

هالة قادر آغا

Skin Lesion Thesis - ResNet18 %5EM LIME %28FINAL 4 chapters%29.DOC6.docx

 Midocean University

Document Details

Submission ID

trn:oid:::2945:406444321

Submission Date

Aug 8, 2026, 6:15 PM GMT+3

Download Date

Aug 9, 2026, 4:04 AM GMT+3

File Name

Skin Lesion Thesis - ResNet18 ^M LIME (FINAL 4 chapters).DOC6.docx

File Size

9.2 MB

23 Pages

5,842 Words

34,058 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely human generated as AI generated and likely AI generated as human generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.



Title
Explainable Multimodal Skin Lesion
Classification Using ResNet-18 and TabNet with
LIME

Name: Alanoud Alnjrany **ID No:** 261001412

Name: Fatimah Hasan Alghamdi **ID No:** 222000322

Name: Halah Qader Agha **ID No:** 253001696

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

College of Informatics
Midocean University

Supervisor Name: D\Sherif Mostafa.....

Collage of Informatics, Midocean University, Comoros

Email.....

Explainable Multimodal Skin Lesion Classification Using ResNet-18 and TabNet with LIME

Acknowledgements

First, we would like to express our sincere thanks and appreciation to our supervisor, Dr. Sherif Mostafa, for all the guidance, assistance, and patience he provided. His scientific guidance and constructive criticism were indispensable to the progress of our research and enabled us to complete this project. We are also grateful to the faculty and staff of the department for providing a safe and encouraging environment, the necessary resources, and access to dermatological image repositories. We also acknowledge our gratitude to the International Skin Imaging Collaboration (ISIC), whose free images allowed us to compile the dataset used in this study.

On a personal level, we express our deep gratitude to our families for their unwavering support and encouragement during the long hours dedicated to this research. Their trust in us was a fundamental factor in making this achievement possible. Finally, we extend our thanks and appreciation to everyone who contributed in any way to the completion of this project.

Abstract

Skin Cancer is still the most widespread cancer worldwide, and the rate of cancer research is still increasing monthly. Both melanin and non-melanin skin cancer can be fatal if not detected early, so early detection can greatly affect the patient. Due to the difficulty in interpreting images of surgical diseases, there is no scientific evaluation of the same lesion, and early stages often overlap between cases, and diagnosis depends partly on observational experience.. Deep learning reduces that gap. Convolutional neural networks such as ResNet-18 already reach classification performance on skin images that, in some contexts, nears that of trained clinicians. Their main drawback is a lack of interpretability: they deliver a prediction without explaining the rationale, which limits clinical acceptability. This thesis bridges these two needs. A ResNet-18 model, initialised with ImageNet weights, is fine-tuned via transfer learning on a curated, multi-source collection of 878 dermatological images

across ten skin-condition labels (nine used in the working splits(. The dataset includes malicious, benign, and infectious organisms. Because it is small in size, this dataset uses techniques to increase and weight categories to enhance generalization and protect underrepresented categories, while the program highlights the image regions upon which each decision was based. The objective is apparent but difficult to achieve: clear, high-precision results with clear and concise explanations of the model.

Contents

List of Abbreviations	6
List of Figures.....	7
List of Tables	8
Chapter 1: Introduction.....	9
Overview.....	9
Problem Definition	9
Thesis Objectives.....	10
Thesis Contribution	10
Thesis Organization	10
Chapter 2: Background and Related Work.....	11
Skin Cancer and Dermoscopy	11
Machine Learning and Deep Learning for Image Classification.....	11
Residual Networks and ResNet-18.....	11
Transfer Learning	12
The Class-Imbalance Problem.....	12
Explainable Artificial Intelligence (XAI).....	12
Related Work	14
Summary and Research Gap.....	16
Chapter 3: Dataset Selection.....	17
Choice of Dataset.....	17
Source and Composition.....	17
Characteristics.....	17
Class Distribution and the Imbalance Challenge.....	17
Accessibility	18
Chapter 4: Proposed Framework	19
Overview of the Framework.....	19
Methodology Overview:.....	19
Framework Flow:.....	20
Data pre-processing	20
Data Augmentation and Class Balancing	20
Classification Model: Pre-trained ResNet-18.....	21
Training Setup	21
Explainability with LIME.....	21
Evaluation Strategy.....	22
Tools and Technologies	22
References.....	23

List of Abbreviations

Abbreviation	Meaning
AD	Atopic Dermatitis
AI	Artificial Intelligence
AK	Actinic Keratosis
ANN	Artificial Neural Network
AUC	Area Under the ROC Curve
BKL	Benign Keratosis
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
DF	Dermatofibroma
DL	Deep Learning
FN	False Negative
FP	False Positive
Grad-CAM	Gradient-weighted Class Activation Mapping
ISIC	International Skin Imaging Collaboration
LIME	Local Interpretable Model-Agnostic Explanations
MEL	Melanoma
ML	Machine Learning
NV	Melanocytic Nevus
ReLU	Rectified Linear Unit
ResNet	Residual Network
ROC	Receiver Operating Characteristic
SCC	Squamous Cell Carcinoma
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine
TL	Transfer Learning
TRC	Tinea / Ringworm / Candidiasis
VASC	Vascular Lesion
XAI	Explainable Artificial Intelligence

List of Figures

Figure 1: Class distribution of the curated skin-condition dataset.....	17
Figure 2: Proposed framework (ResNet-18 + LIME).....	18

List of Tables

Table 1: Summary and comparison of related studies	14
Table 2: Class distribution of the curated skin-condition dataset	17
Table 3: Main hyper-parameters of the proposed configuration	20

Chapter 1: Introduction

Overview

Skin cancer is one of the most commonly diagnosed cancers worldwide and has risen in incidence over the last few decades. It occurs when skin cells multiply in uncontrolled numbers after long-term exposure to sunlight or tanning devices, often from exposure to ultraviolet radiation from sunlight or sunburns. Skin cancer is divided into melanoma and non-melanoma groups. Melanoma is less common but accounts for most skin cancer deaths as it is most likely to metastasize. The good thing is that early diagnosis of skin cancer is usually highly treatable, so very early detection can significantly improve survival. Hence, accurate early diagnosis is essential. (4, 19, 21)

Clinical assessment usually starts with a visual inspection of the lesion using a handheld magnifier, a tool that enables the examiner to see the underlying structures which are invisible to the naked eye. Dermoscopy helps with better diagnostic detail, but its images are subjective as well. Judgement is also dependent on training and experience, and studies repeatedly show that even experienced dermatology experts may disagree about the same case. Early-stage lesions can look alike and make diagnosis more difficult. The workload of suspect lesions is increasing and dermatology professionals are scarce in most areas outside large metropolitan areas. These pressures have driven interest in computer-aided diagnosis (CAD) tools. The intention is not to replace dermatologists; rather to provide a fast and reliable second opinion that is available to the general public at large.

In the last decade, deep learning, specifically convolutional neural networks (CNNs), has changed what CAD systems can do. ResNet, VGG, and EfficientNet have been seen to perform well on benchmark datasets comparable to board-certified dermatologists. ResNet-18 is a compact residual network that can be trained quickly, adapt to mid-sized datasets with transfer learning, and should be a realistic choice for a project of this scope. (4, 19, 21)

But predictive performance alone is not enough. Many deep networks function as opaque mechanisms: they provide a confident label without providing a reason why it is so. In medical settings, that opacity is problematic. A clinician who gets an unexpected malignancy prediction has little basis for trusting it, and regulators are not inclined to accept systems that are not justified. Explainable Artificial Intelligence (XAI) aims to fill this gap. Methods like LIME provide visual maps of image regions that helped with the prediction, converting a label into a human-reviewable argument. This thesis is about that union: a fine-tuned ResNet-18 for classification and LIME for providing a visual representation of each decision. (10, 14, 15)

Problem Definition

There are two intertwined issues. First, skin-condition classification is hard; malignant and benign lesions can be different only slightly, inflammatory and infectious diseases can be similar and many dermatology collections are heavily biased towards common classes and not rare ones. A classifier trained on such biased data tends to be biased towards majority classes, leaving rare but clinically important patient classes underrepresented. In order to correct this, we employ a curated and roughly balanced dataset but the trade-off is that we get fewer images per class overall and we need to use transfer learning and augmentation.

Second, even a very accurate model has little clinical value if it cannot explain its predictions. Clinicians need to know the basis for a diagnosis, especially when cancer is a possibility. There is a tension between raw performance and interpretability: top-performing CNNs are often opaque,

and more transparent models often sacrifice accuracy. The challenge is therefore to build a classifier that can handle all ten curated categories (and the rare ones), and provide per-prediction explanations that clinicians can assess.

Thesis Objectives

With this in mind, this thesis aims at the following goals:

1. Develop a multi-class classifier for the ten skin-condition categories in the curated dataset using a pre-trained ResNet-18 and transfer learning.
2. Address the small dataset size and mild class imbalance by data augmentation, transfer learning, and class weighting so that under-represented classes are not overlooked.
3. Integrate LIME so every prediction is accompanied by a visual explanation of image regions that impacted the decision.
4. Use predictive analytics to evaluate the model with accuracy, precision, recall, F1-score, and AUC to assess whether the LIME results match clinical reasoning.
5. Show that acceptable accuracy and interpretability can be achieved in such a way dermatologists can use it.

Thesis Contribution

In this work, we aim to achieve:

1. An end-to-end explainable skin-condition classification pipeline that combines a fine-tuned ResNet-18 with LIME on a curated 10-category dermatological dataset.
2. A data efficiency and balance approach to augmenting and weighting the data to produce robust results for each class on a small dataset, while preserving the overall accuracy.
3. A qualitative assessment of LIME explanations to verify that the model attends to the lesion rather than irrelevant background skin or imaging artefacts (such as hair, rulers, ink).
4. Evidence and literature-supported that a lightweight architecture like ResNet-18 is suitable for the task and is suitable for resource-constrained hardware.

Thesis Organization

The rest of this paper follows a simple structure. Chapter 2 provides the background and the recent works that are relevant to this study and also compares methods and results. Chapter 3 presents the skin condition dataset in detail, its origin, composition and the issues that make it challenging. Chapter 4 outlines the proposed approach step by step, from preparing images and handling imbalance to fine-tuning ResNet-18 with LIME and evaluation. After the model has been implemented and tested, we will discuss the results and conclusions.

Chapter 2: Background and Related Work

Background and Related Work. This chapter delves deeper into the technical and scientific foundations that underpin the project, ranging from the medical aspects of skin lesion diagnosis to deep learning and explainable AI techniques relevant to the proposed framework. The latter half reviews recent studies closely related to the problem, summarising their contributions and contrasting their approaches to identify the gap this thesis intends to fill.

Skin Cancer and Dermoscopy

Most skin cancer is classified into melanoma and non-melanoma in the clinic and skin cancer types. Melanoma is caused by melanocytes (pigment-producing cells) and is less common, but the majority of deaths are due to its aggressive metastatic behaviour. Non-melanoma cancers are mainly basal cell carcinoma and squamous cell carcinoma, which are more common but less aggressive in nature. In addition to malignant tumors, there are benign lesions - common moles, benign keratoses, dermatofibromas and vascular marks that are benign and may look like cancerous tumors.

Dermoscopy is the primary method for studying pigmented lesions. With magnification and controlled illumination, often with immersion fluid or polarisation to reduce glare, a dermoscope shows colours and structural patterns which are invisible to the unaided eye: pigment networks, dots and globules, streaks, vascular arrangements. Clinicians often distil these features into heuristics like the ABCD rules - asymmetry, border, colour, differential structures - although they can be described, grading is difficult and so the task is suited for pattern-learning methods like deep learning.

Machine Learning and Deep Learning for Image Classification

The previous image-analysis systems were based on hand-crafted features, where researchers wrote code to find measurements of colour, texture, and shape and fed the descriptors into classifiers such as support vector machines or random forests. This approach, however, was limited by the quality of the features built, and designing good descriptors for medical images is time-consuming. Deep learning removed this bottleneck by learning hierarchical features directly from pixels. Convolutional Neural Networks (CNNs) are the backbone of image deep learning: stacked convolutional layers detect local patterns, pooling layers reduce spatial dimensions and summarise activations, and non-linearities like ReLU are used to model complex relationships. The neural networks learn basic edge and color detectors of images and then the deeper layers build more complex textures and lesion structures, and the fully connected layers map these structures to class probabilities. Architectures from Alex-Net and VGG to the residual networks have established CNNs as the standard image classification technology for medicine and beyond.

Residual Networks and ResNet-18

As networks deepened in search of higher accuracy, researchers came across the degradation problem: beyond a certain depth, adding layers degraded performance due to loss of gradients in backpropagation. Residual Networks (Res-Nets) incorporated identity shortcut connections, so each layer learns a residual (a tiny correction) with respect to its input, rather than a new

transformation. These skip connections preserve gradient flow and enable very deep networks to be trained successfully. ResNet-18 is the compact model of the family with 18 weight layers and around eleven million parameters. It is efficient and relatively inexpensive in memory and less prone to overfitting on mid-sized datasets, which makes it popular in medical imaging and is why we chose it for our study. (6)

Transfer Learning

Training deep networks from scratch demands large volumes of labelled data which are rarely available in medical areas. Transfer learning addresses this problem by starting from a model pre-trained on a large, general dataset like ImageNet. The early layers of such a model already capture generic visual primitives - edges, textures, shapes - and are useful for many tasks. The final classification layer is replaced and the network is fine-tuned on the new medical data, typically using a low learning rate to adapt general features to the target domain without destroying them. Transfer learning is fundamental to improving accuracy and reducing training time for skin lesion tasks and is the heart of our method. (3, 8)

The Class-Imbalance Problem

Real-world thermoscopic collections tend to be skewed because common lesion types are photographed far more frequently than rare ones. To see a network with thousands of examples for one class and only a few hundred from another biases the model towards the majority class, which is problematic when rare classes are clinically significant. A few alternatives include augmentation to increase minority classes (by rotation, flipping, scaling and color jitter), re-sampling (oversampling minorities or under sampling majorities), and cost-sensitive losses (weighted cross-entropy makes mistakes on rare classes more costly; focal loss focuses on harder-to-classify examples). The dataset used here is largely balanced and so we can remove a few residual skew in our model. (5, 9, 20)

Explainable Artificial Intelligence (XAI)

Explainable AI is a set of approaches that explains the reasoning behind a complex model. For image classifiers, post-hoc methods are popular: they interpret the trained model without altering its structures. Grad-CAM traces gradients into the last convolution layer to produce a coarse heatmap of attended regions. SHAP uses game-theoretic methods to model the contribution of feature groups. LIME, the approach we are using in our work, explains a single prediction by perturbing the input: it divides the image into super-pixels, turns subsets of them on and off, observes any changes in the output of the model and fits a simple, interpretable model to that local behavior. The result is a ranked map of super-pixels and their influence on the prediction. LIME is model-agnostic with a natural fit to perceptual image patches and its outputs are easy to use and interpret by clinicians. Its strong local fidelity in medical imaging is also very strong. (10, 14, 15)

AI-Driven Skin Lesion Analysis: A 5-Stage Classification Pipeline

This pipeline details the systematic progression of skin lesion analysis using Artificial Intelligence, moving from initial raw image preparation through deep feature extraction and specialized tabular learning to reach a final classification and an explainable model output.

Stage 1:

Method
Image Augmentation & Pre-processing

Purpose
Optimizing image quality and consistency



Input: Raw skin lesion images

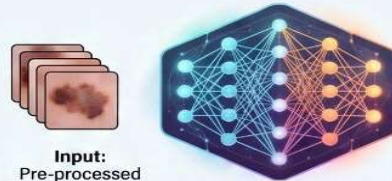


Output: Standardized (resized/normalized) images

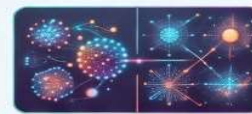
Stage 2:

Method
Deep Feature Extraction

Purpose
Capturing intricate visual patterns and textures



Input: Pre-processed image data

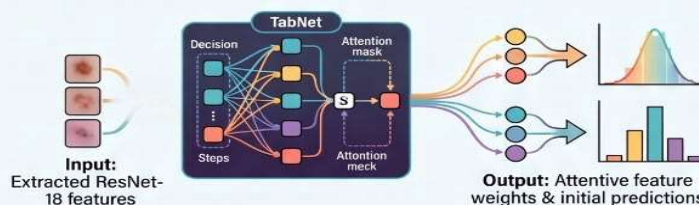


Output: High-dimensional feature vectors

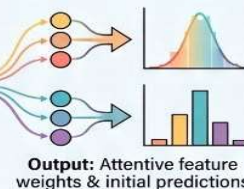
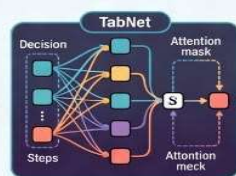
Stage 3:

Method
Tabular Learning via TabNet

Purpose
Filtering redundant data and prioritizing discriminative features



Input: Extracted ResNet-18 features

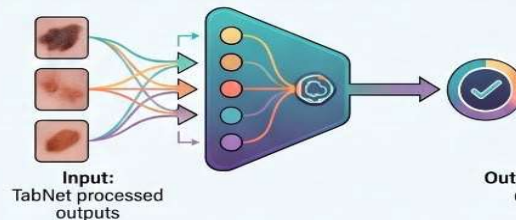


Output: Attentive feature weights & initial predictions

Stage 4:

Method
Final Classification

Purpose
Mapping optimized features to diagnostic categories



Input: TabNet processed outputs

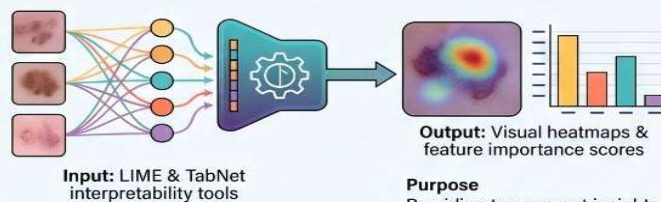
Specific skin disease labels

Output: Specific skin disease labels

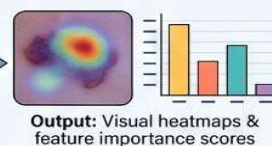
Stage 5:

Method
Model Explainability (XAI)

Purpose
Providing transparent insights into decision-making



Input: LIME & TabNet interpretability tools



Output: Visual heatmaps & feature importance scores

Related Work

Deep learning research on skin lesion classification has evolved from initial architecture testing to clinical demonstration. Initial progress in this field focused on improving CNN classification. Deep residual learning was achieved with Yu et al. (2017) by adding residual or skip connections to Res-Net architectures to overcome the optimization bottlenecks of very deep networks. By addressing the vanishing gradient problem, these connections made deeper models more effective and provided a higher classification accuracy. They demonstrated the attractiveness of deep residual learning on ISBI 2016 and further demonstrated the potential of deep residual networks in medical imaging.

Esteva et al. (2017) published a landmark study which showed how important massive datasets are to the performance of deep learning models. By training an Inception-v3 network on a vast dataset of more than 129,000 clinical images with more than 2,000 skin diseases, they were able to match the diagnostic accuracy of board-certified dermatologists in skin cancer diagnosis. This was an impressive accomplishment that demonstrates deep learning's potential to be a clinical decision-making tool. However, since both of these early works focus on the classification accuracy rather than interpretability, they left the decision-making process of the models opaque, which has compromised clinical trust and has underlined the need for explainable artificial intelligence (XAI).

Recent research in the field of AI medical models has shifted from a very early stage of developing models to working with the deployment validation benchmarked against clinical practitioners and developing standardized evaluation scales. A key study on skin cancer detection is often cited in education literature, and in fact this AI model is much more efficient than most clinicians but still has clear clinical application boundaries.

At the same time, Tschandl et al. (2018) have released the HAM10000 dataset to address the serious shortage of high-quality training data. The dataset was collected from over 10,000 thermoscopic images of pigmented lesions in seven types of clinics. By providing a diverse open-access benchmark, HAM10000 provided researchers with a standard framework to train stronger models, run fair comparisons and build systems that generalize well to new data. It remains a foundational dataset for testing deep learning tools in dermatology today.

Table 1: *Summary and comparison of the most relevant related studies.*

Study (Year)	Model / Technique	Dataset	Explainability	Reported performance / key finding	Limitations
Yu et al. (2017)	Very deep residual networks (>50 layers)	ISBI 2016	None	Early proof that ResNet depth helps melanoma recognition	Small benchmark dataset; no explainability; limited generalization evaluation.
Esteva et al. (2017)	Inception-v3 CNN	129,450 clinical images	None	Matched 21 dermatologists	Used mainly clinical images rather than dermoscopy; black-box model; high computational cost.
Haenssle et al. (2018)	Inception-v4 CNN	Dermoscopic melanoma set	None	CNN AUC 0.86 vs dermatologists 0.79	Focused mainly on melanoma detection; limited interpretability; external validation was limited.
Tschandl et al. (2018)	Dataset contribution	HAM10000 (10,015; 7 cls)	None	Standard dermoscopy benchmark	Dataset imbalance and duplicated visual patterns; not a classification model.
Brinker et al. (2019)	CNN (ResNet-based)	12,378 melanoma images	None	Beat 136 of 157 dermatologists	Melanoma-focused evaluation; no explainability; possible dataset bias.
Combalia et al. (2019)	Dataset contribution	BCN20000	None	Realistic in-the-wild lesions for ISIC 2019	Class imbalance remained; dataset quality varies; not a predictive framework.
Codella et al. (2019)	Challenge protocol	ISIC 2019 (8 cls)	None	Defined the eight-class benchmark	Benchmark only; severe class imbalance; no model architecture proposed.
Gessert et al. (2020)	EfficientNet ensemble + metadata	ISIC 2019 (8 cls)	None	Won ISIC 2019; top accuracy, heavy	Large ensemble requires high computational resources and is unsuitable for real-time deployment.
Kassem et al. (2020)	Modified GoogLeNet + TL	ISIC 2019 (8 cls)	None	~94.9% accuracy; imbalance addressed	Relied heavily on augmentation; no explainability; limited external validation.
Mahbod et al. (2020)	Multi-scale multi-network ensemble	ISIC 2018	None	~86.2% balanced accuracy	Complex architecture; computationally expensive; difficult to deploy on resource-limited devices.
Nigar et al. (2022)	Deep CNN + LIME	ISIC 2019 (8 cls)	LIME	~94.5% acc.; explanations clinically aligned	LIME explanations may be unstable and computationally intensive; Grad-CAM not evaluated.
Munjal et al. (2024)	Inception-v3 + Grad-CAM + LIME	HAM10000 aug. (7 cls)	Grad-CAM + LIME	~96% acc.; AUC ~99.8%	Performance depends on heavy augmentation; explainability evaluated qualitatively rather than quantitatively.
Gayatri & Aarthy (2024)	DL + augmentation / regularisation	ISIC 2019 (8 cls)	Saliency-based XAI	Reduced overfitting on imbalance	Saliency maps provide limited explanation quality; no comparison with multiple XAI methods.
Mia et al. (2025)	Lightly tuned DenseNet-169	ISIC (multi-class)	None	Balanced classification via augmentation	No explainability; dependence on augmentation; limited real-world clinical validation.
Shakya et al. (2025)	VGG19 / ResNet-18 / MobileNet-V2	Dermoscopic images	LIME	ResNet-18 competitive and efficient	LIME only; relatively small dataset; no external clinical testing.

Summary and Research Gap

Taken together, these studies show a clear pattern. Deep CNNs with transfer learning now understand skin lesions to the same level as dermatologists, and ISIC 2019 is a standard benchmark for multi-class tasks. Two issues keep recurring: class imbalance — tackled by augmentation, re-sampling or cost-sensitive losses — and interpretability. Explainable AI is beginning to take hold, but there are still many top-performing models that remain black boxes and only a few works (Nigar et al., 2022; Munjal et al., 2024) combine competitive accuracy with image-level explanations and they do so with heavy backbones. Our thesis starts from this point of view: we use a lightweight ResNet-18 on a curated and roughly balanced dataset, use augmentation and class weighting to reduce the remaining skew and LIME to explain every decision so that accuracy and trust are achieved together rather than as opposing goals.

Chapter 3: Dataset Selection

Choice of Dataset

In this work, we use a curated multi-source skin-condition dataset drawn from open dermatological image archives. The selection was intentional. It provides a dataset of 10 clinically relevant categories, such as malignant and benign neoplasms but also common inflammatory and infectious diseases, and the classification problem is broader and more typical of routine dermatology. It is roughly balanced across classes, which reduces majority-class bias common in public collections, and because it is extracted from public repositories the experiments will be reproducible. (1, 2, 18)

Source and Composition

The dataset is generated from a number of public sources. The dermatoscopic images of the classic types of lesion (melanoma, melanocytic nevus, benign and actinic keratoses, dermatofibroma, squamous cell carcinoma and vascular lesions) are mainly from ISIC and clinical photographs provide examples for inflammatory and infectious types (atopic dermatitis and the tinea/ringworm/candidiasis group) and the healthy (normal-skin) reference class. The combination of these sources allows for different imaging equipment, anatomical sites and capture situations and is therefore more realistic and challenging than any single source.

Characteristics

The active dataset comprises 878 images assigned to training and validation splits, organised under a ten-class taxonomy: actinic keratosis, atopic dermatitis, benign keratosis, dermatofibroma, a healthy (normal-skin reference) category, melanocytic nevus, melanoma, squamous cell carcinoma, the tinea/ringworm/candidiasis group, and vascular lesion (Table 2). Images are colour photographs with varying resolutions and aspect ratios, so they are rescaled to a uniform input size prior to training. The taxonomy encompasses malignant (melanoma, squamous cell carcinoma), pre-malignant (actinic keratosis), benign (melanocytic nevus, benign keratosis, dermatofibroma, vascular lesion) and non-neoplastic (atopic dermatitis, tinea/ringworm/candidiasis) conditions. The healthy class serves as a normal-skin reference and is reserved in the current splits, to be filled as more images are gathered. This work relies solely on the visual content of images rather than any patient metadata.

Class Distribution and the Imbalance Challenge

And unlike the highly biased public data, the curated dataset is quite balanced: the most populated classes contain around one hundred images with the tinea/ringworm/candidiasis group slightly smaller at seventy-six examples. The largest populated class is only about 1.3 times the size of the smallest and is far less extreme than the scale of ISIC 2019. This is because of the majority-class balance and the main problem is the relative size of the sample and not the imbalance. So, in Chapter 4, we focus on transfer learning and augmentation to improve generalisation and introduce light weights for the remaining skew. (5, 20)

Table 2: Class distribution of the curated skin-condition dataset across the training and validation splits.

ID	Category	Train	Val.	Total	Share
0	Actinic keratosis	80	20	100	11.4%
1	Atopic Dermatitis	81	21	102	11.6%
2	Benign keratosis	80	20	100	11.4%
3	Dermatofibroma	80	20	100	11.4%
4	Healthy (normal-skin reference)	—	—	reserved	—
5	Melanocytic nevus	80	20	100	11.4%
6	Melanoma	80	20	100	11.4%
7	Squamous cell carcinoma	80	20	100	11.4%
8	Tinea / Ringworm / Candidiasis	56	20	76	8.7%
9	Vascular lesion	80	20	100	11.4%
	Total (populated categories)	697	181	878	100%

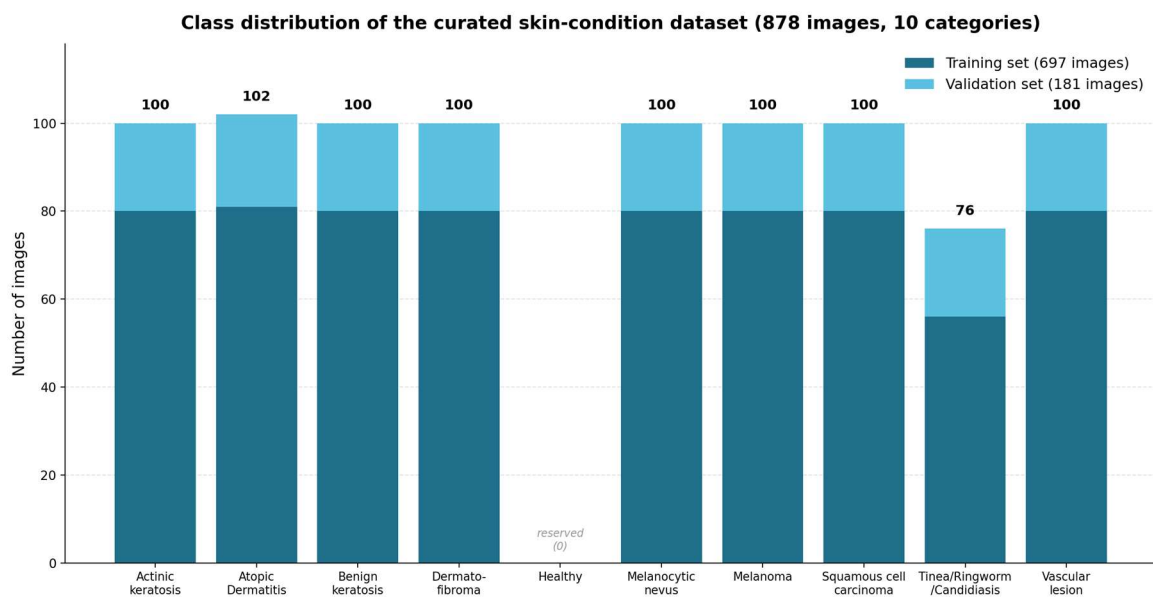


Figure 1: Class distribution of the curated skin-condition dataset, showing the approximately balanced categories and the reserved healthy reference category.

Accessibility

The images are sourced from freely available dermatology data (most notably the ISIC website, but also from the public clinical image collections) and are solely used for academic analysis. Because data and class definitions are available in public, the dataset can be reconstructed and the experiments are replicated for transparent model development and evaluation as discussed in the next chapter.

Chapter 4: Proposed Framework

Overview of the Framework

The proposed pipeline accepts a raw dermatological image and outputs two products simultaneously: a predicted skin-condition label and a visual justification for that prediction. The system proceeds through sequential stages: data preparation, imbalance handling, classification by a pre-trained ResNet-18, explainability via LIME, and evaluation. The following sections describe each stage and the tools employed.

Proposed Framework — Explainable Skin-Lesion Classification (ResNet-18 + LIME)

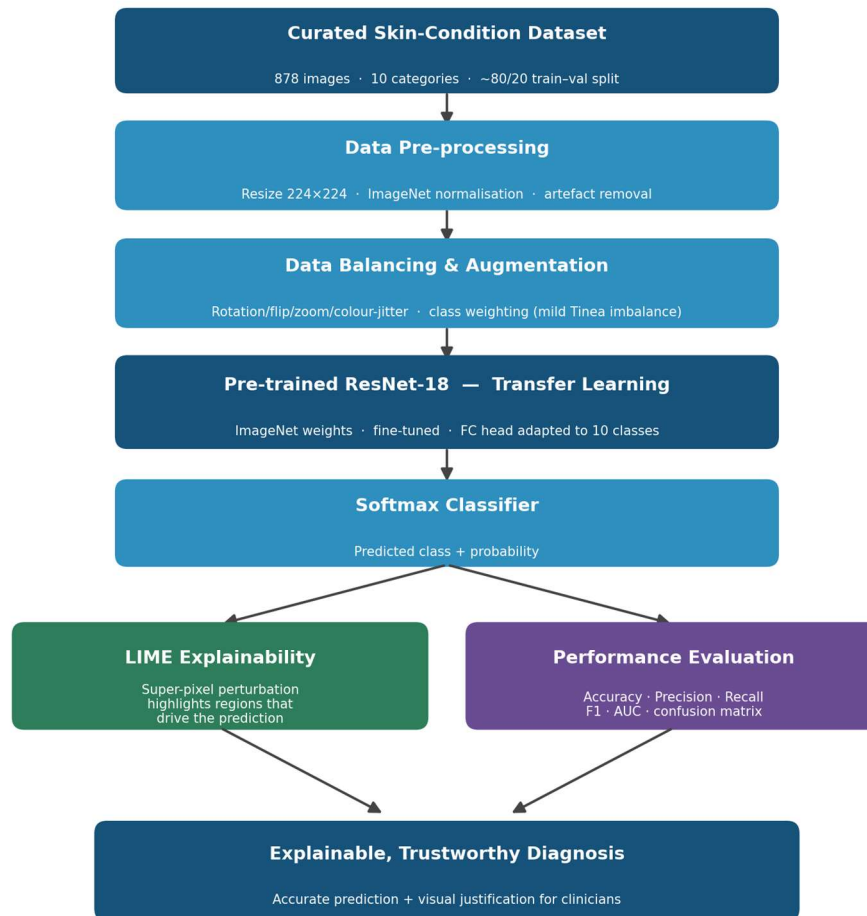
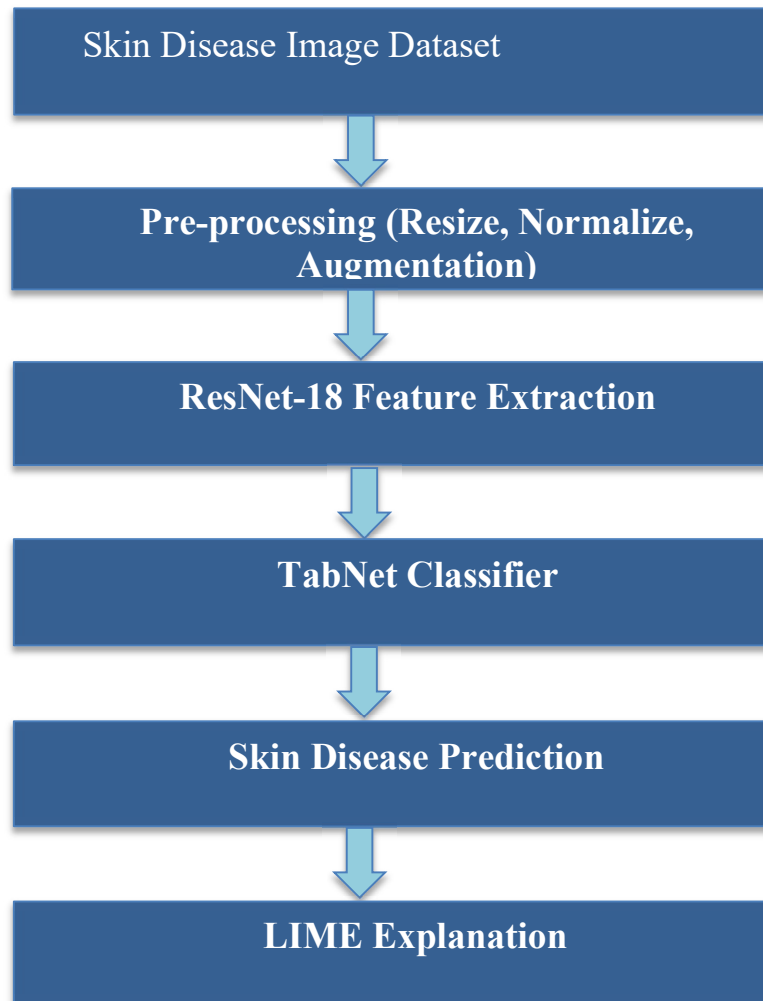


Figure 2: The proposed framework for explainable skin-lesion classification using a pre-trained ResNet-18 and LIME.

Methodology Overview:

The proposed framework combines a pre-trained ResNet-18 model with TabNet to classify skin diseases. First, ResNet-18 is used to extract deep features from skin images after the pre-processing stage. These deep features are provided to TabNet that will select the most relevant features through its attentive feature selection mechanism and performs the classification task. To make the interpretation of the model more interpretable, LIME is used to explain the prediction results by highlighting the image regions that contribute most to the final decision. TabNet provides feature importance scores that help to identify the features that affect the classification process.

Framework Flow:



Data pre-processing

Given the diversity in image size and quality, the initial stage standardizes input. The image dimensions are resized to 224×224 pixels (i.e., the input dimension of ResNet-18) and pixel intensities are normalized with ImageNet mean and standard deviation to maintain the same value as the model was pre-trained on. We avoid obvious dermatological clutter (hair, ink marks, rulers and dark lens vignetting) where possible since these can influence learning and thus the curated collection is divided into training and validation sets (around 80% and 20%, i.e., 697 training and 181 validation images) so that the same class is captured across the two sets. (3, 6)

Data Augmentation and Class Balancing

Two complementary methods can maximise learning from the small and relatively balanced data set. Firstly, data augmentation leads to plausible variants of lesions via random rotation, horizontal and vertical flips, zooms and relatively minor brightness/contrast changes, thus making the training set larger and acting as a regulariser to reduce overfitting. Secondly, light class weighting compensates for the small under-representation of the tinea/ringworm/candidiasis group, since the loss function is weighted inversely to class frequency so errors on smaller classes have a greater penalty. Focal loss is also considered as an alternative to further emphasise difficult examples. Together these techniques reduce the risk that the network will neglect clinically important minority categories.

Classification Model: Pre-trained ResNet-18

The core classifier is a ResNet-18 with ImageNet weights. Convolutional backbone (residual blocks with skip connections) is retained and fine-tuned with dermatology images and the original 1,000-way final layer is replaced with a fully connected layer of 10 neurons (one per class) and a softmax to find probabilities. Fine-tuning takes place with a low learning rate in such a way that general visual features from ImageNet are gently adapted to dermatology images. ResNet-18 is chosen as the ideal choice as compared to larger networks because it is accurate (trains faster and uses less compute), which makes it suitable for the dermatology field. (3, 6, 8)

Training Setup

Training uses the Adam optimiser with a weighted cross-entropy loss, images supplied in mini-batches and a learning-rate schedule that decays over time. Validation performance is checked every epoch and early stopping halts training when improvement ceases, preventing overfitting and saving resources. Table 3 lists sensible starting hyper-parameters to be explored empirically in the implementation phase.

Table 3: Main hyper-parameters of the proposed configuration.

Setting	Value
Input image size	224 x 224 x 3
Backbone network	ResNet-18 (pre-trained on ImageNet)
Output layer	Fully connected, 10 neurons + softmax
Loss function	Cross-entropy with class weighting (focal loss considered)
Optimizer	Adam
Initial learning rate	1e-4 with scheduler
Batch size	32
Epochs	Up to 50, with early stopping
Train / validation split	≈ 80 % / 20 % (697 / 181 images, stratified)
Framework	Py Torch + torch-vision

Explainability with LIME

After the classifier makes the prediction, LIME is used to explain it. Its method first divides the image into super-pixels (patches of neighboring pixels with similar properties) and generates many perturbed copies of the image, where random parts of the super-pixels are masked out. The trained ResNet-18 is evaluated on each perturbed image and LIME captures how the predicted probability of the target class changes as different regions are removed. Based on these results, LIME finds a simple and interpretable surrogate model that will locally approximate the deep model, and rank super-pixels on the basis of their importance. The result is an overlay that highlights the most influential areas of the image that are most responsible for the decision. If these areas correspond to the lesion itself, confidence in the model increases; if they point to background or artefacts (hair, rulers, ink), the explanation shows potential shortcomings— the transparency clinicians demand. (14)

Evaluation Strategy

As a qualitative result, we will evaluate our framework on both quantitative and qualitative grounds. Quantitatively, we will report classification performance in terms of accuracy as well as per-class recall and F1 (with average summaries), so that good results on common classes cannot mask poor performance on rarer ones; ROC AUC and confusion matrices will also help to better understand which classes are confused. Recall on malignant classes is of priority, as missing a cancerous lesion is worse than a false alarm. Qualitatively, we will look at LIME outputs to see if the model is targeting clinically relevant regions of the image and whether that is a good way to measure our model for use in practice.

Tools and Technologies

We will implement our model and training pipeline in Python using Py-Torch and torchvision, which feed the pre-trained ResNet-18 weights. We will use OpenCV, Pillow and Albumentations for image processing and augmentation; NumPy and pandas will hold the data; and scikit-learn will provide evaluation metrics. LIME and scikit-image will take care of explanations and super-pixel segmentation; Matplotlib will show the outcomes. We'll train on a GPU using Google Colab or a similar environment to keep the turnaround time on the test manageable.

References

- [1] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., and Halpern, A. "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)." arXiv preprint arXiv:1902.03368, 2019.
- [2] Combalia, M., Codella, N. C. F., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., Carrera, C., Barreiro, A., Halpern, A. C., Puig, S., and Malvehy, J. "BCN20000: Dermoscopic Lesions in the Wild." arXiv preprint arXiv:1908.02288, 2019.
- [3] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. "ImageNet: A Large-Scale Hierarchical Image Database." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 248-255, 2009.
- [4] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. "Dermatologist-level classification of skin cancer with deep neural networks." Nature, 542(7639), 115-118, 2017.
- [5] Gayatri, E., and Aarthy, S. L. "Reduction of overfitting on the highly imbalanced ISIC-2019 skin dataset using deep learning frameworks." Journal of X-Ray Science and Technology, 32(3), 2024.
- [6] He, K., Zhang, X., Ren, S., and Sun, J. "Deep Residual Learning for Image Recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778, 2016.
- [7] Kassem, M. A., Hosny, K. M., and Fouad, M. M. "Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning." IEEE Access, 8, 114822-114832, 2020.
- [8] Krizhevsky, A., Sutskever, I., and Hinton, G. E. "ImageNet Classification with Deep Convolutional Neural Networks." Advances in Neural Information Processing Systems (NeurIPS), 25, 1097-1105, 2012.
- [9] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollar, P. "Focal Loss for Dense Object Detection." Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2980-2988, 2017.
- [10] Lundberg, S. M., and Lee, S.-I. "A Unified Approach to Interpreting Model Predictions." Advances in Neural Information Processing Systems (NeurIPS), 30, 4765-4774, 2017.
- [11] Mia, M. M., et al. "Exploring Augmentation Strategies for Balanced Skin Lesion Classification: An Explainable Lightly Tuned DenseNet-169 Architecture." 2025.
- [12] Munjal, G., Bhardwaj, P., Bhargava, V., et al. "SkinSage XAI: An explainable deep learning solution for skin lesion diagnosis." Health Care Science, 3(6), 438-455, 2024.
- [13] Nigar, N., Umar, M., Shahzad, M. K., Islam, S., and Abalo, D. "A Deep Learning Approach Based on Explainable Artificial Intelligence for Skin Lesion Classification." IEEE Access, 10, 113715-113725, 2022.
- [14] Ribeiro, M. T., Singh, S., and Guestrin, C. "Why Should I Trust You?: Explaining the Predictions of Any Classifier." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144, 2016.
- [15] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618-626, 2017.
- [16] Shakya, M., Patel, R., and Joshi, S. "A comprehensive analysis of deep learning and transfer learning techniques for skin cancer classification." Scientific Reports, 15, 4633, 2025.
- [17] Simonyan, K., and Zisserman, A. "Very Deep Convolutional Networks for Large-Scale Image Recognition." International Conference on Learning Representations (ICLR), 2015.
- [18] Tschandl, P., Rosendahl, C., and Kittler, H. "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." Scientific Data, 5, 180161, 2018.
- [19] Brinker, T. J., Hekler, A., Enk, A. H., et al. "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task." European Journal of Cancer, 113, 47-54, 2019.
- [20] Gessert, N., Nielsen, M., Shaikh, M., Werner, R., and Schlaefer, A. "Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data." MethodsX, 7, 100864, 2020.
- [21] Haenssle, H. A., Fink, C., Schneiderbauer, R., et al. "Man against machine: diagnostic

performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists." *Annals of Oncology*, 29(8), 1836-1842, 2018.

[22] Mahbod, A., Schaefer, G., Wang, C., Dorffner, G., Ecker, R., and Ellinger, I. "Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification." *Computer Methods and Programs in Biomedicine*, 193, 105475, 2020.

[23] Yu, L., Chen, H., Dou, Q., Qin, J., and Heng, P. A. "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks." *IEEE Transactions on Medical Imaging*, 36(4), 994-1004, 2017.

[24] Database <https://www.kaggle.com/datasets/ujjwalsinha01/skin-disease-classification-dataset>