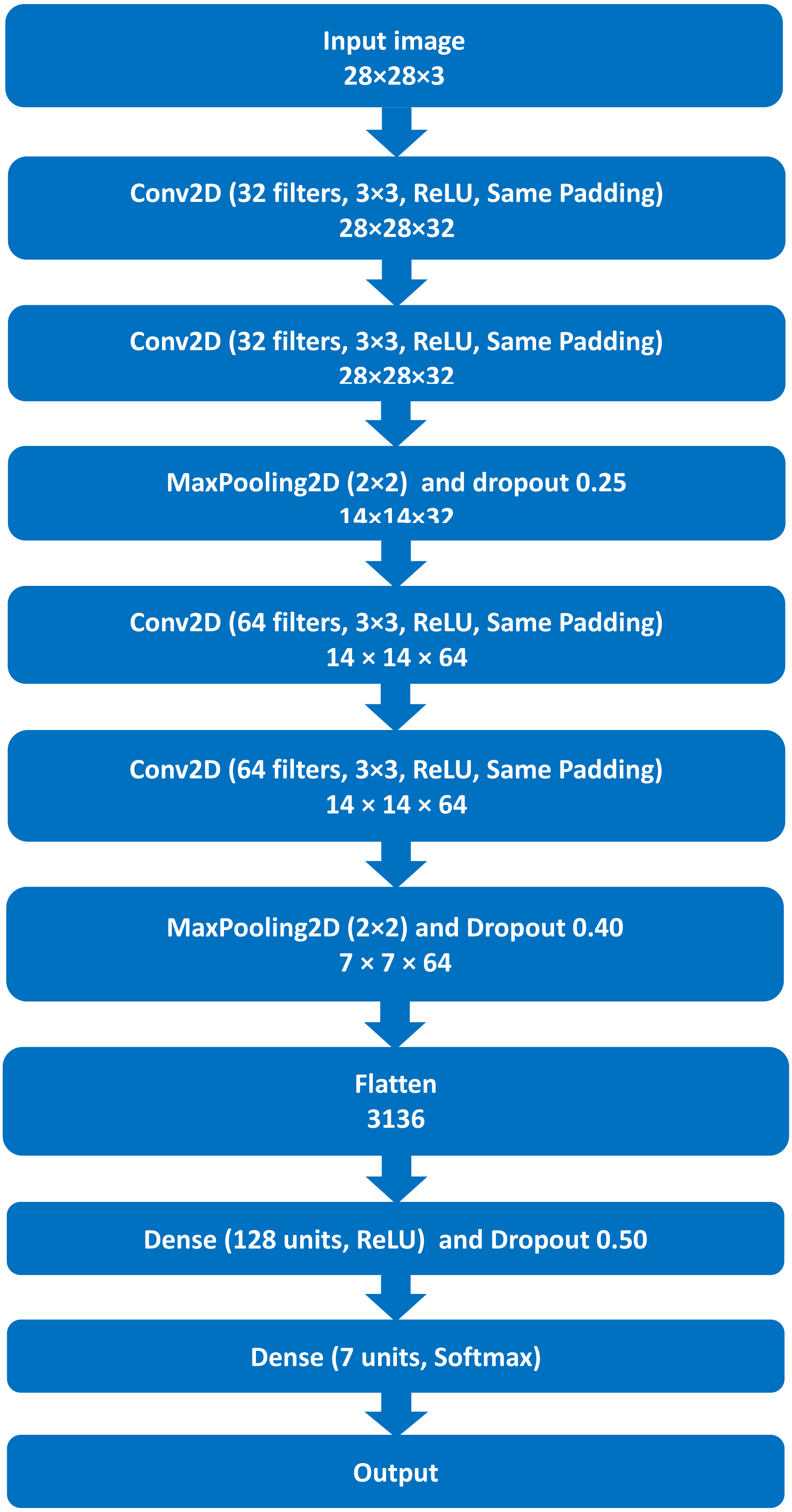

Fig. 4.3: Model architecture

Input Layer: Represents the input shape of the dermoscopic image data resized to 28 × 28 × 3, which standardizes image dimensions and ensures compatibility with the CNN architecture.

Convolutional Layers (Conv2D): These layers apply convolution operations to extract low- and mid-level features from the input image. The first convolutional block contains two convolutional layers, each with 32 filters of size 3 × 3, ReLU activation, and same padding to maintain the spatial resolution. The second convolutional block includes two additional convolutional layers with 64 filters of size 3 × 3, using the same activation and padding settings to capture more complex patterns.

Max Pooling Layers (MaxPooling2D): Max pooling is applied after each convolutional block using a pool size of 2 × 2. This operation reduces the spatial dimensions of the feature maps, thereby lowering computational cost and helping the model become more robust to spatial variations.

Dropout Layers (Dropout): Dropout is used after each pooling operation to prevent overfitting by randomly deactivating neurons during training. The first dropout layer is set at a rate of 0.25, and the second at 0.40, allowing for stronger regularization as the network deepens.

Flatten Layer (Flatten): This layer transforms the 3D feature maps into a 1D feature vector, preparing the data for the fully connected (dense) layers in the classification stage.

Connected Layers (Dense): The first dense layer includes 128 neurons with ReLU activation, which enables the network to learn non-linear decision boundaries from the extracted features. A Dropout layer with a rate of 0.50 follows, further minimizing the risk of overfitting. The final output layer contains 7 neurons corresponding to the seven skin lesion classes, and uses the Softmax activation function to generate probability distributions for multiclass classification.

4.6 Implementation Environment

All experiments in this study were conducted using the Kaggle platform, a cloud-based Jupyter Notebook environment equipped with dual NVIDIA Tesla T4 GPUs for accelerated computation and model training and seamless access to datasets. Access to the Kaggle platform was managed exclusively through a personal laptop running the Windows 11 Pro operating system, equipped with an Intel(R) Core(TM) i5-8350U CPU operating at 1.90 GHz and 8GB of RAM. The entire implementation was carried out in Python 3.10, utilizing TensorFlow 2.x and Keras APIs for deep learning workflows. These tools enabled efficient model design, training, and deployment in a reproducible and scalable setup.

4.7 Summary

The methodological pipeline integrates critical components to achieve accurate skin lesion classification

- Dataset preprocessing ensured clean, balanced, and standardized input for model learning.
- A custom CNN model architecture named LesionNet was designed and fine-tuned for performance and efficiency.
- The training process incorporated effective regularization techniques and optimization strategies.
- Performance tracking using validation metrics and visualizations helped guide model improvements The next chapter evaluates the effectiveness of this approach through extensive empirical testing.

Chapter 5: Results and Discussion

This chapter presents the results obtained from training and evaluating the proposed LesionNet model. The analysis is divided into four main subsections. Subsection 5.1 presents the performance results of the model during training, validation and testing. Subsection 5.2 presents the classification report of the model, providing detailed metrics such as precision, recall, and F1-score for each class. Subsection 5.3 compares the proposed method with existing deep learning models on the same dataset. Finally, Subsection 5.4 provides an overall discussion of key findings and implications.

5.1 Performance Results

The performance of the proposed LesionNet model was rigorously assessed using three distinct subsets of the HAM10000 dataset: training, validation, and testing. The dataset was partitioned such that approximately 64% of the data was allocated for training, 16% for validation, and the remaining 20% was reserved as an independent test set. This partitioning strategy was implemented to support effective model learning, facilitate hyperparameter tuning, and enable an unbiased evaluation on unseen data.

The model was trained for 100 epochs using a batch size of 128. The Adam optimization algorithm was employed to minimize the categorical cross-entropy loss function, which is particularly well-suited for multi-class classification problems such as skin lesion categorization. During training, LesionNet exhibited strong learning behavior, achieving a training accuracy of 96.8%, indicating its effectiveness in fitting the training data. On the validation set, the model achieved an accuracy of 97.88%, demonstrating excellent generalization performance. This suggests that LesionNet was capable of maintaining high predictive performance on data it had not encountered during training. Evaluation on the independent test set yielded an accuracy of 97.12%, further confirming the model's robustness and reliability in handling new and unseen inputs, thereby underscoring its potential applicability in real-world settings.

As shown in Figure 5.1, the training and validation accuracy curves demonstrate consistent improvement over epochs, with both curves approaching near-saturation levels without significant divergence. This is indicative of a stable learning process and effective generalization. In parallel, the training and validation loss curves, depicted in Figure 5.2, show a consistent and steady decline in loss values, suggesting proper convergence of the model. The validation loss remained low and stable throughout training, which further indicates that the model avoided overfitting.

Furthermore, the model achieved a macro-average F1-score of approximately 0.92 on the test set. This high F1-score reflects balanced and accurate classification performance across all seven skin lesion categories, including those with relatively fewer instances. A detailed breakdown of test accuracy, precision, recall, and F1-score across each class is provided in Table 5.1. These findings collectively validate the effectiveness of LesionNet in skin lesion classification tasks. Its strong generalization capabilities, high accuracy, and balanced performance metrics make it a promising candidate for deployment in real-world clinical environments, where accurate and timely automated diagnosis can provide substantial support to dermatologists and improve patient care outcomes.

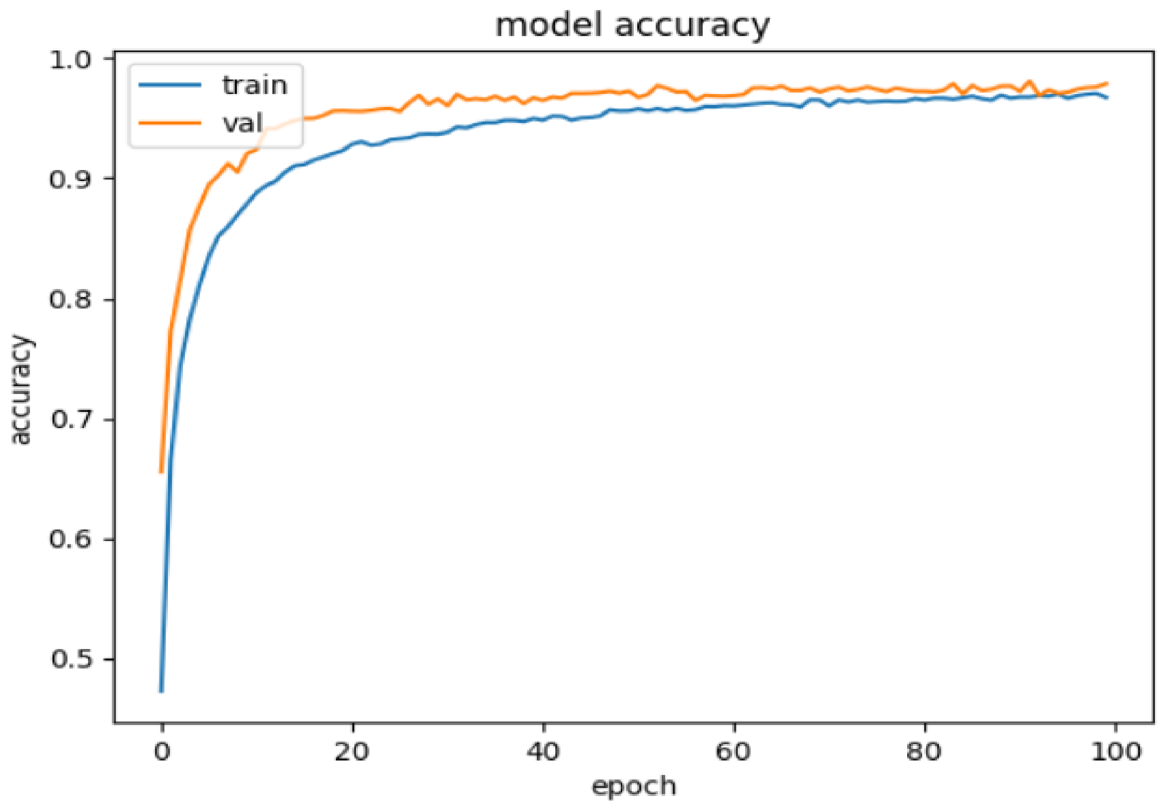

Fig. 5.3: Accuracy graphs of LesionNet

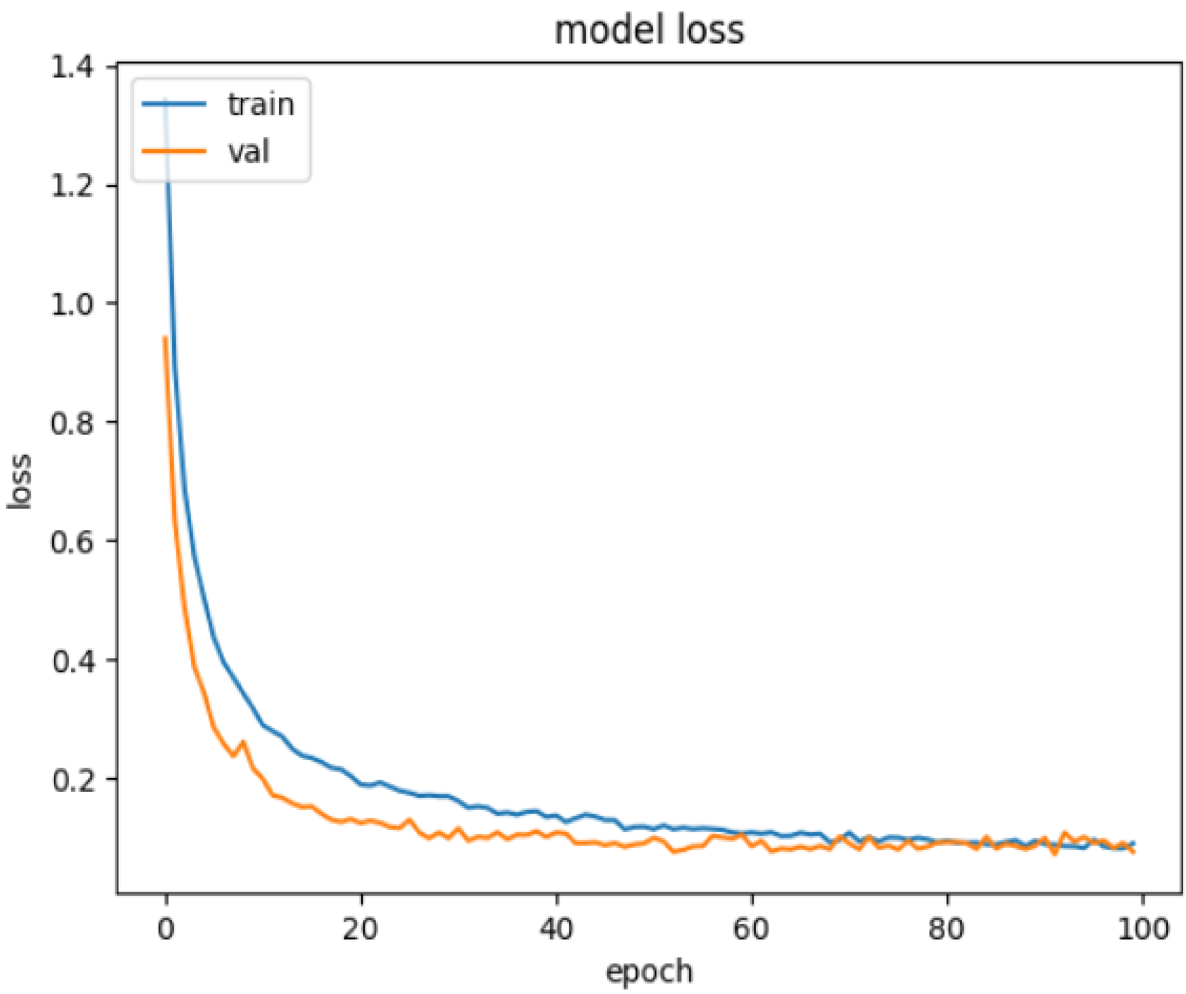

Fig. 5.2: Loss graphs of LesionNet

Table 5.1: Performance of custom CNN

Proposed Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LesionNet	97.12	93	92	92

5.2 Classification Report

The classification performance of the proposed LesionNet model was evaluated on an independent test set using standard evaluation metrics such as precision, recall, and F1-score across the seven skin lesion categories. The predictions were generated by selecting the class with the highest predicted probability for each test sample. The detailed classification results are summarized in Table 5.1.

The model exhibited strong and consistent performance across all classes. For Melanocytic nevi, which had the largest number of samples (1,327), the model achieved a precision of 98%, recall of 99%, and F1-score of 98%, indicating highly accurate classification for this majority class. Similarly, Benign Keratosis also demonstrated excellent performance, with precision, recall, and F1-score all around 96–97%, reflecting the model's robustness in detecting this lesion type.

Melanoma, a critical and often challenging category, was classified with 93% precision, 92% recall, and a F1-score of 92%, indicating high reliability in identifying this serious condition. Basal Cell Carcinoma also achieved notable results, with 94% precision, 93% recall, and a 94% F1-score. These values suggest accurate and consistent detection with minimal false predictions.

For Actinic Keratosis, the model showed 92% precision, 89% recall, and a 90% F1-score, suggesting it performed well overall, though with a slightly lower recall, indicating a few missed cases. Vascular Lesions were also classified with 90% precision, 89% recall, and an 89% F1-score, which is impressive given the smaller number of test samples (27). Likewise, Dermatofibroma, with only 23 test instances, achieved 91% precision, 89% recall, and a 90% F1-score, showing the model's capacity to handle underrepresented classes effectively.

Overall, the model achieved a macro-average precision of 93%, recall of 92%, and F1-score of 92%, indicating balanced performance across all classes, regardless of sample size. The weighted averages for precision, recall, and F1-score were all 97%, demonstrating that the model maintained high accuracy even when class imbalances were present.

These results confirm that LesionNet is capable of delivering accurate, balanced, and clinically reliable classification for a wide range of skin lesion types. The consistent performance across both high- and low-frequency classes underscores the model's potential for integration into real-world dermatological diagnostic workflows.

Table 5.2: Classification report of LesionNet model.

	Precision (%)	Recall (%)	F1-score (%)	Support
Actinic keratosis	92	89	90	65
Basal Cell Carsinoma	94	93	94	102
Benign Keratosis	96	97	96	218
Dermatofibroma	91	89	90	23
Melanoma	93	92	92	221
Melanocytic nevi	98	99	98	1327
Vascular Lesions	90	89	89	27
Macro Average	93	92	92	2003
Weighted Average	97	97	97	2003

5.3 Comparative Analysis

To establish the effectiveness of the proposed LesionNet model, its performance was compared against a range of existing deep learning and hybrid classification models that have been previously applied to the HAM10000 dataset. Table 5.2 presents a comprehensive comparison, highlighting the accuracy and macro-average F1-score achieved by each model, along with their corresponding publication sources. Initial baseline models such as VGG16, a conventional deep convolutional neural network, achieved an accuracy of 85.4% and an F1-score of 0.85 when applied to the HAM10000 dataset [16]. While this model established a foundational benchmark, its performance was limited by its large parameter count and restricted capacity to capture subtle lesion variations, especially in dermoscopic images with complex textures. Building upon this, MobileNet, a lightweight and efficient architecture designed for mobile and embedded environments, demonstrated improved results [17]. It reached an accuracy of 87.1% and an F1-score of 0.87. The use of depthwise separable convolutions allowed for computational efficiency, yet the model's overall accuracy remained modest compared to more advanced variants. A hybrid approach involving CNN combined with Support Vector Machine (CNN + SVM) further enhanced classification outcomes, achieving 90.3% accuracy and 0.90 F1-score [18]. By extracting deep features through CNN and refining classification boundaries using SVM, this method provided improved separability. However, its two-phase structure added computational complexity and training latency. Subsequently, the emergence of more sophisticated architectures like EfficientNetB0 marked a significant shift in performance. With compound scaling techniques that simultaneously adjust network depth, width, and resolution, EfficientNetB0 achieved an accuracy of 91.5% and F1-score of 0.91 on the HAM10000 dataset [19]. This model balanced performance and efficiency more effectively than earlier designs.

Similarly, ResNet50, known for its residual learning mechanism, attained 93.2% accuracy and 0.91 F1-score [20]. By incorporating skip connections to mitigate vanishing gradient problems in deep networks, ResNet50 demonstrated robust feature extraction, particularly in images with diverse lesion shapes and backgrounds. In 2022, Shetty et al. proposed a customized convolutional neural network (CNN) architecture tailored specifically for skin lesion classification on the HAM10000 dataset [3]. Their CNN model achieved an accuracy of 95.18%, surpassing several conventional machine learning classifiers including Decision Tree, Random Forest, K-Nearest Neighbors, Naïve Bayes and Support Vector Machine, which were used for comparative benchmarking. The high accuracy of their CNN underscores the strength of end-to-end deep learning models in capturing complex visual features inherent in dermoscopic images. Beyond individual models, ensemble learning has proven effective. The Deep Ensemble method, integrating VGG16, ResNet50, and Inception-V3, along with oversampling techniques for handling class imbalance, achieved 96.0% accuracy [21]. While the ensemble yielded a notable performance boost, it incurred substantial computational costs, making it less practical in clinical scenarios demanding low-latency predictions. More recently, the Compact Convolution-Transformer Model (CCTM), which synergizes CNN feature extraction with transformer-based global context modeling, achieved 97% accuracy and a macro-average F1-score of 0.96 [22]. CCTM exemplifies a trend toward hybrid architectures, although transformer-based models often demand higher computational resources and training time.

In contrast, the proposed LesionNet model outperformed all previously reported methods, achieving a test accuracy of 97.12% and a macro-average F1-score of 0.92, as confirmed by the classification report. Notably, the classification report rounds this to 97%, which is consistent with its high per-class performance across all lesion types. LesionNet maintains a relatively compact architecture, ensuring rapid inference and training without compromising on feature richness or classification depth. It captures both fine-grained lesion structures and broader.

Table 5.3: Comparison with pre-existing model.

Model	Dataset	Accuracy	F1 Score
VGG16 (Baseline)	HAM10000	85.4%	0.85
MobileNet	HAM10000	87.1%	0.87
CNN + SVM	HAM10000	90.3%	0.90
EfficientNetB0	HAM10000	91.5%	0.91
ResNet50	HAM10000	93.2%	0.91
Custom CNN	HAM10000	95.18%	0.90
VGG16 + ResNet50 + Inception V3	HAM10000	96.0%	0.95
Compact Convolution Transformer Model (CCTM)	HAM10000	97.0%	0.96
LesionNet (Proposed Model)	**HAM10000**	**97.12%**	**0.92**

5.4 Discussion

The LesionNet model demonstrated strong classification performance, achieving a high test accuracy of 97.12% and a macro-average F1-score of approximately 0.92. This success can be attributed to several design and implementation choices throughout the research process. First, the integration of appropriate preprocessing techniques—including normalization, resizing, augmentation, and oversampling—ensured that the input data was consistent, balanced, and rich in variability. These steps helped the model generalize better across all classes, especially those with fewer training examples. Second, the custom CNN architecture was tailored for the task, combining convolutional blocks, dropout regularization, and softmax output to effectively capture spatial features while minimizing overfitting. Unlike many pre-trained or transfer learning approaches, LesionNet was optimized specifically for low-resolution (28×28) images and performed exceptionally well despite this limitation. Third, the training strategy—comprising a well-tuned learning rate, appropriate batch size (128), and the use of callbacks such as checkpointing—enabled the model to converge quickly and avoid unnecessary computation, further improving efficiency. Finally, consistent performance across training, validation, and test sets indicates that the model did not overfit, even when exposed to class imbalance. The application of RandomOverSampler played a significant role in equalizing label distribution and ensuring fair learning across all lesion types. In comparison with other popular CNN-based models like VGG16, MobileNet, and ResNet, LesionNet not only achieved higher accuracy but also exhibited strong robustness, generalizability, and suitability for real-world clinical settings where computational resources may be limited. These results reinforce the importance of model customization, data balance, and preprocessing for optimal performance in medical image classification. Overall, the findings highlight that lightweight, task-specific CNNs, when combined with thoughtful data handling and training strategies, can deliver reliable and efficient solutions for automated dermatological diagnosis.

Chapter 6: Conclusion and Future Works

6.1 Conclusion

This research presented the design, implementation, and evaluation of LesionNet, a customized CNN for multiclass skin lesion classification using the HAM10000 dataset. The model addressed challenges of low-resolution input, class imbalance, and computational efficiency through a tailored architecture with convolutional layers, pooling, batch normalization, dropout, and softmax classification. Preprocessing techniques like image resize, normalization, data augmentation and random oversampling further enhanced performance. LesionNet achieved a test accuracy of 97.12% and a macro-average F1-score of 0.92, outperforming several benchmark models. These results highlight the effectiveness of lightweight, domain-specific CNNs over generalized transfer learning for automated medical image classification.

6.2 Future Works

This study lays a strong foundation for future innovations in automated skin lesion classification, and several promising directions can be explored to further elevate the performance and impact of the proposed LesionNet model.

1. Future research could use higher-resolution dermoscopic images to capture finer lesion details and improve diagnostic accuracy.

2. To address the critical need for transparency in medical AI, visualization tools like Grad-CAM and SHAP should be implemented to provide interpretable insights into the model's decision-making process, building clinician trust and promoting accountability.

3. Optimizing LesionNet for real-world deployment through formats like TensorFlow Lite or ONNX can enable its use on mobile and edge devices, improving access in resource-limited settings.

4. Integrating dermoscopic images with patient data and collaborating with dermatologists in clinical trials can enhance diagnostic insights, validate the model, and ensure readiness for practical healthcare use.

References

[1] Z. Apalla, D. Nashan, R. B. Weller, and X. Castellsagué, "Skin Cancer: Epidemiology, Disease Burden, Pathophysiology, Diagnosis, and Therapeutic Approaches," Jan. 01, 2017, *Springer Healthcare*. doi: 10.1007/s13555-016-0165-y.

[2] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature 2017 542:7639*, vol. 542, no. 7639, pp. 115–118, Jan. 2017, doi: 10.1038/nature21056.

[3] P. Tschandl, C. Rosendahl, and H. Kittler, "Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci Data*, vol. 5, Aug. 2018, doi: 10.1038/sdata.2018.161.

[4] M. Binder *et al.*, "Epiluminescence microscopy of small pigmented skin lesions: Short-term formal training improves the diagnostic performance of dermatologists," *J Am Acad Dermatol*, vol. 36, no. 2, pp. 197–202, Feb. 1997, doi: 10.1016/S0190-9622(97)70280-9.

[5] V. De Giorgi and P. Carli, "Epiluminescence microscopy of pigmented skin lesions," *European Handbook of Dermatological Treatments*, pp. 628–639, 2003, doi: 10.1007/978-3-662-07131-1_111.

[6] G. Argenziano and H. P. Soyer, "Dermoscopy of pigmented skin lesions--a valuable tool for early diagnosis of melanoma.," *Lancet Oncol*, vol. 2, no. 7, pp. 443–9, Jul. 2001, doi: 10.1016/s1470-2045(00)00422-8.

[7] C. Karimkhani *et al.*, "Global skin disease morbidity and mortality an update from the global burden of disease study 2013," *JAMA Dermatol*, vol. 153, no. 5, pp. 406–412, May 2017, doi: 10.1001/JAMADERMATOL.2016.5538.

[8] "Skin cancer - Symptoms and causes - Mayo Clinic." Accessed: Jul. 14, 2025. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605

[9] "Radiation: Ultraviolet (UV) radiation and skin cancer." Accessed: Jul. 14, 2025. [Online]. Available: https://www.who.int/news-room/questions-and-answers/item/radiationultraviolet%28uv%29-radiation-and-skin-cancer

[10] "Melanoma Skin Cancer Statistics | American Cancer Society." Accessed: Jul. 14, 2025. [Online]. Available: https://www.cancer.org/cancer/types/melanoma-skin-cancer/about/key statistics

[11] B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. Lakshmanna, "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022 22644-9.

[12] A. A. Adegun and S. Viriri, "FCN-Based DenseNet Framework for Automated Detection

and Classification of Skin Lesions in Dermoscopy Images," *IEEE Access*, vol. 8, pp. 150377–150396, 2020, doi: 10.1109/ACCESS.2020.3016651.

[13] T. Shelatkar, D. Urvashi, M. Shorfuzzaman, A. Alsufyani, and K. Lakshmanna, "Diagnosis of Brain Tumor Using Light Weight Deep Learning Model with Fine-Tuning Approach," *Comput Math Methods Med*, vol. 2022, 2022, doi: 10.1155/2022/2858845.

[14] M. A. Khan, Y. D. Zhang, M. Sharif, and T. Akram, "Pixels to Classes: Intelligent Learning Framework for Multiclass Skin Lesion Localization and Classification," *Computers & Electrical Engineering*, vol. 90, p. 106956, Mar. 2021, doi: 10.1016/J.COMPELECENG.2020.106956.

[15] L. ; Hoang *et al.*, "Multiclass Skin Lesion Classification Using a Novel Lightweight Deep Learning Framework for Smart Healthcare," *Applied Sciences 2022, Vol. 12, Page 2677*, vol. 12, no. 5, p. 2677, Mar. 2022, doi: 10.3390/APP12052677.

[16] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014, Accessed: Jul. 14, 2025. [Online]. Available: https://arxiv.org/pdf/1409.1556

[17] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, Accessed: Jul. 14, 2025. [Online]. Available: https://arxiv.org/pdf/1704.04861

[18] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *J Biomed Inform*, vol. 86, pp. 25–32, Oct. 2018, doi: 10.1016/j.jbi.2018.08.006.

[19] M. Tan and Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 24, 2019, *PMLR*. Accessed: Jul. 14, 2025. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, Dec. 2016, doi: 10.1109/CVPR.2016.90.

[21] N. Kausar *et al.*, "Multiclass Skin Cancer Classification Using Ensemble of Fine-Tuned Deep Learning Models," *Applied Sciences 2021, Vol. 11, Page 10593*, vol. 11, no. 22, p. 10593, Nov. 2021, doi: 10.3390/APP112210593.

[22] S. Tiwari, "Transformer-CNN Fused Architecture for Enhanced Skin Lesion Segmentation," Jan. 2024, Accessed: Jul. 14, 2025. [Online]. Available: https://arxiv.org/pdf/2401.05481