Hybrid Deep Learning System for Nail Disease Diagnosis Combining Synthetic Data, Explainable AI and Multi-Model Architectures

A dissertation submitted for the Degree of Master of Computer Science

T. M. D. R. M. Attanayake

University of Colombo School of Computing

2025

Declaration

Name of the student: T. M. D. R. M. Attanayake
Registration number: 2022/MCS/004
Name of the Degree Programme: Master of Computer Science
Project/Thesis title: Hybrid Deep Learning System for Nail Disease Diagnosis Combining Synthetic Data, Explainable AI and Multi-Model Architectures

1. The project/thesis is my original work and has not been submitted previously for a degree at this or any other University/Institute. To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.

2. I understand what plagiarism is, the various types of plagiarism, how to avoid it, what my resources are, who can help me if I am unsure about a research or plagiarism issue, as well as what the consequences are at University of Colombo School of Computing (UCSC) for plagiarism.

3. I understand that ignorance is not an excuse for plagiarism and that I am responsible for clarifying, asking questions and utilizing all available resources in order to educate myself and prevent myself from plagiarizing.

4. I am also aware of the dangers of using online plagiarism checkers and sites that offer essays for sale. I understand that if I use these resources, I am solely responsible for the consequences of my actions.

5. I assure that any work I submit with my name on it will reflect my own ideas and effort. I will properly cite all material that is not my own.

6. I understand that there is no acceptable excuse for committing plagiarism and that doing so is a violation of the Student Code of Conduct.

Signature of the Student	**Date (DD/MM/YYYY)**
	25/06/2025

Certified by Supervisor(s)

This is to certify that this project/thesis is based on the work of the above-mentioned student under my/our supervision. The thesis has been prepared according to the format stipulated and is of an acceptable standard.

	Supervisor 1	**Supervisor 2**	**Supervisor 3**
Name	Mr. Gihan P. Seneviratne		
Signature			
Date	25/06/2025		

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Mr. Gihan P. Seneviratne, for their invaluable guidance, continuous support and encouragement throughout this research. Their insightful feedback and expertise have been instrumental in shaping the direction of this thesis and I am truly grateful for their patience and dedication.

I extend my heartfelt thanks to the academic and administrative staff of the University of Colombo School of Computing (UCSC) for providing the necessary resources, knowledge and assistance that facilitated the successful completion of this research. Their commitment to academic excellence and research has greatly contributed to my learning journey.

I am also profoundly grateful to my family for their unwavering support, patience and encouragement throughout this endeavor. Their belief in me has been a constant source of motivation, and I could not have completed this work without their love and understanding.

To my friends, who stood by me through every challenge, offered words of encouragement and provided a much-needed balance between work and life. I truly appreciate their kindness, support and companionship.

Finally, I am thankful to everyone who has contributed to this research in any way, whether through direct assistance, thoughtful discussions or simply by inspiring me to persevere.

T. M. D. R. M. Attanayake
2025/03/16

ABSTRACT

Nail abnormalities serve as crucial indicators of a wide range of systemic, dermatological and nutritional disorders, necessitating early and precise detection for effective medical intervention. However, conventional diagnostic methods rely on subjective visual assessments by healthcare professionals which can lead to inconsistencies and delays in diagnosis. This research presents an AI-powered Nail Abnormality Detection System which has been designed to automate the identification and classification of five common nail abnormalities which are clubbing, spoon nails (koilonychia), black nails, splinter hemorrhages and nail pitting alongside normal nails. By integrating advanced deep learning techniques, this study addresses key challenges in medical image analysis including data scarcity, model interpretability and real-time diagnostic efficiency.

The proposed system utilizes a hybrid deep learning framework combining Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) to enhance feature extraction by improving both classification accuracy and generalization. Given the limited availability of high-quality annotated datasets, this research also explores synthetic data generation using diffusion models to augment training data while ensuring balanced representation of various nail conditions. In addition, Explainable AI (XAI) techniques such as Grad-CAM, Ablation-CAM and SIDU, are employed to provide visual interpretability, enhancing model transparency and enabling medical professionals to validate AI-driven predictions. Furthermore, YOLO-based segmentation is integrated to precisely localize affected regions by facilitating severity assessment and aiding in clinical decision-making.

Comprehensive experimental evaluations demonstrate that the system achieves great classification accuracy, high precision and recall across multiple datasets while outperforming traditional diagnostic approaches. The findings establish that the combination of deep learning and computer vision techniques significantly enhances diagnostic reliability, offering a scalable, accessible, and cost-effective solution for early disease detection. The study not only advances the field of AI-driven dermatological diagnostics but also lays the groundwork for future research in AI-assisted healthcare applications. With its ability to provide rapid, consistent and interpretable results, this system has the potential to revolutionize telemedicine, point-of-care screening and large-scale population health monitoring.

The pipeline begins with a Nail-Filter CNN that screens photographs for relevance, achieving 98 % overall accuracy, 100 % recall for nail images and 100 % precision for non-nail images (F1 = 0.98), thus eliminating most irrelevant inputs before further processing. A hybrid classification stage then combines ViTs and CNNs: the ViT model—trained on both real and diffusion-generated synthetic images—reaches 95 % overall accuracy with class-wise precision/recall up to 97 % / 96 %, while the best CNN verifier (Inception-ResNet-V2, Adam, LR = 0.001) delivers 94 % accuracy. Their outputs are fused by a dual-stage consensus algorithm that boosts reliability without added latency. For localization and severity assessment, an instance-segmentation module based on YOLOv8-Nano produces pixel-level masks (mAP = 86.6 %, precision = 93.6 %, recall = 82.1 % on a 600-image set), improving performance by nearly 4 % over YOLOv7. Throughout, XAI tools generate heat-maps so clinicians can verify that model attention aligns with pathological regions. Collectively, these results confirm that the hybrid framework delivers more than 95 % diagnostic accuracy, robust segmentation and full interpretability while running in real time on commodity GPUs—offering a cost-effective route to population-scale nail-disease screening and establishing a foundation for broader AI-assisted dermatological applications.

Table of Contents

Declaration ... i

ACKNOWLEDGEMENTS ... ii

ABSTRACT ... iii

LIST OF FIGURES ... viii

LIST OF TABLES ... x

ABBREVIATIONS ... x

CHAPTER 1: INTRODUCTION ... 1

1.1 Problem Domain ... 1

1.2 Problem Statement ... 2

1.3 Motivation ... 5

1.4 Research Questions ... 6

1.5 Research Objectives ... 7

1.6 Research Contribution ... 7

1.7 Scope ... 9

CHAPTER 2: LITEREATURE REVIEW ... 11

2.1 Background ... 11

Nail Pitting ... 11

Black Line ... 12

Splinter haemorrhages ... 13

Nail Clubbing ... 13

Spoon Nails (Koilonychia) ... 14

2.3 Significance of Related Work ... 16

2.4 Limitations ... 17

2.5 Relation to Current Research and Identified Gaps ... 18

2.6 Explainable Artificial Intelligence (XAI) Methods ... 20

2.7 The Need for Explainability in AI ... 21

2.8 Classification of XAI Methods 22
2.8.1 Model-Agnostic Methods 22
2.8.2 Model-Specific Methods 23
2.9 Image-Based XAI Methods 24
2.9.1 Attribution-based Methods 24
2.9.2 Non-Attribution Methods 26
2.10 Evaluation of XAI Methods 27
2.11 Applications of XAI in Medical Imaging 29
2.12 Challenges in Future Directions of XAI 30
2.13 Comparative Analysis 31
CHAPTER 3: METHODOLOGY 36
3.1 Overview 36
3.2 Multi-model Architecture 37
3.3 Methodology Analysis 38
3.3.1 Nail Filter Model 38
3.3.2 ViT Nail Diagnosis Model 41
3.3.3 CNN Nail Diagnosis Model (Verification Model) 44
3.3.4 Dual-Stage Nail Diagnosis & Verification Algorithm 50
3.3.5 Nail Abnormality Detection and Segmentation Model 52
3.3.6 Visual Saliency Algorithm 57
Vision Transformers (ViT) and Attention Maps 57
Convolutional Neural Networks (CNNs) and Saliency Maps 57
Differences Between Attention Maps (ViT) and Saliency Maps (CNNs) 58
Pros and Cons Between Attention Maps (ViT) and Saliency Maps (CNNs) 58
Attention Map Generation Using ViT 59
Convolutional Neural Networks (CNNs) and Saliency Maps 65
Ablation-CAM 68
CHAPTER 4: EVALUATION 70

4.1 Evaluation for Nail Filter Model......70
4.2 Evaluation of ViT Nail Diagnosis Model (Verification Model) 72
4.2.1 Experiment 1 72
4.2.2 Experiment 2 74
4.2.3 Best Hyperparameters 76
4.3 Evaluation of CNN Nail Diagnosis Model 77
4.3.1 Experiment 1 77
4.3.3 Experiment 3 78
4.3.4 Experiment 4 78
4.3.5 Experiment 5 80
4.3.6 Best Hyper parameter......82
4.4 Evaluation of Nail Abnormality Detection and Segmentation Model 83
4.5 Results of Attention Map Generation Using ViT......88
4.6 Results of Convolutional Neural Networks (CNNs) and Saliency Maps 89
CHAPTER 5: CONCLUSION AND FUTURE WORK......92
5.1 Conclusion......92
5.2 Future Work 93
REFERENCES I

LIST OF FIGURES

Figure 1. Toenail Pitting 12
Figure 2. Fingernail Pitting 12
Figure 3. Black Line Abnormality 13
Figure 4. Splinter haemorrhages 13
Figure 5.Nail Clubbing 14
Figure 6. Spoon Nails 14
Figure 7. High-level Architecture Diagram 38
Figure 8. Confidence Score Normalization Process 50
Figure 9. Weighted Confidence Score Calculation 51
Figure 10. Weighted Score Calculation for Disagreement Cases 51
Figure 11. Final Confidence Score 51
Figure 12. Annotation of Nail Black Line Using Roboflow 54
Figure 13. Annotation of Nail Splinters Using Roboflow 54
Figure 14. Image Reading and Preprocessing Code 59
Figure 15. Patch Creation for Vision Transformer 60
Figure 16. Positional Encoding in Vision Transformer 61
Figure 17. Multi-Head Self-Attention in Vision Transformer 62
Figure 18. Transformer Block for Multi-Head Attention Aggregation 63
Figure 19.Extracting Attention Weights from Vision Transformer 63
Figure 20. Generating Heatmap from Attention Weights 64
Figure 21. Classification Report for Nail Filter Model 71
Figure 22. Confusion Matrix for Nail Filter Model 72
Figure 23. Confusion Matrix for ViT Nail Diagnosis Model 73
Figure 24. Classification Report for ViT Nail Diagnosis Model 73
Figure 25. Confusion Matrix for Optimized ViT Nail Diagnosis Model 75
Figure 26. Classification Report for Optimized ViT Nail Diagnosis Model 75
Figure 27. Sample Images from Nail Diagnosis Dataset 76
Figure 28. Confusion Matrix for Optimized Learning Rate (0.01) 79
Figure 29. Classification Report for Optimized Learning Rate (0.01) 80
Figure 30. Confusion Matrix for Best-Performing Configuration (Adam, LR=0.001) 81
Figure 31. Classification Report for Best-Performing Configuration (Adam, LR=0.001) 82
Figure 32. Sample Images from Optimized Nail Diagnosis Dataset 83
Figure 33. Training and Validation Metrics for YOLOv7 and YOLOv8 Models 84
Figure 34. Instance Segmentation Results for Nail Abnormality Detection 87
Figure 35. Attention Map for Nail Abnormality Detection for samples 88

Figure 36. Interpretability and Visualization of Nail Abnormality Detection Using XAI Methods..... 89
Figure 37. XAI-Based Model Interpretability Visualizations 90

LIST OF TABLES

Table 1. Comparative Analysis 31
Table 2. Automated Random Image Augmentation Parameters 43
Table 3. Model Experiments Comparison 47
Table 4. Differentiation of ViT vs. CNNs 58
Table 5. Pros and Cons of ViT vs. CNNs 58
Table 6. Learning Rate vs. Accuracy for ViT Nail Diagnosis Model 72
Table 7. Optimizer Comparison for ViT Nail Diagnosis Model 74
Table 8. Accuracy Comparison of CNN-Based Transfer Learning Models 77
Table 9. Accuracy of Top Three Models After Architectural Modifications 78
Table 10. Accuracy Improvement After Architectural Enhancements 78
Table 11. Learning Rate Optimization Results 79
Table 12.Optimizer Comparison for Final Experiment 81
Table 13. YOLOv7 and YOLOv8 models results 84

ABBREVIATIONS

AI – Artificial Intelligence

CNN – Convolutional Neural Network

DL – Deep Learning

EHR – Electronic Health Records

ML – Machine Learning

NLP – Natural Language Processing

ViT – Vision Transformer

XAI – Explainable Artificial Intelligence

YOLO – You Only Look Once

mAP – Mean Average Precision

LR – Learning Rate

RMSprop – Root Mean Square Propagation

XGBoost – Extreme Gradient Boosting

SVM – Support Vector Machine

ANN – Artificial Neural Network

ReLU – Rectified Linear Unit

LoRA – Low-Rank Adaptation

SD – Stable Diffusion

AMPS – Automated Medical Palmistry System

KNN – K-Nearest Neighbors

NIPS – Nail Image Processing System

IoU – Intersection over Union

Adadelta – Adaptive Delta Optimization

CHAPTER 1: INTRODUCTION

1.1 Problem Domain

Nails are not just cosmetic features of the human body but also, they serve crucial biological and functional purposes. They act as protective barriers, shielding the sensitive tissues beneath and enable precision tasks such as scratching, gripping, and grooming. Structurally, nails consist of keratin which is a tough protein forming a complex matrix including the visible nail plate, the nail matrix, the nail bed and surrounding grooves. Their intricate structure highlights their significance in daily functions and overall well-being. Beyond these roles, nails are also widely recognized in the medical field as valuable diagnostic tools.

Medical professionals often assess an individual's health by examining the condition of their nails. Changes in nail color, texture, thickness and shape can indicate a variety of health conditions. For example, systemic illnesses like cardiovascular and respiratory diseases frequently manifest as nail abnormalities which makes nails a valuable diagnostic tool. Common conditions include Beau’s lines, spoon-shaped nails (koilonychia), clubbing and nail pitting. Each of these is linked to specific health concerns: Beau’s lines may suggest systemic illness or trauma, clubbing is often associated with lung disease and nail pitting is commonly seen in psoriasis or other autoimmune disorders. In addition to their functional role, nails act as essential health indicators, helping clinicians detect potential underlying diseases (Podell et al., 2023 ; “7 fingernail problems not to ignore,” 2023).

Although nails have significant diagnostic potential, nail health evaluation still largely depends on visual inspections by healthcare providers. These assessments are often subjective, limited by human observation and susceptible to inconsistencies. Moreover, early-stage nail changes can be subtle and difficult to detect without advanced imaging or specialized training. This highlights the urgent need for automated diagnostic systems that offer high accuracy and scalability by ensuring a standardized evaluation process, minimizing human error and enabling early detection of abnormalities. The limitations of conventional diagnostic methods highlight a significant gap in medical technology which could be effectively addressed through advancements in computer vision and artificial intelligence (AI) (“12 nail changes a dermatologist should examine,” n.d.).

This challenge is further compounded by the scarcity of large, annotated datasets needed to train AI models effectively. Collecting and labeling images of nail conditions across diverse populations is both time consuming and expensive. Additionally, variations in nail appearance due to factors such as ethnicity, age and underlying health conditions add to the complexity. Therefore, innovative solutions are required to expand data resources and utilize advanced machine learning architectures capable of managing these challenges (Pandit et al., 2013).

1.2 Problem Statement

Although nails are important for diagnosing health conditions, research and technology for automated nail disease detection are still lacking. Current diagnostic methods predominantly rely on visual assessments performed by healthcare professionals which are often subjective, inconsistent and prone to human error. These traditional approaches can overlook subtle and early-stage indicators of nail abnormalities resulting in delayed diagnosis and potentially compromising patient outcomes (Pandit et al., 2013).

There is a significant gap in developing integrated diagnostic systems that combine robust image segmentation, precise feature-based classification and effective early-stage detection (Mente, 2017). Existing methods are typically limited in scope, focusing on isolated features such as color or texture without incorporating a holistic analysis that could improve diagnostic reliability. Furthermore, segmentation techniques in prior studies have often lacked the precision necessary to isolate nail features from complex backgrounds while classification models have struggled to achieve high accuracy due to insufficient data diversity and model complexity.

This research aims to address these gaps by developing an automated system capable of detecting five common nail abnormalities: clubbing, spoon nails (koilonychia), black nails, splinter hemorrhages and nail pitting ("7 fingernail problems not to ignore," 2023). Each of these conditions has distinct health implications as below.

- ***Clubbing***: Associated with chronic lung diseases, heart conditions, and gastrointestinal disorders. Failure to diagnose early can lead to missed opportunities for treating underlying serious health issues.
- ***Spoon Nails (Koilonychia):*** Often linked to iron deficiency anemia or other systemic diseases, early detection can prompt necessary medical interventions to address nutritional deficiencies or underlying conditions.
- ***Black Nails***: Can be a sign of trauma, fungal infections, or more concerning issues such as melanoma. Misdiagnosis or delayed detection of melanomas can have life-threatening consequences.

- ***Splinter Hemorrhages***: Tiny blood clots under the nail bed, potentially indicative of systemic conditions like endocarditis or vascular disorders. Early identification is critical to managing these underlying conditions.
- ***Nail Pitting***: Frequently associated with psoriasis and autoimmune disorders. Early diagnosis can facilitate the initiation of treatment to manage chronic conditions and prevent further complications.

The exact computer science problem addressed by this research involves the development of an integrated, automated system for the accurate detection, classification and assessment of nail abnormalities using advanced computer vision and deep learning techniques. This problem can be broken down into several technical challenges (Ansari and Vijay, 2023):

I. **Automated Image Segmentation and Classification:** Accurately segmenting nail regions from complex and diverse image backgrounds is a major technical challenge. Nail images often vary in lighting, skin tone and shape by making the segmentation process more difficult. The key challenge is to develop a reliable model that can consistently isolate nail features across different conditions while ensuring that only the relevant regions are analyzed. Furthermore, after segmentation, these images must be classified into specific abnormality categories such as clubbing, spoon nails, black nails, splinter hemorrhages and nail pitting. This requires a robust classification system capable of extracting meaningful features while effectively handling the variations found in real-world datasets.

II. **Data Scarcity and Synthetic Data Generation:** Developing machine learning models for nail disease detection is challenging due to the limited availability of large, annotated datasets. Collecting and labeling high-quality images of nail abnormalities from diverse populations is both time consuming and expensive. To overcome this limitation, this research focuses on generating realistic synthetic nail images to expand training datasets and enhance the model’s ability to generalize effectively. This requires the use of diffusion models and other generative techniques to create diverse and representative synthetic data, ensuring improved performance in real-world applications.

III. **Hybrid Multi-Model Architecture:** A key challenge in this research is developing a hybrid multi-model architecture that leverages the strengths of both Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs). While CNNs are highly effective at capturing local features, ViTs excel in recognizing global patterns within images. Combining these models for complementary feature extraction and classification requires solving complex computational problems, including model optimization, efficient fusion strategies, and hyperparameter tuning.

IV. **Explainable AI (XAI) for Model Interpretability:** Another essential challenge in this research is ensuring that the diagnostic system is both interpretable and transparent. For medical professionals and users to trust the AI model's decisions, they need to understand how it arrives at its conclusions. This requires incorporating Explainable AI (XAI) techniques, such as saliency maps and attention mechanisms, to highlight the specific nail regions and features that influence predictions. However, striking a balance between high model performance and interpretability remains a complex and ongoing challenge in the field of AI.

V. **Anomaly Detection and Severity Assessment**: Beyond classification, the system must also assess the extent and severity of nail abnormalities. This requires implementing YOLO-based models for real-time instance segmentation and anomaly detection, enabling the identification of specific problem areas such as discoloration or abnormal growth patterns. Developing models that can perform this task efficiently and accurately is particularly challenging due to the high variability in nail appearances and disease presentations.

So that, the primary focus is developing a comprehensive AI-driven diagnostic system that addresses these challenges through:

- Accurate segmentation and classification of nail images using a hybrid deep learning approach.

- Synthetic data generation to overcome the scarcity of labeled data and improve model robustness.
- Model interpretability through Explainable AI techniques, ensuring trust and transparency.
- Efficient anomaly detection and severity assessment to provide meaningful and actionable insights for medical professionals and users.

Ultimately, this research aims to create an integrated and efficient solution that leverages cutting-edge computer vision and machine learning technologies, transforming nail disease diagnostics into a more accurate, interpretable, and accessible process.

1.3 Motivation

This research is driven by the crucial role that early and accurate detection of nail abnormalities plays in improving individual health outcomes. Nails serve as vital indicators of systemic and dermatological health conditions and therefore the changes in their appearance often provide the first warning signs of underlying diseases. Despite this, current diagnostic practices remain limited, relying heavily on subjective visual inspections by healthcare professionals. These traditional methods lack the precision and consistency needed for early diagnosis which can lead to delayed treatment and poorer health outcomes for patients.

The ability to automate and improve the accuracy of nail disease detection has far-reaching implications for healthcare systems worldwide. Early detection of conditions like clubbing which is associated with heart and lung diseases, or black nails which could indicate severe trauma or melanoma, can be lifesaving. Similarly, diagnosing conditions such as spoon nails linked to anemia or nail pitting associated with autoimmune disorders in their early stages can prompt timely medical intervention. By providing a reliable and accessible diagnostic tool, the burden on healthcare professionals can be reduced and patients can benefit from faster and more accurate assessments.

This research is driven by a desire to push the boundaries of computer vision and machine learning technologies. The advancement of deep learning models such as Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs), alongside innovative segmentation techniques like YOLO, is a great opportunity to revolutionize medical diagnostics. Combining these technologies with synthetic data generation and Explainable AI (XAI) methods not only improves diagnostic accuracy but also makes the process transparent and interpretable for users and healthcare professionals. The integration of these cutting-edge technologies into a comprehensive diagnostic system is one of the key motivators for this research (Mente, 2017).

A key motivation for this research is the challenge posed by limited and imbalanced datasets, which has hindered the development of accurate AI models in the medical field. To address this issue, advanced techniques such as fine-tuned diffusion models can be used to generate synthetic data, enhancing the diversity and robustness of training datasets. By tackling this challenge, this research aims to establish a new standard for addressing data scarcity in medical imaging applications (Meshram, 2016).

Beyond technological advancements, there is a pressing need for accessible and scalable diagnostic solutions, especially in resource-limited areas. Many communities worldwide lack access to dermatology specialists and advanced diagnostic tools for making early detection and treatment challenging. An automated nail disease detection system can bridge this gap by providing accurate and timely health assessments to individuals regardless of their location or financial situation. The potential to improve healthcare accessibility and equity, ultimately making a meaningful impact on people's lives, is the key motivation behind this research.

1.4 Research Questions

The research questions which are addressing from this research is as follows.

I. How can artificial intelligence be used to improve the accuracy and reliability of automated nail disease diagnosis?
II. What role can synthetic data play in overcoming the challenges of limited and imbalanced medical image datasets?
III. How can object detection and segmentation techniques help in identifying and assessing nail abnormalities more effectively?
IV. In what ways can Explainable AI enhance the trust and usability of AI-driven medical diagnostic systems?

1.5 Research Objectives

The research objective is to develop a comprehensive and automated AI-powered diagnostic system that accurately detects, classifies, and assesses five common nail abnormalities sunch as clubbing, spoon nails (koilonychia), black nails, splinter hemorrhages and nail pitting using a hybrid deep learning architecture combining Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), while addressing challenges of data scarcity, model interpretability and real-time performance through the integration of synthetic data generation, Explainable AI (XAI) and YOLO-based segmentation.

1.6 Research Contribution

This research will significantly contribute to the fields of computer vision, deep learning and medical diagnostics through the development of an innovative AI-driven system for the automated detection and assessment of nail abnormalities. The key contributions of this research are outlined below (Maniyan and Shivakumar, 2018 ; Rani et al., 2018; Maniyan et al., 2018):

I. **Development of a Hybrid Deep Learning Model:** This study introduces a novel multi-model architecture that effectively integrates Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs) for the accurate classification of nail abnormalities. By leveraging the complementary strengths of these models— CNNs for capturing local features and ViTs for understanding global image contexts—this research advances the state-of-the-art in medical image analysis. The hybrid model improves diagnostic accuracy and generalizability across diverse nail images, setting a new benchmark for disease classification.

II. **Synthetic Data Generation for Robust Model Training:** To address the challenge of limited and imbalanced datasets, this research proposes a custom synthetic data generation approach using fine-tuned diffusion models. These models can generate realistic and diverse nail images that represent various abnormalities. The use of synthetic data enhances the robustness of the machine learning models, mitigating the impact of data scarcity and class imbalance. This contribution provides a framework for other medical imaging applications facing similar data limitations.

III. **Explainable AI (XAI) for Transparent Diagnostics:** This research emphasizes model interpretability by integrating Explainable AI (XAI) techniques. Saliency maps for CNNs and attention mechanisms for ViTs are implemented to highlight the features and regions of nail images that influenced the model's predictions. This transparency fosters trust in the diagnostic system among medical professionals and users by providing clear visual explanations of the AI's decision-making process, a critical advancement in the adoption of AI in healthcare.

IV. **YOLO-Based Anomaly Detection and Severity Assessment:** The research contributes to real-time nail disease assessment with advanced YOLO models for anomaly detection and segmentation. By pinpointing areas of concern, such as discoloration, splinter hemorrhages, or abnormal growth patterns, the system provides a detailed analysis of the severity of nail conditions. This feature not only enhances diagnostic precision but also enables actionable insights for healthcare professionals, improving the overall effectiveness of disease management.

V. **Comprehensive and Scalable Diagnostic Tool**: The system developed in this research is designed to be user-friendly, scalable and adaptable to real-world medical environments. The intuitive interface allows users to easily upload images for analysis while the backend ensures efficient and accurate processing. By combining state-of-the-art computer vision technologies with a focus on usability and scalability, the research offers a practical solution that can be deployed in various healthcare settings by making advanced diagnostics accessible to a broader population.

VI. **Framework for Future Research in Medical Imaging**: The methodologies and techniques developed in this research provide a foundation for future advancements in medical imaging. The integration of hybrid deep learning models, synthetic data augmentation, and XAI approaches can be extended to other medical conditions, contributing to the broader field of AI-driven healthcare.

At the end, this research introduces an advanced automated nail disease detection system designed to bridge critical gaps in current diagnostic practices. By integrating state-of-the-art image segmentation, robust feature-based classification, synthetic data generation and Explainable AI, the system provides a comprehensive and reliable solution. These innovations not only improve the accuracy and transparency of diagnostic models but also contribute to making healthcare more accessible and scalable for a wider population.

1.7 Scope

The scope of this research is defined by the development and evaluation of a comprehensive, automated system for detecting and assessing common nail abnormalities using advanced computer vision and machine learning techniques. This study focuses on the following key components and limitations (Maniyan et al., 2018; Dhanashree et al., 2022; Nithya et al., 2019) :

I. **Target Abnormalities**: The system is specifically designed to detect six prevalent nail conditions including five disease types and normal nails.

- ***Clubbing***: Abnormal curvature of the nails, often linked to cardiovascular or respiratory diseases.
- ***Spoon Nails (Koilonychia):*** A concave shape associated with iron deficiency anemia or metabolic conditions.
- ***Black Nails***: Dark streaks or patches, potentially indicative of trauma, fungal infections, or melanoma.
- ***Splinter Hemorrhages***: Small blood clots under the nail bed, which may signal systemic vascular issues.
- ***Nail Pitting***: Small depressions on the nail surface, commonly associated with psoriasis or other autoimmune disorders.
- ***Normal Nails***: Smooth, evenly colored nails with a consistent shape, indicating healthy nail growth and overall well-being.

II. **Image Segmentation and Classification:** The research covers the implementation of hybrid deep learning models, including Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs), to perform precise segmentation and feature-based classification of nail images. The system aims to capture both local and global features of the nail to improve diagnostic accuracy.

III. **Synthetic Data Generation:** The study addresses the issue of limited datasets by employing synthetic data generation techniques, such as fine-tuned diffusion models, to augment the training data. This aspect of the research is limited to the generation of realistic nail images that correspond to the five target abnormalities.

IV. **Explainable AI (XAI) Techniques:** The research includes the integration of Explainable AI techniques to make the system's predictions transparent and interpretable. This involves visual saliency maps for CNNs and attention mechanisms for ViTs, which are limited to explaining the model's decision-making process for the specified nail conditions.

V. **User Interface and Accessibility**: The research focuses on developing a user-friendly interface that allows individuals to upload nail images for automated analysis. The system is designed to be scalable and deployable in various healthcare settings, including clinics and telemedicine platforms, but it is not intended for use as a standalone diagnostic tool without medical supervision.

And the exclusion of this scope is as below.

I. The system does not address other dermatological or systemic conditions that may also manifest in the nails.
II. The research does not cover the real-time deployment and integration of the system into existing healthcare IT infrastructure, although the potential for future integration is discussed.
III. The ethical implications and patient data privacy considerations are noted but not deeply explored within the scope of this research.

Concluding the introduction, the scope of this research lays the groundwork for future advancements, including the integration of additional nail conditions, improving the model's ability to generalize across diverse populations and evaluating its real-world application in clinical settings

CHAPTER 2: LITEREATURE REVIEW

The application of nail image processing for early disease detection has gained substantial momentum in recent years, driven by significant advancements in digital image processing and machine learning technologies. These advancements have made it possible to enhance diagnostic accuracy and efficiency, providing a promising alternative to traditional manual inspection methods. Nail abnormalities can serve as indicators of various health conditions, making their timely detection crucial for early intervention and treatment.

Pioneering research efforts have introduced a variety of innovative methodologies, each contributing valuable insights into the field. These efforts range from automated color analysis to complex machine learning-based classification systems, highlighting the versatility of digital approaches in diagnosing nail-related health issues. Despite the progress made, numerous challenges remain, such as the limitations of existing feature extraction methods and the need for more diverse datasets to improve model generalizability.

As the field evolves, it becomes increasingly important to address these challenges by refining the techniques used for feature extraction, segmentation, and classification. The integration of multimodal data and more sophisticated algorithms has the potential to further enhance diagnostic precision. By building on the foundation laid by earlier studies, current and future research can push the boundaries of what is possible in automated disease detection through nail image analysis.

2.1 Background

Below is a short background of each nail abnormality selected for this research.

Nail Pitting

Nail pitting is an abnormality in the nail when you have tiny dents in fingernails or toenails shown in Figure 1 and Figure 2. Nail pits refer to shallow indentations on the surface of the nail plate, which differ in their appearance and location. These depressions signify an imperfection in the outermost layer of the nail plate, originating from the proximal nail matrix.

Nail pitting can happen due to variety of reasons, such as

- Psoriasis: Nail pitting is a common symptom of psoriasis, an autoimmune condition that affects the skin and nails.
- Alopecia areata: This is another autoimmune condition that can cause hair loss and nail pitting.
- Eczema: skin condition that can cause nail pitting and other nail abnormalities.
- Connective tissue disorders: Certain connective tissue disorders, such as lupus or rheumatoid arthritis, can cause nail pitting.
- Infections: Fungal infections of the nails
- Trauma: Injury to the nail bed can cause nail pitting.

Figure 1. Toenail Pitting

Figure 2. Fingernail Pitting

Black Line

Black line is also known as 'Melanonychia' in medical terms. It appears as dark line or streak on the nail plate shown in Figure 3. The line usually appears on the nail plate and extends from the base of the nail to the tip. The most general reason to have Melanonychia nail abnormality is overproduction of melanin in the nail matrix. In the rare cases, it can be a sign of a type of skin cancer.

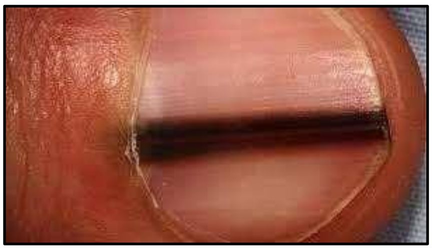


Figure 3. Black Line Abnormality

Splinter haemorrhages

Splinter haemorrhages is a nail abnormality that appears as thin, reddish brown or black vertical lines on the nails and resemble a splinter, hence the name. This can be seen on both finger and toenails. Most of the times this happens because of tiny blood clots which are bleeding under the nail as shown in Figure 4. However, this abnormality also happens due to some medical conditions.

Figure 4. Splinter haemorrhages

Nail Clubbing

Nail clubbing is a medical condition marked by the enlargement and rounding of the fingertips and nails leading to changes in their shape. As a result, the nails become thicker and the angle between the nail bed and nail plate increases as shown in the Figure 5. This condition is often associated with underlying medical conditions such as heart disease, liver disease, gastrointestinal disorders, and certain genetic conditions.

Figure 5.Nail Clubbing

Spoon Nails (Koilonychia)

Spoon nails develop due to changes in the nail's shape shown in Figure 6, resulting in a concave or spoon-like depression. The edges of the nails curve upward, resembling the shape of a spoon. This condition is often associated with iron deficiency and can also indicate underlying health issues such as heart disease, hypothyroidism (underactive thyroid), diabetes, and other systemic conditions.

Figure 6. Spoon Nails

The absence of a comprehensive, automated approach that addresses these abnormalities efficiently highlights the need for a novel diagnostic system. This system would not only improve accuracy and consistency but also facilitate early medical intervention, ultimately enhancing patient outcomes.

2.2 Contributions of Previous Work

The previous works in the field of nail image processing for disease detection have collectively contributed significantly to advancing diagnostic methods. These contributions can be grouped based on their focus areas, each addressing specific aspects of the broader diagnostic process.

Pandit, Bazar, and Shah (Pandit et al., 2013) made an initial breakthrough by developing an automated model to analyze nail colors, which significantly reduced the subjectivity and inconsistency associated with human visual assessments. Their approach was one of the first to demonstrate how automated analysis could surpass manual inspection in terms of accuracy, providing a foundation for more objective diagnostic tools. However, their model predominantly focused on color features, which limited its scope in detecting more complex nail abnormalities that may also involve texture and shape variations.

Mente and Marulkar (Mente, 2017) conducted a comprehensive review of various image processing techniques used in analyzing nail characteristics. Their work was instrumental in emphasizing the significance of integrating multiple features, such as color, texture, and pliability, for disease prediction. They highlighted the inefficiencies inherent in traditional manual inspection methods and showcased the potential of digital image processing to enhance diagnostic precision. Despite this valuable contribution, their study did not explore specific image processing techniques indepth, leaving room for future research to identify the most effective methods.

Meshram (Meshram, 2016) took a significant step forward by expanding the diagnostic scope to include both nail and palm images through the Automated Medical Palmistry System (AMPS). By integrating palm analysis with nail image processing, Meshram demonstrated that a multimodal approach could potentially yield more accurate diagnostic outcomes. This work broadened the applicability of nail image processing beyond standalone nail analysis, which could be highly beneficial for identifying systemic health conditions. However, the complexity of handling multiple image types presented challenges in ensuring consistency and processing efficiency.

Ansari and Vijay (Ansari and Vijay, 2023) explored the use of computer-aided disease detection techniques specifically through fingernail image processing, demonstrating that digital methods could achieve higher accuracy than traditional visual inspections. Their research demonstrated the utility of extracting and evaluating nail characteristics, particularly color, texture, and pliability, to predict various health issues. Nonetheless, their work focused primarily on improving accuracy without adequately addressing the computational efficiency of the algorithms, an aspect that could impact the practical usability of their system in real-time applications.

Maniyan and Shivakumar (Maniyan and Shivakumar, 2018 ; Rani et al., 2018) introduced and enhanced the Nail Image Processing System (NIPS) in two notable stages. Initially, they employed K-Nearest Neighbors (NIPS-K) to diagnose nail abnormalities based on extracted features like color, shape, and texture. Subsequently, they advanced their work by incorporating a Multiclass Support Vector Machine (NIPS-McS), which allowed the system to handle multiple disease classes simultaneously, thereby increasing its diagnostic capacity. Their work highlighted the importance of sophisticated classification algorithms in enhancing diagnostic accuracy. However, they did not conduct a comparative study of different shape feature extraction methods, leaving an important gap in understanding which methods could deliver the best results for specific abnormalities.

Overall, these contributions collectively represent a progression from basic feature extraction and automated color analysis to more complex classification systems that integrate multiple diagnostic parameters. The advancements made by each study underscore the versatility and potential of digital image processing in healthcare, laying the groundwork for further development of more sophisticated diagnostic tools that could ultimately improve the reliability and accessibility of healthcare services.

2.3 Significance of Related Work

The studies have made significant contributions to the field of nail image processing for disease detection, highlighting both the potential and challenges of applying digital image processing technologies to healthcare.

The introduction of automated systems, such as the one by Pandit, Bazar, and Shah (Pandit et al., 2013), significantly reduced the subjectivity associated with traditional manual inspections. By automating the analysis of nail images, these systems laid the foundation for more reliable diagnostic tools that can consistently outperform human visual assessments, especially in terms of accuracy. This breakthrough was pivotal in transitioning the field toward digital methodologies that minimize human error.

Mente and Marulkar (Mente, 2017) emphasized the importance of integrating multiple features—such as color, texture, and shape—into the diagnostic process. Their work demonstrated that a holistic approach to feature utilization could significantly improve diagnostic precision. By focusing on the diversity of features, they helped establish that relying solely on color analysis is insufficient for comprehensive disease detection. Their findings set a precedent for using multiple image features to enhance the depth and accuracy of disease predictions.

Meshram's Automated Medical Palmistry System (AMPS) (Meshram, 2016) introduced a novel approach by incorporating both nail and palm analysis. This multimodal system broadened the scope of nail image processing to potentially identify systemic health conditions that may be reflected in both palm and nail features. This was a significant step toward developing diagnostic systems that provide more comprehensive health assessments, leveraging a wider array of biometric indicators. The inclusion of multiple data types indicated a movement towards more sophisticated models capable of capturing nuanced health information from different parts of the body.

Maniyan and Shivakumar (Maniyan and Shivakumar, 2018 ; Rani et al., 2018) addressed a key challenge in nail image processing—diagnosing multiple disease types concurrently. They expanded the capabilities of nail image processing systems by first using K-Nearest Neighbors (NIPS-K) and later incorporating a Multiclass Support Vector Machine (NIPS-McS). This progression enabled the systems to handle more complex diagnostic scenarios, effectively increasing their applicability in real-world settings where patients may present with multiple overlapping conditions. Their work underscored the importance of robust classification methods that can manage the inherent variability in biological data.

These contributions collectively underscore the value of automation, diverse feature integration, multimodal approaches, and sophisticated classification algorithms in the field of nail image processing. They highlight the field's progression from basic automated analysis to more complex, integrated systems that can handle multiple conditions with greater precision. The pioneering research efforts have established a foundation for future work aimed at improving the diagnostic reliability, scope, and accuracy of nail image analysis tools.

2.4 Limitations

While the advancements in nail image processing for disease detection are promising, several limitations persist that hinder the full potential of these technologies.

One significant limitation is the Color Spectrum Limitation. Many of the existing models are restricted to identifying a limited range of nail colors, which can result in the oversight of subtle variations that are crucial for diagnosing specific diseases. This limitation underscores the need for more advanced color feature extraction methods capable of capturing a broader spectrum of colors and subtle distinctions between them.

Another key issue is the Accuracy Rates achieved by current models. For example, systems like the one developed by Rani (Rani et al., 2018) reached a 65% accuracy rate, which is not sufficient for reliable clinical applications. Higher accuracy is crucial to ensure that automated systems can be trusted as a primary or supportive diagnostic tool. The limited accuracy highlights the need for improved machine learning models, better feature selection, and more effective training processes.

Dataset Diversity is also a major concern. The performance of machine learning models heavily relies on the quality and diversity of their training datasets. Many of the existing studies have used datasets that lack diversity in terms of demographic information, including age, gender, and ethnicity. Expanding the datasets to encompass a wider range of demographics and a broader variety of nail diseases is vital for developing models that can generalize effectively to a larger population. Without diverse datasets, these models may fail to perform accurately across different patient groups, thereby limiting their practical usability.

The Shape Feature Extraction methods used in current research are another area that requires further attention. Detecting specific nail abnormalities, such as clubbing and spoon nails, relies on accurate shape feature extraction. However, existing studies have not extensively compared various shape feature extraction techniques to determine which is the most effective. This gap is particularly important because the precision of shape feature extraction can significantly impact the overall accuracy of the diagnostic system. Future work should focus on systematically evaluating different extraction methods to identify the most reliable approach for identifying complex nail conditions.

In addition, there are challenges associated with Computational Efficiency and Real-Time Application. While several studies have demonstrated promising accuracy rates, few have addressed the computational demands of these models. Real-time analysis is crucial for practical implementation in clinical settings. The lack of optimization in many current algorithms limits their usability, as high computational requirements can make real-time diagnostics impractical.

Future research should focus on integrating multi-modal data (e.g., combining nail analysis with other biometric indicators), enhancing feature extraction techniques, especially for shape features and employing more sophisticated machine learning algorithms to improve diagnostic accuracy and reliability. Additionally, establishing standardized image segmentation and feature extraction protocols can contribute to more consistent and accurate disease prediction models. By addressing these limitations, future developments can pave the way for more robust, accurate, and practical diagnostic tools in the field of nail image processing.

2.5 Relation to Current Research and Identified Gaps

The current research aims to bridge several of the critical gaps identified in previous studies by addressing specific limitations related to nail abnormality detection, feature extraction, and image segmentation. Despite numerous advancements in the field of nail image processing, there remains a clear gap in effectively detecting particular nail abnormalities, such as clubbing and spoon nails, using image processing techniques alone. Previous studies have primarily focused on color features and basic shape characteristics without thoroughly investigating the optimal techniques for extracting shape-related features, which are essential for accurately diagnosing more complex nail conditions.

Shree (Dhanashree et al., 2022) and Indi and Patil Indi and Patil, 2019) provided insights into using shape features such as perimeter, compactness, and eccentricity; however, their approaches lacked a comprehensive comparison of various shape feature extraction methods. The absence of a systematic evaluation of these methods means that the most effective shape feature extraction technique remains undetermined, thereby limiting the potential for accurate and reliable automated detection of specific nail abnormalities. This research directly addresses this gap by experimenting with multiple shape feature extraction techniques to determine the most effective approach, ensuring a more accurate classification of nail abnormalities like clubbing and spoon nails.

Additionally, the existing research has not sufficiently explored the integration of advanced machine learning models capable of handling multiple disease classes with a high degree of accuracy. While some studies, such as those by Maniyan and Shivakumar (Maniyan and Shivakumar, 2018 ; Rani et al., 2018), introduced systems that could classify multiple disease types, they did not delve into the comparative effectiveness of different machine learning algorithms for specific nail conditions. This current research aims to enhance classification accuracy by employing and comparing advanced machine learning techniques, thereby improving upon the existing methods for handling diverse nail conditions.

Image segmentation is another area that has not been fully explored for nail image processing. Image segmentation is crucial for accurately isolating the nail region from the rest of the image, which directly impacts the effectiveness of subsequent feature extraction and classification processes. Kaur et al. (Lee and Yang, 2018) and Muhammad (Khan, 2014) analyzed different segmentation techniques, highlighting the strengths and limitations of approaches like threshold segmentation, watershed transformation, and hybrid segmentation. However, no standard segmentation technique has been established that consistently provides optimal results for nail image scenarios. This research aims to address this gap by evaluating and optimizing segmentation techniques that are particularly suited for nail analysis. Improved segmentation will enhance the quality of extracted features, thereby contributing to more accurate disease classification.

Moreover, there is a need for a more holistic approach to feature integration. Many prior works have either focused solely on color or have inadequately combined different features without leveraging their collective potential. The current study aims to integrate multiple features—color, texture, and shape—alongside experimenting with different extraction methods to create a more comprehensive feature set. This holistic approach is expected to provide better diagnostic precision and address the shortcomings of feature selection in earlier research.

By systematically addressing these identified gaps, the current research intends to develop an automated system that not only improves upon previous models in terms of feature extraction and classification accuracy but also provides a standardized approach to detecting nail abnormalities. This research contributes to the field by enhancing the robustness, accuracy, and reliability of automated nail image analysis, ultimately paving the way for more effective early disease detection tools.

2.6 Explainable Artificial Intelligence (XAI) Methods

Artificial Intelligence (AI) and Deep Learning (DL) have transformed the landscape of machine learning by enabling the development of advanced solutions for complex problems across various domains such as medical imaging, natural language processing (NLP), autonomous systems and predictive analytics. These technologies have significantly improved decision-making capabilities, enhancing diagnostic accuracy in healthcare, streamlining financial risk assessments and optimizing security protocols. However, despite their remarkable performance, AI and DL models often function as "black boxes" which means their internal decision-making processes remain opaque and difficult to interpret. This lack of transparency raises critical concerns regarding the reliability, trustworthiness and ethical implications of AI-driven decisions, particularly in high-stakes applications such as healthcare, where misdiagnoses can have life-threatening consequences, finance where biased credit-scoring models can lead to unfair lending practices and security where AI-driven surveillance systems may infringe on privacy rights.

To address these challenges, Explainable Artificial Intelligence (XAI) has emerged as a vital research area aimed at making AI models more interpretable, transparent, and accountable. XAI techniques focus on elucidating how AI models arrive at specific decisions, providing insights into their reasoning processes. This is essential for fostering user trust, regulatory compliance and ethical AI deployment. By integrating methods such as feature attribution, model visualization and rule-based explanations, XAI enhances the interpretability of AI systems, enabling stakeholders including researchers, policymakers and end-users to better understand and validate AI-driven outcomes. As AI continues to evolve, the advancement of XAI will play a crucial role in bridging the gap between model performance and human interpretability while ensuring that AI systems operate in a fair, explainable, and responsible manner.

2.7 The Need for Explainability in AI

The black-box nature of AI and Deep Learning (DL) models arises from their inherently complex architectures which often consist of millions of parameters, intricate layer connections and non-linear transformations. These sophisticated structures enable models to achieve exceptional accuracy in tasks such as image recognition, speech processing and predictive analytics. However, this complexity also makes it challenging to decipher how these models reach their conclusions, limiting their widespread adoption in critical domains where transparency and interpretability are essential.

For instance, in medical imaging, clinicians must understand the rationale behind an AI model's classification of an image as malignant or benign to make informed treatment decisions and avoid potential misdiagnoses (Arrieta et al., 2020; Pawar et al., 2020). Similarly, in finance, credit scoring and fraud detection models require explainability to ensure fair decision-making and prevent biased outcomes that may disproportionately affect certain groups. In autonomous systems, understanding why an AI-powered vehicle makes specific navigation decisions is crucial for ensuring safety and accountability.

Beyond improving trust, explainability plays a crucial role in debugging AI models, detecting biases and ensuring compliance with ethical and legal standards. The importance of explainability is further underscored by regulatory frameworks such as the European Union's General Data Protection Regulation (GDPR) which grants individuals the right to receive explanations for automated decisions that impact them (Daglarli, 2020). Failure to provide such transparency can lead to legal and ethical challenges, potentially hindering AI adoption in regulated industries.

As a result, Explainable AI (XAI) has emerged as a pivotal area of research and industry focus, aiming to make AI systems more transparent, interpretable and accountable. Researchers and practitioners are actively developing methods such as feature attribution, saliency maps, rule-based models and surrogate explainability techniques to bridge the gap between model performance and human interpretability. By enhancing explainability, XAI fosters trust, facilitates regulatory compliance and enables AI systems to be deployed responsibly in high-stakes applications (Arrieta et al., 2020; Pawar et al., 2020; Daglarli, 2020).

2.8 Classification of XAI Methods

XAI methods can be broadly categorized into model-agnostic and model-specific approaches, each with its strengths and limitations (Gohel et al., 2021; Raghavan et al., 2023).

2.8.1 Model-Agnostic Methods

Model-agnostic methods are versatile techniques that can be applied to any machine learning model, regardless of its underlying architecture. These methods analyze the relationship between input features and model predictions without requiring direct access to the model's internal parameters or structure. By offering interpretability without altering the model, they are widely used in domains such as healthcare, finance and automated decision-making systems. Common model-agnostic methods include:

- **LIME (Local Interpretable Model-agnostic Explanations) (Gohel et al., 2021):** LIME generates local explanations by perturbing the input data and observing the resulting changes in the model's predictions. It then approximates the model's behavior in a localized region using a simpler, interpretable model, such as linear regression or decision trees. This approach allows for better human interpretability by providing feature importance values. However, LIME's effectiveness heavily depends on the choice of the surrogate model and the quality of the perturbations, making it sensitive to parameter tuning. Additionally, LIME is more reliable for explaining individual predictions rather than providing global interpretability for the entire model.

- **SHAP (SHapley Additive exPlanations) (Gohel et al., 2021):** SHAP is based on cooperative game theory and assigns each feature a **Shapley value**, which represents its contribution to the model's output. Unlike LIME, SHAP provides globally consistent and theoretically grounded explanations, making it a robust interpretability tool. It helps in understanding feature importance across multiple predictions while ensuring fairness and transparency in AI-driven decisions. However, SHAP can be computationally expensive, particularly for models with a large number of features or deep architectures, as it requires evaluating multiple permutations of input values to compute precise Shapley values.

- **Perturbation-Based Methods:** These methods systematically modify the input data and analyze how the model's predictions change, helping to identify which parts of the input are most influential in decision-making. One example is **RISE**

(Randomized Input Sampling for Explanation) (Petsiuk et al., 2018) which generates saliency maps by randomly occluding portions of the input image and measuring the effect on model predictions. Perturbation-based methods are particularly useful in computer vision tasks where visual interpretability is needed. However, they struggle with precise localization of important features and can introduce noise if the perturbations are not carefully controlled.

By providing post-hoc explanations without modifying the original model, these model-agnostic methods enhance the interpretability of AI-driven systems, making them more transparent and trustworthy in real-world applications.

2.8.2 Model-Specific Methods

Model-specific methods are designed for particular model architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) and transformers. These techniques leverage the model's internal structure such as activation layers and weight distributions to generate more accurate and computationally efficient explanations. Unlike model-agnostic approaches, model-specific methods provide deeper insights into the decision-making process by utilizing internal gradients, attention mechanisms and neuron activations. Common model-specific interpretability methods include:

- **Grad-CAM (Gradient-weighted Class Activation Mapping) (Raghavan et al., 2023):** Grad-CAM generates heatmaps that highlight the most influential regions in an input image by computing the gradients of the predicted class with respect to the feature maps of the last convolutional layer. These heatmaps provide visual explanations, making Grad-CAM a widely used tool in medical imaging, object detection and facial recognition. However, while it effectively identifies relevant regions, it struggles with fine-grained localization, particularly in cases where multiple objects overlap or where subtle patterns contribute to the prediction.

- **Ablation-CAM (Ablation-based Class Activation Mapping) (Desai and Ramaswamy, 2020):** Ablation-CAM improves upon Grad-CAM by systematically removing or masking portions of the model's feature maps to evaluate their impact on the final prediction. Instead of relying on gradient-based calculations, it measures the contribution of individual feature maps more directly, leading to more precise and robust explanations. This method is particularly useful for deep CNN architectures but remains limited to specific types of neural networks by making it less generalizable compared to Grad-CAM.

- **Layer-wise Relevance Propagation (LRP):** LRP redistributes the model's output back to its input features by propagating relevance scores layer by layer through the network. This method is particularly effective for understanding how individual neurons and features contribute to the final prediction in deep neural networks. LRP has been widely used in finance, bioinformatics, and NLP applications where precise feature attribution is required. However, its complexity increases with deeper architectures, making computational efficiency a challenge in very large models.

Model-specific methods provide more detailed and architecture-aware explanations than model-agnostic techniques, making them essential for applications that demand high interpretability, such as medical AI, self-driving cars, and fraud detection systems. By leveraging internal model structures, these methods enhance trust in AI systems and facilitate better debugging, model refinement, and regulatory compliance.

2.9 Image-Based XAI Methods

Image-based XAI methods are essential for interpreting deep learning models in computer vision tasks. These methods can be categorized into attribution-based and non-attribution-based techniques (Zamzmi et al., 2023).

2.9.1 Attribution-based Methods

Attribution-based methods generate visual explanations by identifying and highlighting the regions of an image that contribute most to the model's prediction. These methods help in understanding which features are influential in decision-making and making them essential for applications in medical imaging, object detection and autonomous systems. Attribution-based techniques are particularly useful for explaining convolutional neural networks (CNNs) and deep learning models applied to visual tasks. They can be broadly categorized into the following groups:

Gradient-Based Methods: These techniques utilize gradient information to determine which input features most strongly influence the model's output. Grad-CAM (Gradient-weighted Class Activation Mapping) and Integrated Gradients (IG) are widely used approaches in this category.

- **Grad-CAM** computes the gradients of the predicted class with respect to the feature maps in the last convolutional layer, generating saliency maps that highlight the most important regions of an image. It is commonly applied in medical imaging for tumor detection and in self-driving cars for object recognition. However, Grad-CAM can struggle with fine-grained localization and may produce misleading explanations if the gradients are noisy.
- **Integrated Gradients (IG)** addresses some of these limitations by accumulating gradient information across multiple interpolated inputs, offering a more stable and theoretically grounded approach. IG is particularly useful in scenarios where precise feature importance is needed, such as medical diagnosis and financial decision-making.

Perturbation-Based Methods: These methods systematically modify input data and observe changes in model predictions to infer feature importance. Common approaches include LIME (Local Interpretable Model-agnostic Explanations) and RISE (Randomized Input Sampling for Explanation).

- **LIME** perturbs input samples and trains a surrogate interpretable model (e.g., a linear regression) to approximate the model's local decision-making behavior. While it is effective for structured and tabular data, it may struggle with high-dimensional image data where feature dependencies are complex.

- **RISE** generates saliency maps by randomly occluding parts of the image and measuring the effect on model predictions. It is robust to noisy gradients, making it useful for applications such as face recognition and anomaly detection in surveillance systems. However, it may not always precisely localize important features, especially in cluttered images.

Layer-wise Relevance Propagation (LRP): LRP redistributes relevance scores across the model's layers to attribute the final output back to the input features. This method is particularly effective for understanding the hierarchical feature contributions in deep neural networks. LRP is widely used in medical diagnostics, financial modeling, and explainable AI applications, where precise feature attribution is required. However, as model complexity increases, LRP computations become more challenging, requiring careful implementation to maintain accuracy.

Attribution-based methods bridge the gap between model predictions and human interpretability, enabling trustworthy AI applications in critical domains. While gradient-based methods offer efficient and scalable explanations, perturbation-based methods provide model-agnostic and robust interpretations. Combining multiple attribution techniques can enhance the reliability of AI-driven decisions, ensuring greater transparency and accountability.

2.9.2 Non-Attribution Methods

Non-attribution methods focus on understanding a model's decision-making process by analyzing high-level concepts, prototypes or counterfactual explanations, rather than directly attributing importance to input features. These methods offer a more human-interpretable perspective, making AI models more transparent and relatable, especially in domains where stakeholders require meaningful explanations rather than raw feature attributions. Non-attribution methods can be broadly categorized into:

Concept-Based Methods: These techniques link model predictions to high-level,human-understandable concepts rather than low-level features such as pixel values or numerical attributes. A prominent approach is Testing with Concept Activation Vectors (TCAVs) which quantifies the influence of a predefined concept (e.g., "striped texture" in an image classifier) on the model's predictions.

- **TCAV** is widely used in fields like medical AI, where researchers analyze whether a deep learning model relies on known pathological markers (e.g., "cell irregularity" for cancer detection). It is also useful in bias detection, helping assess whether a model associates certain gender or racial features with predictions in applications like hiring algorithms.

Counterfactual-Based Methods: Counterfactual explanations describe how an input needs to be minimally changed for the model to produce a different prediction. By identifying these small alterations, counterfactual methods provide actionable insights into model decisions.

- **Guided by Prototypes** and the Contrastive Explanations Method (CEM) generate perturbed versions of the input to determine the critical factors influencing the prediction.

- For instance, in **loan approval models**, a counterfactual explanation might suggest, "If the applicant had an income of $50,000 instead of $45,000, the loan would have been approved." These methods are valuable in finance, healthcare, and legal applications, where transparency and fairness are crucial.

Prototype-Based Methods: Prototype-based methods explain decisions by presenting representative examples (prototypes) that closely resemble the input, allowing humans to compare how similar cases were classified.

- **Prototypical Part Network (ProtoPNet)** identifies key parts of an input image and associates them with known examples, making it particularly useful in medical diagnosis (e.g., matching a tumor to previously diagnosed cases).

- **Maximum Mean Discrepancy Critic (MMD-Critic)** selects diverse prototypes from the dataset to serve as reference cases, improving the interpretability of AI models used in image classification and anomaly detection.

- These methods are commonly applied in **computer vision, fraud detection, and product recommendation systems**, where users benefit from understanding how a model relates new inputs to previously seen data.

By leveraging concepts, counterfactual reasoning, and prototypes, non-attribution methods offer interpretable explanations that align with human intuition. They are particularly valuable in applications where users need meaningful justifications for AI decisions, enhancing trust, fairness, and transparency in machine learning systems.

2.10 Evaluation of XAI Methods

The effectiveness of Explainable AI (XAI) methods is typically evaluated using both qualitative and quantitative metrics to ensure that explanations are interpretable, reliable, and applicable across various domains. These evaluation techniques help determine whether an XAI method successfully conveys the reasoning behind a model's decision-making process and whether it can be trusted in high-stakes applications.

Qualitative evaluation primarily involves the visualization of explanations, such as saliency maps and heatmaps, to assess interpretability. These visualizations allow domain experts to determine if the AI model is focusing on meaningful regions within the input data. For instance, in medical imaging, heatmaps help highlight critical areas in X-ray or MRI scans, allowing radiologists to assess whether the model's predictions align with their clinical knowledge (Lin et al., 2021). Another form of qualitative evaluation includes feature importance plots, which illustrate how different input variables influence the model's predictions, helping stakeholders understand model behavior in applications such as finance and healthcare (Lin et al., 2021). Additionally, user studies and expert feedback are commonly used to gauge how effectively AI explanations improve human decision-making, particularly in scenarios where AI-generated insights assist in complex reasoning processes (Lin et al., 2021).

Quantitative evaluation provides objective and measurable metrics to assess the reliability and robustness of XAI methods. One commonly used metric is Intersection over Union (IoU) (Lin et al., 2021), which measures the overlap between the predicted bounding box (highlighted by the XAI method) and the actual region of interest. A higher IoU score indicates a better alignment between the explanation provided by the XAI method and the true causal features influencing the model's prediction. IoU is widely applied in medical diagnosis, autonomous driving, and security systems, where precise feature localization is critical for trust and accuracy (Lin et al., 2021).

Beyond localization, other quantitative metrics include robustness to adversarial attacks, which evaluates whether slight modifications to the input data significantly alter the generated explanations. This metric is crucial in security-sensitive applications such as fraud detection, biometric authentication, and cybersecurity, where adversarial robustness ensures model reliability (Lin et al., 2021). Another important factor is computational efficiency, which assesses the time and resource consumption of generating explanations. Some XAI techniques, such as SHAP, offer highly detailed attributions but require significant computational power, whereas others, such as LIME, provide faster but potentially less stable explanations (Lin et al., 2021).

Furthermore, the faithfulness of an explanation is evaluated to determine how well an XAI method reflects the actual decision-making process of the AI model rather than providing a simplified or misleading approximation. Techniques such as perturbation analysis and counterfactual reasoning help assess whether explanations remain valid across various input conditions (Lin et al., 2021). Lastly, user satisfaction studies play a critical role in determining the practical utility of XAI in real-world applications. Surveys and usability testing help assess whether AI-generated explanations are clear, actionable, and trustworthy for end-users, such as medical professionals, financial analysts, or policymakers, ensuring that AI systems align with human expectations (Lin et al., 2021).

By integrating both qualitative and quantitative evaluation techniques, researchers and practitioners can ensure that XAI methods provide transparent, trustworthy, and interpretable insights, fostering greater acceptance and ethical deployment of AI systems in high-stakes environments.

2.11 Applications of XAI in Medical Imaging

XAI methods have been extensively applied in medical imaging to enhance the interpretability of deep learning models, enabling clinicians and researchers to better understand AI-driven diagnostic decisions. These techniques provide insights into how models classify medical images, ensuring transparency and reliability in critical healthcare applications.

For instance, (Imouokhome et al., 2023) demonstrated the effectiveness of XAI techniques such as Integrated Gradient, GradientShap, and Occlusion in interpreting classification results in medical imaging tasks. These methods help highlight the most influential features in medical scans, allowing radiologists to verify whether an AI model focuses on clinically relevant regions. By generating saliency maps and feature attributions, these techniques improve trust in AI-based diagnostic tools, particularly in tasks such as tumor detection, pneumonia classification, and diabetic retinopathy screening.

Similarly, (Kashefi et al., 2024) explored explainability methods for visual transformers, emphasizing the importance of attention mechanisms in understanding model decisions. The study demonstrated how transformer-based architectures, commonly used in medical imaging analysis, allocate attention across different image regions, aiding in tasks such as lesion detection and organ segmentation. By visualizing the attention weights, researchers can assess whether the model correctly prioritizes medically significant features, thus improving interpretability and clinical acceptance.

Additionally, (Cantone et al., 2023) highlighted the role of XAI methods such as Class Activation Mapping (CAM), Gradient-weighted Class Activation Mapping (Grad-CAM), and Grad-CAM++ in explaining CNN-based medical image classification. These methods provide heatmaps that localize critical regions in images, making them particularly useful in radiology, pathology, and dermatology applications. For example, Grad-CAM has been widely used to interpret deep learning models for skin cancer classification, lung disease detection, and brain

MRI analysis by clearly indicating which image regions contributed most to the model's prediction.

2.12 Challenges in Future Directions of XAI

Despite significant advancements, several challenges persist in the field of Explainable Artificial Intelligence (XAI), limiting its widespread adoption in high-stakes applications such as healthcare, finance, and autonomous systems. Addressing these challenges is crucial to improving the interpretability, reliability, and efficiency of AI-driven decision-making processes.

One of the key challenges is accurate localization, as many XAI methods struggle to precisely identify the most relevant features in complex tasks. Techniques such as saliency maps and Grad-CAM often highlight broad regions rather than pinpointing specific areas responsible for a model's decision. This limitation is particularly problematic in applications like medical imaging, where precise feature localization is essential for diagnostic accuracy.

Another major issue is computational efficiency, as some XAI methods require significant computational resources. For instance, SHAP generates highly detailed explanations but is computationally expensive, making it challenging to apply to large datasets or real-time AI systems. The high processing time required for certain methods can limit their practical use, especially in environments where quick decision-making is critical, such as autonomous driving or emergency medical diagnosis.

A further challenge lies in user-centric explanations, as many current XAI techniques generate outputs that are difficult for non-experts to interpret. While methods such as LIME and SHAP provide feature attributions, they often lack intuitive and actionable insights for end-users, such as doctors, financial analysts, or policymakers. Future research should focus on developing explanations that are not only technically accurate but also accessible and meaningful for diverse user groups.

Emerging approaches, such as SIDU (Similarity Difference and Uniqueness) (Muddamsetty et al., 2020), aim to address these challenges by providing more accurate and reliable explanations. By improving feature attribution techniques, enhancing model transparency, and reducing computational costs, SIDU and similar methods contribute to making AI systems more interpretable and trustworthy.

As the field of XAI evolves, ongoing research and development will continue to enhance the usability and effectiveness of AI systems. Techniques such as LIME, SHAP, Grad-CAM, and saliency maps have already advanced the state of the art, offering valuable insights into model decision-making. However, overcoming challenges related to localization accuracy, computational demands, and user-friendliness remains essential for the future of XAI. By addressing these issues, researchers and practitioners can ensure that AI models become more transparent, interpretable, and applicable across various critical domains, ultimately fostering greater trust and adoption of AI-driven technologies.

2.13 Comparative Analysis

Table 1 depicts comprehensive analysis between related similar approaches conducted so far and the proposed approach as below.

Table 1. Comparative Analysis

Feature/Methodology	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**	**[11]**	**[12]**	**[13]**	**[14]**	**[15]**	**[16]**	**[17]**	**[18]**	**Proposed Approach (NailHealth AI)**
Feature Extraction Comparison	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes – Comprehensive comparison of shape feature extraction techniques, including invariant moments, gradient-based descriptors, and contour-based methods.
Segmentation Technique	No	Yes	Yes	No	No	Yes	Yes	Yes	No	No	Yes	No	Yes	No	No	YOLOv8 and YOLOv9 – Utilizes advanced YOLO models for precise instance segmentation and anomaly detection.

Machine Learning Model Used	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Hybrid approach – Combines Vision Transformers (ViT) and Convolutional Neural Networks (CNNs) with ensemble techniques for enhanced accuracy.
Hybrid Model Approach	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes – Integrates CNN and ViT architectures to leverage both local and global feature extraction capabilities.

Synthetic Data Generation	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes – Implements synthetic data generation using fine-tuned Stable Diffusion models (SD 1.5, SDXL, SD2, SD3) with Low-Rank Adaptation (LoRA) and Dreambooth techniques to augment real clinical datasets.
Explainable AI (XAI)	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes – Generates visual saliency maps for CNNs and attention maps for ViTs to provide transparency and enhance user trust.

User Interaction and Engagement	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes – Features an OpenAI-based conversational agent for user queries and generates comprehensive, user-friendly reports.
Severity Assessment	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes – Utilizes YOLO models to quantify disease severity through severity scores and visual highlights of anomalies.
Detailed Reporting with Recommendations	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes – Generates detailed reports that include diagnoses, visual evidence, severity assessments, and medical recommendations.

Handling Dataset Limitations	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes – Combines real and synthetic data to enhance dataset diversity, address class imbalances, and improve model generalization.
Advanced Image Analysis Techniques	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes – Leverages Vision Transformers (ViT), CNNs (e.g., ResNet, Inception, EfficientNet), and YOLOv8/YOLOv9 for sophisticated image analysis and anomaly detection.

CHAPTER 3: METHODOLOGY

The Nail-Scan Tool is an advanced AI-driven system developed to enhance the detection and diagnosis of nail diseases through sophisticated computer vision techniques. This chapter outlines the research methodology used in designing, developing, and evaluating the system, detailing the underlying models, dataset preparation, training procedures, and performance evaluation. The Nail-Scan Tool consists of six key components, each playing a distinct role in the diagnostic pipeline: the Nail Filter Model, CNN Nail Diagnosis Model, ViT Nail Diagnosis Model, Nail Abnormality Detection and Segmentation Model, and the XAI Visual Saliency Algorithm. By employing a multi-model architecture, this system integrates multiple deep learning approaches to overcome individual model limitations, ensuring improved diagnostic accuracy, robustness, and interpretability.

3.1 Overview

The following sections provide a comprehensive overview of each component, describing their objectives, architectures, data processing techniques, training methodologies, and evaluation metrics. The overview of the key components is as below.

I. **Nail Filter Model:** The proposed approach process begins with the Nail Filter Model, which optimizes the diagnostic efficiency by distinguishing relevant nail images from irrelevant or poor-quality inputs. This ensures that only valid and high-quality images proceed to the subsequent stages of analysis.

II. **VIT Nail Diagnosis Model (Verification Model):** This model functions as a secondary layer of analysis, refining the diagnostic accuracy by verifying the initial predictions made by the Nail Diagnosis Model. It validates and confirm the presence of specific clinical conditions for improved reliability.

III. **CNN Nail Diagnosis Model:** The core of the system is the Nail Diagnosis Model which is a deep learning-based tool designed to categorize images into six distinct classes. These include five specific nail disease conditions and a category for normal findings. This model plays a key role in enabling early detection and classification, offering essential support for timely medical intervention.

IV. **Dual-Stage Nail Diagnosis & Verification Algorithm:** The algorithm integrates two stages: a Convolutional Neural Network (CNN) model for initial diagnosis and a Vision Transformer (VIT) model for verification. This approach enhances diagnostic accuracy by refining predictions and confirming the presence of nail conditions by providing a more reliable solution for early detection.

V. **Nail Abnormality Detection and Segmentation Model:** This model specializes in detecting and segmenting nail abnormalities such as black lines and splinters. By focusing on these specific features, the model provides detailed insights into subtle and potentially critical nail health indicators.

VI. **Visual Saliency Algorithm:** The final component namely the **Visual Saliency Algorithm** implements **Explainable AI (XAI)** by generating visual explanations of diagnostic outputs. It highlights salient regions in input images, enhancing transparency, interpretability, and user trust in the system's results.

3.2 Multi-model Architecture

The proposed solution adopts a multi-model architecture which ensures the limitations of individual models are mitigated by leveraging complementary strengths. This collaborative design enhances diagnostic precision, robustness, and overall system reliability. Combining CNN-based models with Vision Transformers (ViTs) for diagnosing nail diseases creates a robust and versatile diagnostic framework by leveraging the complementary strengths of both architectures. CNNs excel at extracting localized features such as texture, color variations, and fine-grained patterns through their convolutional layers, which are critical for detecting subtle abnormalities like fungal infections, ridges, or micro-hemorrhages. In contrast, ViTs utilize self-attention mechanisms to analyze global relationships across the entire nail image, capturing structural patterns such as widespread discoloration, deformation, or symmetry loss that might span large regions. By integrating these approaches, the hybrid model achieves multi-scale analysis where CNNs focus on pixel-level details and spatial hierarchies (e.g., edges → textures → shapes), while ViTs contextualize these features within the broader nail anatomy by enhancing diagnostic accuracy. This synergy also improves robustness, as CNNs inherently handle spatial invariance (e.g., rotation, scaling), and ViTs adapt to diverse imaging conditions (e.g., lighting, angles). The fusion of local and global perspectives mimics clinical workflows, where dermatologists inspect both localized lesions and overall nail health, reducing false positives/negatives. Additionally, pre-trained CNNs and ViTs transfer knowledge from large datasets, compensating for limited medical data and improving generalization. The hybrid model's interpretability is further strengthened by ViT's attention maps, which highlight clinically relevant regions to guide CNN focus. Overall, this combined approach offers a data-efficient, reliable solution for nail disease diagnosis, bridging the gap between detailed feature extraction and holistic context understanding in medical imaging.

Figure 7 is a high-level architecture diagram of the proposed system as shown in the below.

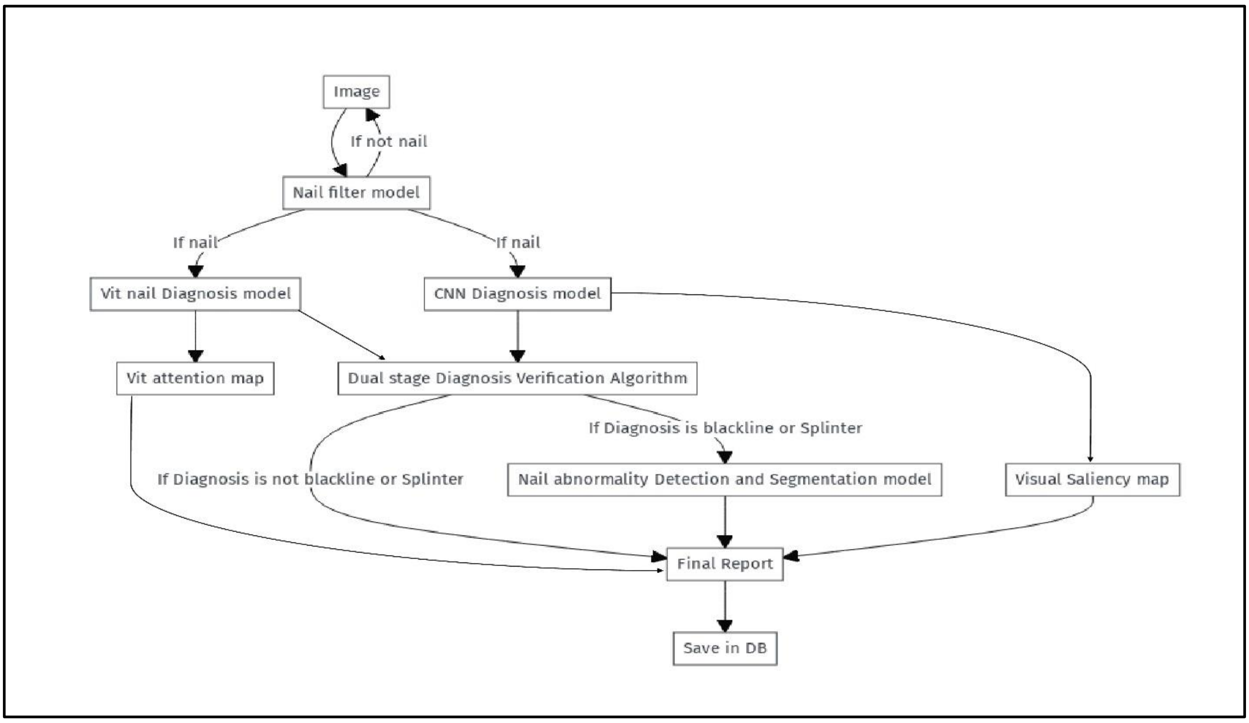


Figure 7. High-level Architecture Diagram

3.3 Methodology Analysis

In this section, a detailed analysis of the six main components mentioned above will be presented. Each component will be examined comprehensively by covering its methodology, including the model's objective, dataset preparation, architecture, training process, and performance evaluation.

3.3.1 Nail Filter Model

The Nail Filter Model is a foundational component of the nail disease diagnostic pipeline, designed to preprocess and filter out images that do not contain nails.

Objective

The primary objective of the Nail Filter Model is to act as an automated gatekeeper by ensuring that only images containing nails are passed on to subsequent stages of the diagnostic process. Non-nail images, such as those of objects, landscapes, or other body parts, are filtered out to prevent unnecessary computational load and to streamline the workflow. By focusing solely on nail images, the model enhances the efficiency and accuracy of the diagnostic system, allowing it to allocate resources effectively and reduce the risk of false positives or irrelevant analyses. This preprocessing step is critical in real-world applications, where input data may include a mix of relevant and irrelevant images.

Dataset Preparation

The development of the Nail Filter Model relied on a carefully curated dataset to ensure robust training and generalization. The dataset was divided into two main categories called nail images and non-nail images.

I. **Nail Images**:

- **Number of Images**: 1500
- **Details**: The nail images encompassed a wide range of conditions, including healthy nails, nails with discoloration, deformities, fungal infections, and other abnormalities. This diversity was essential to ensure the model could recognize nails across various presentations, from subtle texture changes to significant structural deformations. The inclusion of both common and rare nail conditions ensured the model's adaptability to real-world scenarios.

II. **Non-Nail Images**:

- **Number of Images**: 1500
- **Details**: The non-nail images included a diverse collection of objects, landscapes, and other body parts. This variety was crucial to train the model to distinguish nail textures from non-nail textures, such as skin, hair, or everyday objects. By exposing the model to a broad spectrum of non-nail images, it learned to identify nails with high precision, even in the presence of visually similar patterns.

The balanced dataset ensured that the model was not biased towards either category by enabling it to perform effectively in distinguishing nails from non-nails.

Model Architecture

The Nail Filter Model was built using a modified version of the Vision Transformer (ViT) architecture. ViT was chosen for its ability to capture global patterns and longrange dependencies in images through self-attention mechanisms which are particularly useful for identifying complex structures like nails. In order to tailor the ViT architecture for the specific task of nail image identification, the last two layers of the original model were removed. This simplification reduced the model's complexity while retaining its powerful pattern recognition capabilities, making it more efficient for the preprocessing task.

The modified ViT architecture processes input images by dividing them into patches, which are then flattened and fed into the transformer encoder. The self-attention mechanism allows the model to focus on relevant regions of the image, enabling it to distinguish nails from non-nails with high accuracy. This architecture was particularly effective in handling the diverse textures and patterns present in nail images.

Training Mechanism

The training process was designed to ensure the model's robustness and generalization across various imaging conditions. The following steps were undertaken:

I. **Data Augmentation**:

To enhance the model's ability to handle variations in nail images, extensive data augmentation was performed using the ImageDataGenerator from Keras. The augmentation techniques included:

- **Rotation and height shifts** to introduce positional variance, simulating different angles and orientations of nail images.
- **Shear and zoom** to simulate changes in shape and perspective, ensuring the model could recognize nails even when distorted.
- **Fill mode** to handle pixel filling after transformations, maintaining image integrity.
- **Brightness adjustments** to account for varying illumination conditions, which are common in real-world imaging.
- **Vertical and horizontal flips** to eliminate directional biases, ensuring the model could recognize nails regardless of their orientation.

II. **Training Configuration**:

The model was trained over 100 epochs with an early stopping patience value of 15 to ensure thorough learning and convergence. To achieve stable and efficient optimization, multiple training runs were conducted using the Adam, Adadelta, and RMSprop optimizers with learning rates of 0.01, 0.001, and 0.0001. Training was performed on a Tesla P100 GPU with 16GB of memory, allowing efficient handling of the large dataset and computational demands. The combination of data augmentation and optimized training parameters enabled the model to generalize well to unseen data, making it robust and reliable for real-world applications.

3.3.2 ViT Nail Diagnosis Model

The **Vision Transformer (ViT) Nail Diagnosis Model** is an advanced diagnostic system capable of detecting and classifying nail disorders based on image analysis. This model focuses on the assessment of nail images by classifying them into normal (healthy nail) or any of the five **pitting, black line, splinter, clubbing, or spooning** nail pathologies. Its goal is to support the detection of nail disorders at the earliest stage possible, which aids in prompt treatment and improves patients' prognosis. The ViT model, which was initially created to perform natural language processing (NLP) tasks, has been modified to perform tasks within the domain of computer vision. The model's treatment of image patches as order sequences enables the ViT transform to capture the spatial relations and structures within the images of nails, which makes this model highly effective for this diagnostic purpose.

Objective

The Nail Diagnosis Model serves a specific purpose which is to classify nails images into one of the six categories. Capturing high-definition images of the nails, processing them to identify relevant feature, identifying visual patterns linked to certain nail disorders, and classifying the image as either normal or having one of the five disorders are all steps involved in classification. It is essential for the model to achieve a high level of accuracy for these classifications because it is intended to function as an automated diagnosis system in clinical and dermatology practices. With accomplish this aim, it is expected that the Model will be a dependable and cost-effective system for nail disorder diagnosis.

Role of the Model

The ViT Nail Diagnosis Model acts as the principal diagnostic model in the system. Its major functions involve diagnosing nail issues as early as possible to simplify treatment and lessen the manual work performed by dermatologists, saving them both effort and time. Furthermore, the model allows for the mass evaluation of numerous patients' nail conditions in rural or neglected regions. The model enables healthcare professionals to make more educated decisions by giving useful predictions, thus enhancing the quality of care provided to patients. The automated analysis will benefits users clinically and academically.

Dataset Preparation and Preprocessing

For both training and evaluating the ViT Nail Diagnosis Model, the dataset consists of well-captured images of nails with each labeled as ‘normal’ or one of the five nail conditions. The dataset was collected from many places including clinical repositories and skin public databases to integrate diversity in images. The model to be more general and robust, an **Automated Random Image Augmentation** was applied during the training. This involves creating different versions of each image so that it transforms in a real-world manner. These transformations were the result of many model performance trials and error for model optimization.

Automated Random Image Augmentation

The **Automated Random Image Augmentation** process for a given task randomly selects and applies a specified set of transformations to each image in the dataset. These transformations include rescaling, rotation, height shift, shear, zoom, brightness adjustment and flipping as summarized in Table 2. Each image (both original and augmented) is transformed using a randomly selected subset of these transformations. This approach ensures that the model is exposed to a wide range of image variations during training and improving its ability to generalize to unseen data. The transformations were carefully selected through extensive training and testing of the underlying computer vision model which is resulting in the best model performance across multiple variations and settings.

Table 2. Automated Random Image Augmentation Parameters

Transformations	Value range
rescale	1./255
Rotation range	360
Height shift range	0.1
shear range	0.1
zoom range	0.1
fill mode	'nearest'
brightness range	[0.9, 1.2]
vertical flip	True
horizontal flip	True
Image size	299
Batch size	16
Color mode	'rgb'
Class mode	'categorical'
Shuffle	True

Model Architecture

The ViT Nail Diagnosis Model was developed through a systematic process. After evaluating several state-of-the-art computer vision models on the nail dataset, the **Vision Transformer (ViT)** was selected as the foundational model due to its high performance in capturing spatial relationships and patterns in images. The standard ViT model was further refined by modifying specific layers to suit the nail diagnosis task appropriately. These modifications were based on extensive testing and validation on the dataset. The final model architecture includes a **Patch Embedding Layer** to convert image patches into embeddings, a **Transformer Encoder** to capture spatial relationships through self-attention mechanisms and a **Classification Head** to output probabilities for each of the six classes.

Training Methodology

The training methodology for the ViT Nail Diagnosis Model was designed to optimize performance for the multi-class classification task. As loss function, categorical cross-entropy was used and the splits of train (80%), validation (10%), and test (10%) sets were applied. The initial layers of the model were set to utilize ViT weights to prespecify a large portion of the features with pre-trained values and thus speed up the training process. The model was trained for 100 epochs. Furthermore this experiment was conducted in two phases where the first phase focused on optimizing the learning rate (trying 0.1, 0.01, and 0.001) and the second phase aimed to select the best optimizer out of Adam, AdaDelta, and RMSProp.

Evaluation Metrics

The performance of the model was evaluated using a set of standard metrics to ensure its accuracy and reliability. **Accuracy** was used to measure the proportion of correctly classified images, while **precision, recall, and F1-score** were calculated to assess the model's performance for each class. A **confusion matrix** was also generated to provide a detailed breakdown of predictions against actual labels. These metrics were applied to both the validation and test sets to ensure the generalizability and robustness of the model. Finally, the evaluation process confirmed the effectiveness and the productivity of the mode by classifying nail images into the six predefined classes.

3.3.3 CNN Nail Diagnosis Model (Verification Model)

The **CNN Nail Diagnosis Model** employs the most advanced technology available for the image diagnosis and classification of different nail disorders. As mentioned in the previous approach, the model categorizes nail images into six distinct classes as normal (healthy nail) or one of five specific nail conditions namely pitting, black line, splinter, clubbing, and spooning. The goal remains to facilitate early detection, ensuring timely medical intervention and improving patient outcomes.

A variety of deep learning models were considered in this approach. Instead of the **Vision Transformer (ViT)** used previously, this model incorporates with several well-established convolutional neural network (CNN)architectures such as ResNet50, ResNet101, VGG16, VGG19, MobileNetV2, DenseNet121 and InceptionResNetV2. All these models are known for their high performance in image classification tasks and the capability to capture intricate spatial patterns in nail images which are crucial for diagnosing conditions at the early stage.

Objective

As mentioned in the previous approach, the primary objective of this model is to accurately classify nail images into one of six predefined classes. This process involves extracting relevant features from high-resolution images, identifying subtle visual patterns linked to specific nail conditions and assigning each image to the correct category (normal or one of the five conditions). Achieving high accuracy in these predictions is critical to ensure that the model performs effectively in clinical and dermatological applications. Moreover, the model aims to offer a reliable solution for diagnosing nail conditions at scale by meeting this goal.

Role of the model

The model operates similarly to the ViT approach. Its function is to operate as a secondary diagnostic imaging technique for early detection of nail diseases that is essential for treatment decisions. Furthermore, this model improves efficiency and saves time in clinical practice by lessening the burden of complex manual examinations of the nails by dermatologists. Additionally, its scalability enables large-scale screening of nail conditions, especially in remote or underserved areas, where the access to dermatologists is limited. Overall, this model enables accurate forecasting which empowers healthcare providers to take better decisions and delivering timely care to patients**.**

Dataset and Preprocessing

The dataset for this model is same the one used before, containing a collection of well-labeled images of nails as normal or any of the five nail conditions. It was collected from clinical databases and dermatology public repositories for better diversity in terms of image type. The dataset underwent **Automated Random Image Augmentation** which was done as a part of the training to enhance the robustness and generalization of the model. As explained in the previous methodology, the augmentations were made carefully in order to increase the performance of the model by exposing it to a wide range of image variations

As mentioned in the above approach, the **Automated Random Image Augmentation** process randomly applies several transformations to each image during the training process. These transformations (such as rescaling, rotation, and brightness adjustments) ensure that the model is exposed to diverse image variations while enhancing its generalization to unseen data. Moreover, the **ImageDataGenerator** library from **Keras** was again used to apply these transformations by ensuring that the model remains robust under real-world conditions.

Model Architecture

Instead of using the Vision Transformer, this model is built upon a combination of well-established CNN architectures including ResNet50, ResNet101, VGG16, VGG19, MobileNetV2, DenseNet121 and InceptionResNetV2. Each model was properly tested and adapted for the nail diagnosis task. The final architecture consists of a feature extraction layer based on the selected CNN models followed by a classification head that outputs probabilities for the six nail conditions.

Training Methodology

Basically, six experiments were conducted to fine-tune the performance of the model as described in the following.

i. Experiment 1

In Experiment 1, the models mentioned above (ResNet50, ResNet101, VGG16, VGG19, MobileNetV2, DenseNet121 and InceptionResNetV2) were trained for 150 epochs with early stopping (patience = 20). After evaluation, the top three models were selected for further improvements.

ii. Experiment 2: Architectural Updates

Several architectural updates were made to enhance the model in Experiment 2. The base model was loaded with pre-trained weights and all layers of the base model were set to non-trainable to retain the learned knowledge. A Flatten layer was added to convert the output of the base model into a 1-dimensional feature vector. Additionally, a fully connected layer with 512 hidden units and ReLU activation was introduced for helping the model to learn more complex patterns. A Dropout layer with a rate of 0.5 was also added to prevent overfitting by randomly deactivating some neurons during the training process. Finally, a Dense layer with 2 units and SoftMax activation was included as the output layer to classify the two classes in the dataset. These changes provided more flexibility and allowed the model to capture intricate patterns properly compared to the simpler architecture in Experiment 1.

iii. Experiment 3: Key Architectural Changes

There are several key architectural changes introduced from Experiment 3 compared to Experiment 2. Initially, a Global Average Pooling layer was added right after the pre-trained model by averaging the spatial information from the feature maps and reducing dimensionality while retaining critical information. This makes a clear shift from the Flatten layer used in Experiment 2. Also, the Batch normalization layers were added to reduce internal covariate shifts by leading to faster training and more stable learning. Furthermore, Experiment 3 included additional Dense layers (with 256 units and ReLU activation) and Dropout layers (with a 0.5 rate) which helps the model to learn richer representations and generalize better. Moreover, opted to make the last 15 layers of the pre-trained model trainable with respect to the frozen state of the model in Experiment 2 which allowed the upper layers to capture task specific details.

Summarizing, with regards to the results achieved in Experiments 1, 2, and 3 as detailed in Table 3, it was clear that InceptionResNetV2 surpassed both DenseNet121 and MobileNetV2, particularly in precision, recall, F1-score, and general accuracy of the models. This was especially pronounced in Experiment 3, where InceptionResNetV2 scored an overall accuracy of 91%, the highest value across all models evaluated.

Table 3. Model Experiments Comparison

Architecture	Model	Configuration
Experiment 1	DenseNet121	- DenseNet121 + Flatten - Dense(2, activation='softmax')
Experiment 1	InceptionResNetV2	- InceptionResNetV2 + Flatten - Dense(2, activation='softmax')
Experiment 1	MobileNetV2	- MobileNetV2 + Flatten - Dense(2, activation='softmax')
Experiment 2	DenseNet121	- DenseNet121 (with weights='imagenet') - Flatten

		- Dense(512, activation='relu') - Dropout(0.5) - Dense(2, activation='softmax')
Experiment 2	InceptionResNetV2	- InceptionResNetV2 (with weights='imagenet') - Flatten - Activation - Dropout(0.5) - Dense(2, activation='softmax')
Experiment 2	MobileNetV2	- MobileNetV2 (with weights='imagenet') - Flatten - Dense(128, activation='relu') - Dropout(0.5) - Dense(2, activation='softmax')
Experiment 3	DenseNet121	- DenseNet121 (with weights='imagenet', layers frozen except last 15) - GlobalAveragePooling2D - BatchNormalization - Dense(512, activation='relu') - Dropout(0.5) - Dense(256, activation='relu') - BatchNormalization - Dropout(0.5) - Dense(2, activation='softmax')

Experiment 3	InceptionResNetV2	- InceptionResNetV2 (with weights='imagenet', layers frozen except last 15) - GlobalAveragePooling2D - BatchNormalization - Dense(1024, activation='relu') - Dropout(0.5) - Dense(512, activation='relu') - BatchNormalization - Dropout(0.5) - Dense(256, activation='relu') - BatchNormalization - Dropout(0.5) - Dense(2, activation='softmax')
Experiment 3	MobileNetV2	- MobileNetV2 (with weights='imagenet', layers frozen except last 15) - GlobalAveragePooling2D - BatchNormalization - Dense(512, activation='relu') - Dropout(0.5) - Dense(256, activation='relu') - BatchNormalization - Dropout(0.5) - Dense(2, activation='softmax')

iv. **Experiment 4**

Learning rate was fine-tuned in this experiment and attempted with learning rates of 0.01, 0.001, and 0.0001. Moreover, the model was trained three times for a total of 150 epochs with early stopping.

v. **Experiment 5**

As noted in the previous experiment, the selected learning rate for this experiment is 0.0001. Additionally, the optimizer was changed from **Adam** to **RMSprop**. By maintaining the above-mentioned learning rate, five experiments were conducted suing this optimizer. Finally, the model was trained for 150 epochs with early stopping.

3.3.4 Dual-Stage Nail Diagnosis & Verification Algorithm

This algorithm combines predictions from two models named **Inception-ResNetV3** and **Vision Transformer (ViT)** to generate a final diagnosis along with a confidence score. Therefore, the methodology begins by normalizing the confidence scores from both models by scaling them from a range of 0-100 to 0-1. This is done by dividing the scores by 100 to ensure they are comparable for weighted calculations. Basically, three steps are involved in this step.

I. **Normalizing Scores**

Normalization is done by bringing scores between the range of 0-100 to a standard range of 0-1 to guarantee the confidence scores from both models are within a similar range. This is basically done by dividing confidence score of each model by 100 as illustrated in Figure 8.

$$\text{inception\_score\_norm} = \frac{\text{inception\_pred\_score}}{100}$$

$$\text{vit\_score\_norm} = \frac{\text{vit\_pred\_score}}{100}$$

Figure 8. Confidence Score Normalization Process

This normalization technique ensures that both models contribute reasonable manner when their scores are combined in the subsequent steps.

II. **Weighted Average if Models Agree**

If both models agree on the predicted class (i.e., inception_pred_class == vit_pred_class), the function calculates the weighted average of the confidence scores as Figure 9 which w_{vit} represent the predefined weights of Inception-ResNetV3 (0.4) and ViT (0.6) respectively.

$$\text{final\_confidence\_score} = w_{\text{inception}} \times \text{inception\_pred\_score} + w_{\text{vit}} \times \text{vit\_pred\_score}$$

Figure 9. Weighted Confidence Score Calculation

The final diagnosis is the model predicted class as they agree. Therefore, the return values are the predicted class from anyone of the models and the combined confidence score.

III. **Weighted Scores if Models Disagree**

If the models disagree on the predicted class, the function computes the weighted scores for both models by multiplying the normalized confidence scores with their respective weights as Figure 10.

$$\text{weighted\_inception\_score} = w_{\text{inception}} \times \text{inception\_score\_norm}$$
$$\text{weighted\_vit\_score} = w_{\text{vit}} \times \text{vit\_score\_norm}$$

Figure 10. Weighted Score Calculation for Disagreement Cases

The model which has the higher weighted score is selected as the final diagnosis. The final confidence score is determined by multiplying the winning model's weighted score by 100 for reverting it back to the 0-100 scale as shown in the Figure 11.

$$final\_confidence\_score = \max(weighted\_inception\_score, weighted\_vit\_score) \times 100$$

Figure 11. Final Confidence Score

Finally, this methodology ensures that the final diagnosis is based on the model with the highest confidence while also incorporating the predictions from both models for a more reliable decision-making process.

3.3.5 Nail Abnormality Detection and Segmentation Model

The **Nail Abnormality Detection and Segmentation Model** has been developed to diagnose two specific nail conditions called nail plank lines and splinters. Since these conditions have similar visual characteristics making accurate detection and differentiation seems to be challenging task. The model helps improve accuracy in detecting and differentiating these abnormalities. So that, the model this by utilizing advanced **instance segmentation** techniques which detect the abnormalities while precisely outlining the regions affected by these conditions. This system uses two advanced models called **YOLOv8 Instance Segmentation** and **YOLOv7 Instance Segmentation**. Both of these modes are built on the YOLO (You Only Look Once) architecture which is a popular object detection framework. So that the inclusion of instance segmentation allows the model to provide both classification and pixel-level localization of abnormalities which is critical for medical applications.

YOLOv8 Instance Segmentation is a big step forward in the YOLO family of models. It detects objects like nail abnormalities while outlining their exact shape with high precision. This makes it easier to differentiate between similar-looking conditions. Thanks to its advanced neural network design and smarter learning techniques, YOLOv8 delivers more accurate results and faster processing compared to earlier versions.

On the other hand, **YOLOv7 Instance Segmentation** is a powerful and efficient model in the YOLO family that includes instance segmentation capabilities. It is optimized for real-time performance by allowing it to process images quickly without sacrificing accuracy. This makes it particularly effective for detecting small objects such as nail abnormalities which require detailed segmentation. Moreover, YOLOv7 uses advanced techniques to capture fine details which is essential for identifying subtle and irregularly shaped abnormalities. Also its ability to deliver fast and precise segmentation makes it a strong choice for real-time diagnostic applications.

The decision to use instance segmentation models which YOLOv8 and YOLOv7 over traditional YOLO detection models plays a crucial role in the effectiveness of the Nail Abnormality Detection and Segmentation Model. Unlike standard YOLO models which only draw bounding boxes around detected objects and instance segmentation goes a step further by precisely outlining the shape and area of each abnormality. This is especially important for detecting conditions like nail plank lines and splinters which can be faint, irregular or blend in with normal nail features. By capturing fine details, instance segmentation ensures more accurate detection while making it a valuable tool for diagnosis and treatment planning.

Instance segmentation is especially useful in medical image analysis because it helps the model to differentiate between multiple similar abnormalities even when they are close together or overlapping. For example, nail plank lines and splinters often appear in the same areas on the nail and therefore it is difficult to separate them using traditional detection models. However, instance segmentation works at the pixel level by allowing the model to precisely outline each abnormality. This level of detail makes it easier for medical professionals to accurately diagnose and distinguish between these conditions.

Summarizing, the instance segmentation models like YOLOv8 and YOLOv7 outperform traditional YOLO detection models in the Nail Abnormality Detection and Segmentation Model. Their ability to segment at the pixel level provides the precision needed to accurately identify and distinguish subtle nail abnormalities. This makes them particularly valuable for medical diagnostics, enabling more reliable detection, improved diagnosis and therefore better treatment planning for conditions like nail plank lines and splinters.

Dataset Annotation

For training the Nail Abnormality Detection and Segmentation Model, a dataset containing 300 images of splinters and 300 images of nail plank lines has been used. The annotation process was carried out using Roboflow which is a widely used platform for dataset management and image labeling. Each abnormality was manually outlined with polygon annotations to ensure precise labeling of the affected areas. After that,the detailed annotation method was applied individually to each image by allowing the model to learn both the presence and exact boundaries of abnormalities. This level of precision is crucial for instance segmentation as it helps the model to accurately differentiate between different nail conditions as shown in Figure 12 and Figure 13.

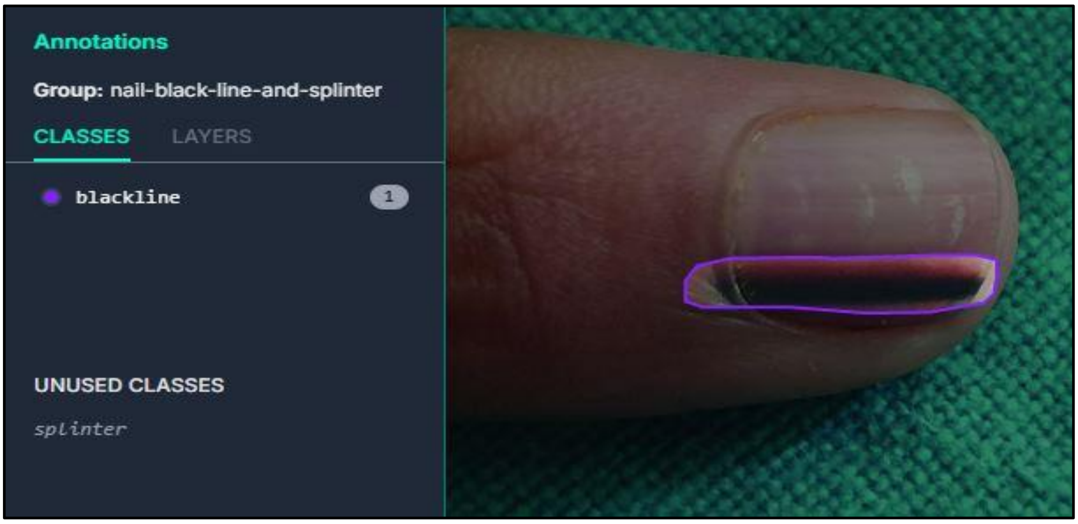


Figure 12. Annotation of Nail Black Line Using Roboflow

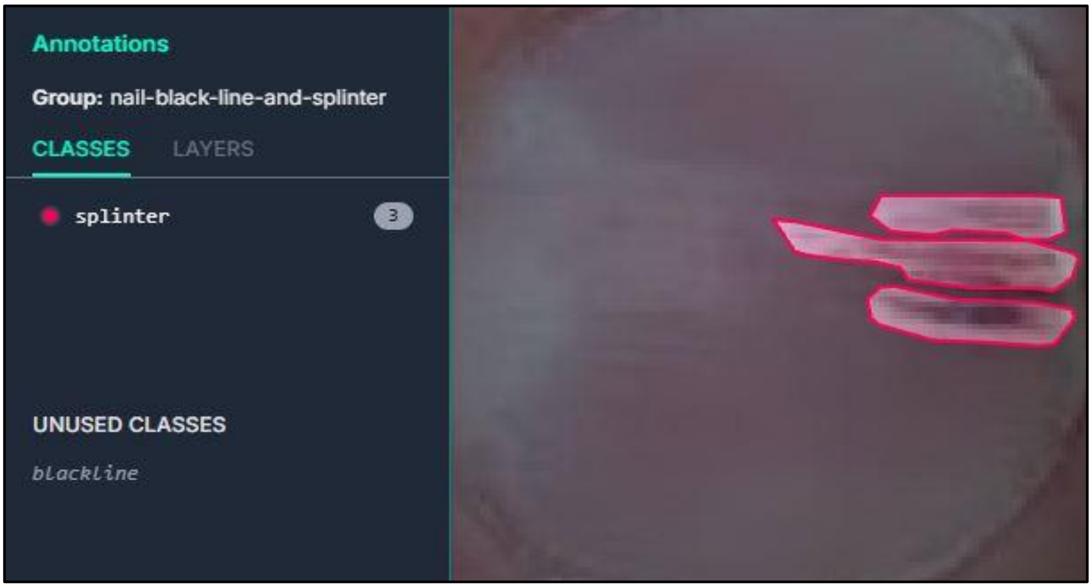


Figure 13. Annotation of Nail Splinters Using Roboflow

Preprocessing

The preprocessing stage is essential for standardizing and preparing images before feeding them into the machine learning model. It ensures consistency of the data by improving the ability of the model to be learnt effectively. The following preprocessing steps were applied in this step.

I. Auto-Orient: This step corrects the orientation of images based on their metadata, such as EXIF tags. If an image was captured upside down or rotated, it is automatically adjusted to the correct upright position. Finally, this ensures that all images are consistently aligned, preventing any misinterpretation by the model due to orientation issues.

II. Resize (Stretch to 640x640): All images were resized to a standard 640x640 pixels to ensure consistency as the model requires uniform input dimensions. The "stretch" method was used for resizing which may slightly distort images if they were not originally square. While this can affect visual proportions, it allows the model to process images efficiently and ensures compatibility with its input layer.

Augmentation

Image augmentation is a crucial technique for expanding the training dataset by applying different transformations to the images. Basically, this enhances the ability of the model to generalize and reducing the risk of overfitting to the original dataset. The following augmentation techniques were applied here.

- **Flip (Horizontal, Vertical)**: Images were flipped horizontally (left to right) and vertically (upside down) while generating mirrored versions of the images. This helps the model to learn invariant features and makes it more robust to changes in orientation.

- **90° Rotate (Clockwise, Counter-Clockwise, Upside Down)**: A range of rotations was applied to the images including 90°, 180°, and 270° rotations. This helps the model to become invariant to orientation and recognize abnormalities from various perspectives.

- **Crop (0% Minimum Zoom, 17% Maximum Zoom)**: A random crop of the image was performed with the zoom level ranging from 0% to 17%. This simulates different framing perspectives and enabling the model to learn features from partial views of the nail abnormalities.

- **Rotation (Between -22° and +22°)**: Random rotations between -22° and +22° were applied to simulate slight tilts in the image while making the model more resilient to minor misalignments during real world data collection.

- **Shear (±6° Horizontal, ±6° Vertical)**: The image was skewed horizontally and vertically by introducing a perspective shift. This simulates how the appearance of abnormalities might change due to slight changes in viewing angles.

- **Saturation (Between -14% and +14%)**: The saturation of the images was adjusted by increasing or decreasing the color intensity. This helps the model to learn features under varying lighting conditions and color contrasts.

- **Exposure (Between -5% and +5%)**: Exposure was modified by slightly increasing or decreasing the brightness of the images. This allows the model to become more robust to different lighting conditions and image exposures that might occur in real world scenarios.

Model Training

Basically, two advanced instance segmentation models called **YOLOv7 Instance Segmentation** and **YOLOv8 Instance Segmentation** have been used for the training. These models were selected for their ability to detect objects while performing pixel-level segmentation needed for accurately identifying and outlining subtle nail abnormalities.

When coming to the training methodology, it is employed categorical cross-entropy loss for handling the multi-class classification aspects efficiently. The dataset was split with 80% allocated for training and the remaining 20% evenly divided between validation and testing. This confirm the model learns effectively while being evaluated on unseen data for accurate performance assessment.

Model Parameters is as follows.

- **Framework**: Ultralytics YOLOv8.0.20
- **Environment**: Python 3.10.12, torch 2.0.1+cu118
- **Epochs**: 100
- **Learning Rate**: 1e-4
- **Hardware**: Tesla P100 and Tesla T4 GPUs

The nano versions of both models (YOLOv7 Nano and YOLOv8 Nano) have been used to optimize training efficiency while maintaining high accuracy. These lightweight versions allowed for faster training times and reduced computational resource usage without compromising performance. Training was conducted for 100 epochs, ensuring the models had sufficient iterations to learn the underlying patterns in the dataset effectively. This approach balanced accuracy and efficiency, making the models well-suited for real-time nail abnormality detection.

3.3.6 Visual Saliency Algorithm

The final component, the Visual Saliency Algorithm, plays a key role in making the system more transparent by incorporating Explainable AI (XAI). It creates visual representations of the diagnostic findings by highlighting important areas in the input images. Also, this improves the interpretability of the results and building user trust by providing clear, visual explanations of the system made decisions.

Vision Transformers (ViT) and Attention Maps

Vision Transformers (ViTs) adapt the transformer architecture which was originally designed for natural language processing to analyze images. In this approach, an image is divided into smaller patches which are then flattened and converted into numerical representations called embeddings. These embeddings are passed through multiple transformer layers where self-attention mechanisms help the model to capture spatial relationships and important features within the image.

Attention maps in Vision Transformers offer a visual representation of the areas which the model focuses on while processing an image. Each attention head within a transformer layer highlights different regions while helping to identify which parts of the image being most influenced by the model predictions. This visualization is especially valuable in tasks like image classification, object detection and segmentation which understands the reasoning behind a model decision.

Convolutional Neural Networks (CNNs) and Saliency Maps

Convolutional Neural Networks (CNNs) are fundamentals in image processing and computer vision due to their ability to recognize spatial patterns and hierarchies within images. They analyze images through a series of convolutional, pooling, and fully connected layers and allowing them to classify images or detect specific features and objects.

Saliency maps are a powerful visualization technique used with CNNs to highlight the most relevant areas of an image that influence the model made decision. They are generated by computing the gradient of the output with respect to the input image by revealing how sensitive the model predictions are to different image regions. Furthermore, this makes them particularly useful for understanding which features play a key role in the model conclusions. Additionally, saliency maps can help debug the model behavior by ensuring meaningful areas of the image for a given task.

Differences Between Attention Maps (ViT) and Saliency Maps (CNNs)

The following Table 4 is an analysis of conceptual approach, visual interpretation and usefulness in model debugging.

Table 4. Differentiation of ViT vs. CNNs

Feature	Attention Maps	Saliency Maps
Conceptual Approach	Based on attention mechanisms within layers.	Based on gradient of the output with respect to the input.
Visual Interpretation	Shows inter-part interactions and layer-wise contributions.	Highlights influential pixels or regions directly.
Usefulness in Model Debugging	Help understand the model's focus across regions and layers, offering insights into decision-making.	Identify which pixels are most influential to predictions, useful for basic visual explanations.

Pros and Cons Between Attention Maps (ViT) and Saliency Maps (CNNs)

The following Table 5 analyses pros and cons between ViT and CNNs.

Table 5. Pros and Cons of ViT vs. CNNs

	Attention Maps (ViT)	Saliency Maps (CNNs)
Pros	- Deep contextual insights - Visualizes complex interactions - Multi-layer and multi-head insights	- Easy to compute and interpret - Directly indicates influential areas - Quick visual feedback
Cons	- Complex to compute and interpret - Computationally intensive	- Can be noisy - Less detailed on interactions between regions

Attention Map Generation Using ViT

Basically, this is a six-step process as described in the following.

I. **Image Collection and Processing**

- **Multiple Source Handling**: Images are collected from a variety of sources, including local storage, web URLs, and in-memory streams. This allows the system to be versatile and adaptable to different application scenarios, ensuring seamless image acquisition regardless of the data source. The process for reading images from various sources is implemented in the code shown in Figure 14.
- **Preprocessing for Uniformity**: Each image is converted to RGB format to maintain consistency in color representation, which is crucial for accurate image analysis. Then they are resized to a specific dimension (like 224x224 pixels) as required for the input layer of the Vision Transformer to ensure that all images are treated equally by the model.

```python
def read(filepath_or_buffer: ImageInputType, size, timeout=None):
    """Read a file into an image object
    Args:
        filepath_or_buffer: The path to the file or any object
            with a `read` method (such as `io.BytesIO`)
        size: The size to resize the image to.
        timeout: If filepath_or_buffer is a URL, the timeout to
            use for making the HTTP request.
    """
    if PIL is not None and isinstance(filepath_or_buffer, PIL.Image.Image):
        return np.array(filepath_or_buffer.convert("RGB"))
    if isinstance(filepath_or_buffer, (io.BytesIO, client.HTTPResponse)):
        image = np.asarray(bytearray(filepath_or_buffer.read()), dtype=np.uint8)
        image = cv2.imdecode(image, cv2.IMREAD_UNCHANGED)
    elif isinstance(filepath_or_buffer, str) and validators.url(filepath_or_buffer):
        with request.urlopen(filepath_or_buffer, timeout=timeout) as r:
            return read(r, size=size)
    else:
        if not os.path.isfile(typing.cast(str, filepath_or_buffer)):
            raise FileNotFoundError(
                "Could not find image at path: " + filepath_or_buffer
            )
        image = cv2.imread(filepath_or_buffer)
    if image is None:
        raise ValueError(f"An error occurred reading {filepath_or_buffer}.")

    image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)
    return cv2.resize(image, (size, size))
```

Figure 14. Image Reading and Preprocessing Code

II. Vision Transformer Model Setup

- **Patch Creation**: The Vision Transformer (ViT) model processes an image by dividing it into multiple smaller patches as depicted in Figure 15, like how sentences are broken down into individual words in Natural Language Processing (NLP). This segmentation allows the model to treat each patch independently while capturing spatial relationships much like tokenization in NLP where each word contributes to the overall understanding of the text.

```python
from tensorflow.keras.layers import Layer, Dense, Reshape

class PatchCreator(Layer):
    def __init__(self, patch_size, num_patches, embedding_dim):
        super(PatchCreator, self).__init__()
        self.patch_size = patch_size
        self.num_patches = num_patches
        self.embedding_dim = embedding_dim
        self.projection = Dense(embedding_dim)

    def call(self, images):
        # Create patches and project to embedding dimension
        batch_size = tf.shape(images)[0]
        patches = tf.image.extract_patches(
            images=images,
            sizes=[1, self.patch_size, self.patch_size, 1],
            strides=[1, self.patch_size, self.patch_size, 1],
            rates=[1, 1, 1, 1],
            padding='VALID')
        patches = Reshape((self.num_patches, patches.shape[-1]))(patches)
        return self.projection(patches)
```

Figure 15. Patch Creation for Vision Transformer

- **Embedding and Position Encoding**: Each patch is then embedded into a higher-dimensional space and supplemented with positional encodings as illustrated in Figure *16*. These encodings provide the model with spatial context, helping it understand where each patch is located relative to others, which is vital for maintaining the spatial hierarchy of features within the image.

```python
class PositionalEncoding(Layer):
    def __init__(self, num_positions, embedding_dim):
        super(PositionalEncoding, self).__init__()
        self.pos_embedding = self.add_weight("pos_embedding", shape=[1, num_positions,

    def call(self, patch_embeddings):
        return patch_embeddings + self.pos_embedding
```

Figure 16. Positional Encoding in Vision Transformer

III. Attention Mechanism of the Transformer

Self-Attention Calculation: Self-attention mechanisms determine how each image patch influences the representation of other patches inside the transformer. The model assigns a set of scores to each patch by indicating its impact on all other patches. As shown in Figure 17, this mechanism allows the model to selectively focus on the most relevant parts of the image. By doing so, it enhances the model's ability to capture meaningful spatial relationships and important features, improving overall image representation.

```python
class MultiHeadSelfAttention(tf.keras.layers.Layer):
    def __init__(self, embed_size, num_heads):
        super(MultiHeadSelfAttention, self).__init__()
        self.embed_size = embed_size
        self.num_heads = num_heads
        self.depth = embed_size // num_heads

        self.wq = tf.keras.layers.Dense(embed_size)
        self.wk = tf.keras.layers.Dense(embed_size)
        self.wv = tf.keras.layers.Dense(embed_size)
        self.dense = tf.keras.layers.Dense(embed_size)

    def split_heads(self, x, batch_size):
        x = tf.reshape(x, (batch_size, -1, self.num_heads, self.depth))
        return tf.transpose(x, perm=[0, 2, 1, 3])

    def call(self, v, k, q, mask):
        batch_size = tf.shape(q)[0]

        q = self.wq(q)  # (batch_size, seq_len, embed_size)
        k = self.wk(k)
        v = self.wv(v)

        q = self.split_heads(q, batch_size)  # (batch_size, num_heads, seq_len_q, depth)
        k = self.split_heads(k, batch_size)
        v = self.split_heads(v, batch_size)

        scaled_attention_logits = tf.matmul(q, k, transpose_b=True)
        scaled_attention_logits /= tf.math.sqrt(tf.cast(self.depth, tf.float32))

        if mask is not None:
            scaled_attention_logits += (mask * -1e9)

        attention_weights = tf.nn.softmax(scaled_attention_logits, axis=-1)

        output = tf.matmul(attention_weights, v)  # (batch_size, num_heads, seq_len_q, de
        output = tf.transpose(output, perm=[0, 2, 1, 3])  # (batch_size, seq_len_q, num_h
        concat_output = tf.reshape(output, (batch_size, -1, self.embed_size))  # (batch_s
```

Figure 17. Multi-Head Self-Attention in Vision Transformer

Aggregation Across Heads: Multiple sets of attention scores (from multiple heads) are combined and allowing the model to gather varied perspectives. As shown in Figure 18, this multi-head approach enables the model to capture different types of features at various positions while enriching the overall feature representation.

```python
class TransformerBlock(tf.keras.layers.Layer):
    def __init__(self, embed_size, num_heads, ff_dim, rate=0.1):
        super(TransformerBlock, self).__init__()
        self.att = MultiHeadSelfAttention(embed_size, num_heads)
        self.ffn = tf.keras.Sequential(
            [tf.keras.layers.Dense(ff_dim, activation="relu"), tf.keras.layers.
        )
        self.layernorm1 = tf.keras.layers.LayerNormalization(epsilon=1e-6)
        self.layernorm2 = tf.keras.layers.LayerNormalization(epsilon=1e-6)
        self.dropout1 = tf.keras.layers.Dropout(rate)
        self.dropout2 = tf.keras.layers.Dropout(rate)

    def call(self, inputs, training, mask):
        attn_output = self.att(inputs, inputs, inputs, mask)
        attn_output = self.dropout1(attn_output, training=training)
        out1 = self.layernorm1(inputs + attn_output)

        ffn_output = self.ffn(out1)
        ffn_output = self.dropout2(ffn_output, training=training)
        return self.layernorm2(out1 + ffn_output)
```

Figure 18. Transformer Block for Multi-Head Attention Aggregation

IV. Extraction and Scaling of Attention Weights

Weight Extraction from Transformers: After the image passes through the transformer model, the attention weights from these layers can be extracted. As illustrated in Figure 19, these weights highlight the importance given to each patch when making a decision, especially mapping out the focus areas of the model.

```python
class ExtractorModel(tf.keras.Model):
    def __init__(self, model):
        super().__init__()
        self.model = model  # the original Vision Transformer model

    def call(self, inputs):
        # Assuming the model's architecture allows access to attention scores
        for layer in self.model.layers:
            if isinstance(layer, TransformerBlock):
                # Get outputs and attention scores
                _, attention_scores = layer.att(inputs, return_attention_scores=True)
                return attention_scores  # Return attention scores for visualization
        return None
```

Figure 19.Extracting Attention Weights from Vision Transformer

Conversion to Visual Heatmap: These extracted weights are then scaled back up to the dimensions of the original image. As shown in Figure 20, the scaled weights are converted into a heatmap which visually represents areas of high and low focus and overlaying this heatmap on the original image gives a direct visual representation of the model's attention distribution.

```
def generate_heatmap(attention_weights, image_size):
    """
    Generates a heatmap from the attention weights and scales it to the image size.
    """
    # Assume attention_weights shape is (num_patches, num_patches)
    num_patches = int(np.sqrt(attention_weights.shape[0]))
    patch_size = image_size // num_patches

    # Reshape attention weights to the image grid
    attention_map = attention_weights.reshape((num_patches, num_patches))
    attention_map = cv2.resize(attention_map, (image_size, image_size), interpolation=cv2.

    # Normalize for display
    attention_map = (attention_map - np.min(attention_map)) / (np.max(attention_map) - np.
    return attention_map
```

Figure 20. Generating Heatmap from Attention Weights

V. Visualization and Analysis

Comparative Display: The original image and its corresponding attention heatmap are displayed side by side and allowing for an easy comparison between the actual visual elements and the regions where the model concentrated its attention. This dual representation provides a clear and intuitive way to assess how accurately the model interprets and processes the image.

Interpretability and Diagnostic Insight: By examining the areas where the model directs its attention, users can assess whether it is focusing on the right features for making predictions. This is especially important in applications like medical diagnosis where the model should prioritize clinically relevant features that indicate a disease rather than being influenced by irrelevant background details. Ensuring the model's attention aligns with meaningful features helps improve trust and reliability in AI driven decision-making.

VI. Real World Applications

Medical Imaging: In the medical field, attention maps can help reassure practitioners that the AI model is focusing on relevant anatomical or pathological features. By visually correlating the model's attention with established diagnostic markers, these maps enhance interpretability and trust while ensuring that the AI supports medical diagnostics in a meaningful and reliable way.

Automated Surveillance: In security applications, attention maps highlight the areas of a scene that the model focuses on the most by providing insights into its decision-making process. This helps in refining the model to better detect anomalies, suspicious activities or critical events by ultimately improving the accuracy and reliability of security systems.

Convolutional Neural Networks (CNNs) and Saliency Maps

Related to this study, several saliency maps such as Grad-CAM, Ablation-CAM, and SIDU (Similarity, Difference and Uniqueness) have been analyzed to achieve model interpretability and currently evaluating explainability using InceptionResNetV2 model.

Similarity Difference and Uniqueness (SIDU) Method for Model Interpretability

The Similarity Difference and Uniqueness (SIDU) method is an advanced technique in computational vision which determines pixel-level saliency using deep learning models. It works by extracting and analyzing the final convolutional layers of deep Convolutional Neural Networks (CNNs) to generate detailed visual explanations of model predictions. SIDU operates through three key steps which are generating feature activation image masks, computing feature importance weights and synthesizing a comprehensive visual representation. This structured approach enhances the interpretability of CNN decisions and therefore making it a valuable tool for understanding deep learning-based image analysis.

I. Generation of Feature Activation Image Masks

The first step in the SIDU method is generating feature activation image masks. This begins by extracting feature activations from the final convolutional layer of a CNN model such as Inception-ResNet-V2. This layer typically has dimensions $n \times n \times N$, where n represents the spatial dimensions and N is the number of feature activations. Each activation map captures how a specific feature responds across the input image and is converted into a binary mask. Thereafter this mask is then unsampled using bilinear interpolation to match with original image dimensions and merged with the input image through point-wise multiplication. This process highlights the localized influence of each feature by laying the groundwork for further analysis and interpretability in the next stages of the SIDU method.

II. Computing Feature Importance Weights

The core analytical process of the SIDU method revolves around computing feature importance weights. This is basically done by evaluating the prediction scores associated with each feature activation image mask using the same predictive model. The goal is to determine the contribution of each feature to the final prediction through two key analyses which are namely similarity differences and uniqueness.

- **Similarity differences** measure how much the influence of a feature deviates from the model's overall prediction for the original image. This helps to identify features that significantly impact the prediction outcome.
- **Uniqueness** assesses which features have a distinct effect on the model's decision by highlighting their importance in shaping the final output.

By integrating these two metrics, the SIDU method derives feature importance weights which quantitatively represent the relevance and impact of each feature on the predicted result. This structured approach enhances the interpretability of deep learning models by providing insights into how different features contribute to decision-making.

III. Visual Explanations for the Prediction

The final step of the SIDU method is generating a visual explanation for the prediction. This has been achieved by combining the weighted feature activation masks into a comprehensive saliency map. Moreover, the weight of each feature has been applied to its corresponding activation mask during this process. The result is a heatmap that highlights the most critical features influencing the decision making by the model.

Apart from that, saliency map serves as a powerful interpretability tool which offers a clear visualization of how deep learning models make predictions. It helps building trust in automated decision-making systems with transparency, especially in critical applications like medical diagnostics and security.

Overall, the SIDU method provides a structured approach to understanding the inner operations of deep learning models. By revealing the specific image features that drive predictions, it promotes greater accountability in AI and supports researchers and practitioners in developing more explainable and trustworthy AI systems.

Gradient-weighted Class Activation Mapping (Grad-CAM)

Gradient-weighted Class Activation Mapping (Grad-CAM) is a technique that enhances the interpretability of convolutional neural networks (CNNs) by visualizing the important regions in an input image which contribute to a prediction. Grad-CAM produces a heatmap highlighting these regions and hence allowing users to understand the focus of the model during decision-making. The methodology flow of Grad-CAM is as follows.

- **Model and Layer Selection**

  Grad-CAM requires a trained CNN model and a specific layer for analysis. In this research, the **last batch normalization layer** has been used instead of the final convolutional layer. Batch normalization layers normalize intermediate activations while preserving spatial information and making them suitable for generating class activation maps.

- **Forward Pass**

  The input image is passed through the model and then the output predictions are obtained. Thereafter the target class is selected for analysis which could be the predicted class or a user-specified class.

- **Gradient Computation**

  A backward pass is performed to compute the gradients of the target class score with respect to the activations of the selected layer. These gradients indicate the importance of each activation for the target class.

- **Weighted Feature Map Combination**

  The computed gradients are used to calculate importance weights for each feature map. These weights are combined with the activations of the selected layer to create a class activation map. This map highlights the areas of the input image that strongly influence the decision.

- **Heatmap Generation**

  The class activation map is resized to match the dimensions of the input image. Also a colormap is applied to convert the activation values into a visually interpretable heatmap.

- **Overlay and Visualization**

  The heatmap is overlaid on the original input image to visually indicate the regions the model focused on for its prediction.

In this research, Grad-CAM has been customized by utilizing the last batch normalization layer which enable us to preserve more feature and spatial information. However, the results produced by Grad-CAM during evaluation were not as effective as those generated by the

SIDU approach. The heatmaps from Grad-CAM appeared less focused and lacked the precision required for clear interpretability. In contrast, SIDU provided more accurate and detailed visualizations while offering a clearer representation of the model's attention and making it a more reliable method for explainability.

Ablation-CAM

Ablation-CAM is an improved visualization technique built upon traditional methods like Grad-CAM which designed to provide more precise and focused interpretations of neural network predictions. Unlike Grad-CAM which relies solely on gradient information, Ablation-CAM systematically removes (ablates) portions from the model feature maps during computation to evaluate their impact on the output. By analyzing how the model predictions change when specific features are removed, this method offers a more reliable representation of the regions that truly influence the decision-making process. The methodology flow is as follows.

I. **Model and Layer Selection**

Similar to Grad-CAM, Ablation-CAM operates on a pre-trained CNN and focuses on a specific layer for generating visualizations. In this research, the **last batch normalization layer** has been selected to retain spatial and feature information essential for localization.

II. **Systematic Ablation**

Ablation-CAM works by systematically disabling (ablating) portions of the feature maps in the selected layer and observing the effect on the output score for the target class. By iterating all over feature maps, it calculates the contribution of each map to the prediction.

III. **Contribution Computation**

The change in the target class score caused by ablation is used to quantify the importance of each feature map. This process ensures that only the most critical activations are considered which resulting in more precise localization of important regions.

IV. **Weighted Feature Map Combination**

The calculated importance of each feature map is used as a weight to combine them. This weighted combination generates the ablation activation map and also highlighting the areas in the input image that have the strongest influence on the model's prediction.

V. **Heatmap Generation**

The ablation activation map is resized to the same dimensions as the input image. After that colormap is applied to create a visually interpretable heatmap like Grad-CAM.

VI. **Overlay and Visualization**

The heatmap is overlaid on the original input image which clearly showing the most relevant regions to the model decision.

In this research, Ablation-CAM has been implemented to enhance the interpretability and overcome some of the limitations of Grad-CAM. By systematically removing contributions from feature maps, Ablation-CAM generated more focused and reliable heatmaps which making it easier to pinpoint the regions that influenced the model's predictions. While it outperformed Grad-CAM in terms of precision and clarity, it still fell short compared to SIDU. SIDU provided sharper and more detailed visualizations by making it the superior choice in our comparative analysis for accurately interpreting the model decision-making process.

CHAPTER 4: EVALUATION

4.1 Evaluation for Nail Filter Model

The final evaluation of the Nail Filter Model highlights its outstanding performance in accurate differentiation between nail and non-nail images. This evaluation is based on a detailed classification report and confusion matrix which demonstrate the effectiveness of the model by ensuring that only the relevant images proceed to further analysis in the nail disease diagnostic pipeline. The model consistently classifies images with a high degree of confidence and minimal errors with an overall accuracy of 98%. The classification report reveals that the model achieved 100% precision for non-nail images which means that every image classified as *non-nail* was completely accurate by eliminating any false positives. In the context of nail images, the model attained a 97% precision score and showed that only a small fraction of non-nail images was misclassified as nail images. This high level of precision ensures that irrelevant images are effectively filtered out by preventing unnecessary processing in later stages of the diagnostic system.

Apart from the precision, the model demonstrates an exceptional recall performance by achieving 97% recall for non-nail images by successfully identifying 97% of all non-nail samples without mistakenly excluding many images. Meanwhile, the nail class achieved a perfect 100% recall by ensuring that every single nail image in the dataset was correctly classified. This indicates that the model effectively captures all relevant nail images without omitting any image which is crucial for ensuring that potentially diseased nails are not overlooked. The F1-score, which balances both precision and recall, achieved an impressive 0.98 for both classes. This indicates that the model maintains a well-balanced approach by avoiding both excessive caution and unnecessary aggressiveness in its classifications. Moreover, it is essential to maintain the balance in medical imaging applications, as misclassifications can result in unnecessary tests or missed diagnoses that could impact patient care.

The confusion matrix provides a detailed evaluation of the model classification accuracy by distinguishing between correct and incorrect predictions. The model accurately identified 248 non-nail images as ***not_nail*** by achieving a 97% accuracy rate for this category. However, 7 non-nail images were misclassified as nail images, resulting in a 3% false positive rate. While this indicates that a small number of irrelevant images may be forwarded to the diagnostic system, the high precision score minimizes the overall impact. In the context of ***nail*** class, the model correctly classified 254 nail images with only one instance of misclassification, where a nail image was mistakenly labeled as ***not_nail***. This corresponds to an exceptionally low false negative rate by ensuring that nearly all nail images are correctly detected and analyzed.

These evaluation results demonstrate the robustness, reliability and practical effectiveness of the Nail Filter Model in real-world applications. The model enhances the image preprocessing stage of the system by achieving high accuracy, strong recall and approximately perfect precision.

Also, the model ensures that no relevant cases are overlooked and minimizing the risk of missing potential signs of disease with 100% recall for nail images. Simultaneously, its high precision for non-nail images prevents unnecessary computational resource usage on irrelevant data while improving the efficiency.

Finally, the Nail Filter Model enhances the diagnostic pipeline by reducing human intervention, speeding up automated analysis, and improving system performance. The model achieves high precision, recall, and F1-score, as illustrated in Figure 21, demonstrating its reliability in distinguishing between nail and non-nail images.

Additionally, the confusion matrix, shown in Figure 22, highlights the model's ability to correctly classify images, with minimal misclassifications. This results in faster and more reliable disease detection, ultimately boosting the accuracy of downstream models and aiding medical professionals in making more informed decisions for patient care.

Classification Report:

	precision	recall	f1-score	support
not_nail	1.00	0.97	0.98	255
nail	0.97	1.00	0.98	255
accuracy			0.98	510
macro avg	0.98	0.98	0.98	510
weighted avg	0.98	0.98	0.98	510

Figure 21. Classification Report for Nail Filter Model

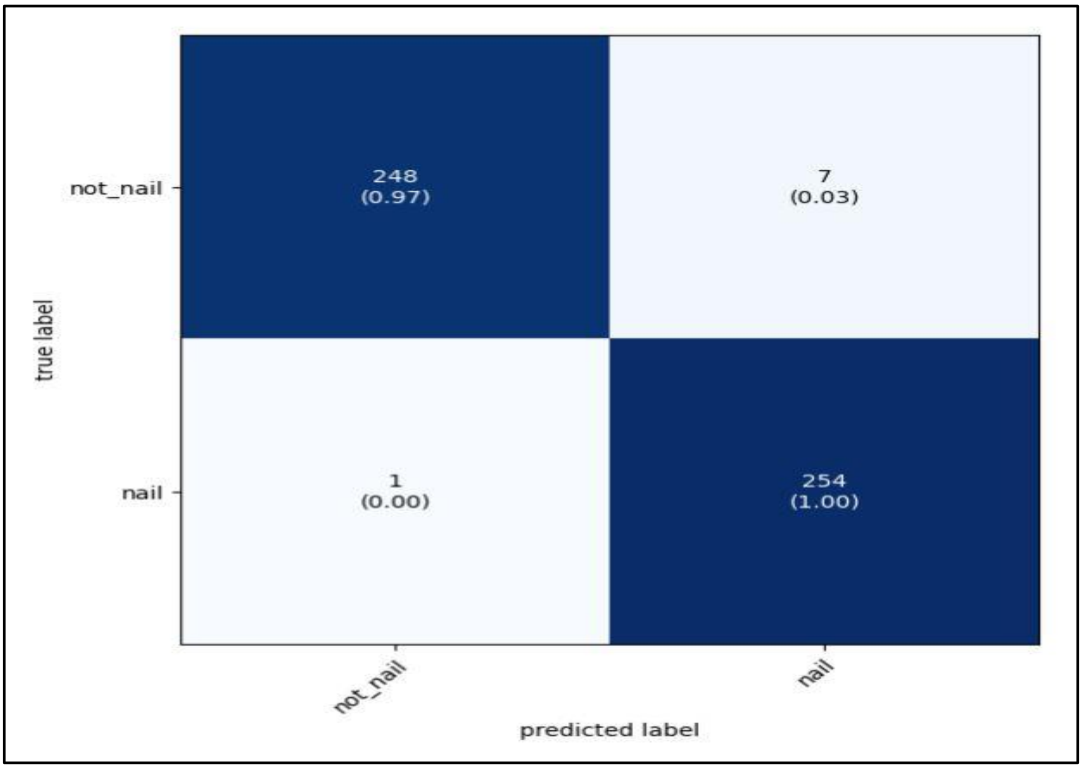


Figure 22. Confusion Matrix for Nail Filter Model

4.2 Evaluation of ViT Nail Diagnosis Model (Verification Model)

There were two experiments conducted for ViT Nail Diagnosis Model as described in the below.

4.2.1 Experiment 1

The objective of the first experiment was to determine the optimal learning rate for the classification model. Three different learning rates were tested which are 0.1, 0.01, and 0.001 using the RMSprop optimizer. The purpose was to analyze how each learning rate impacts the model performance and identifying which one yields the best accuracy. The results shown in Table 6 indicate that a learning rate of 0.1 achieved 93% accuracy while 0.01 reached 94% accuracy and 0.001 resulted in 95% accuracy. Among all of these, 0.001 was chosen as the most suitable choice for further experimentation.

Table 6. Learning Rate vs. Accuracy for ViT Nail Diagnosis Model

Learning Rate	Accuracy
0.1	93%
0.01	94%
0.001	95%

A detailed classification report was generated as shown in Figure 24 for the best-performing learning rate which is 0.001 as shown in the below. Overall, the model achieved 95% of accuracy by demonstrating strong classification capabilities.

Additionally, the confusion matrix, shown in Figure 23 provides insights into the model's predictive performance across different classes. The matrix highlights the correct classifications and misclassifications, demonstrating the model's ability to accurately distinguish between nail abnormalities.

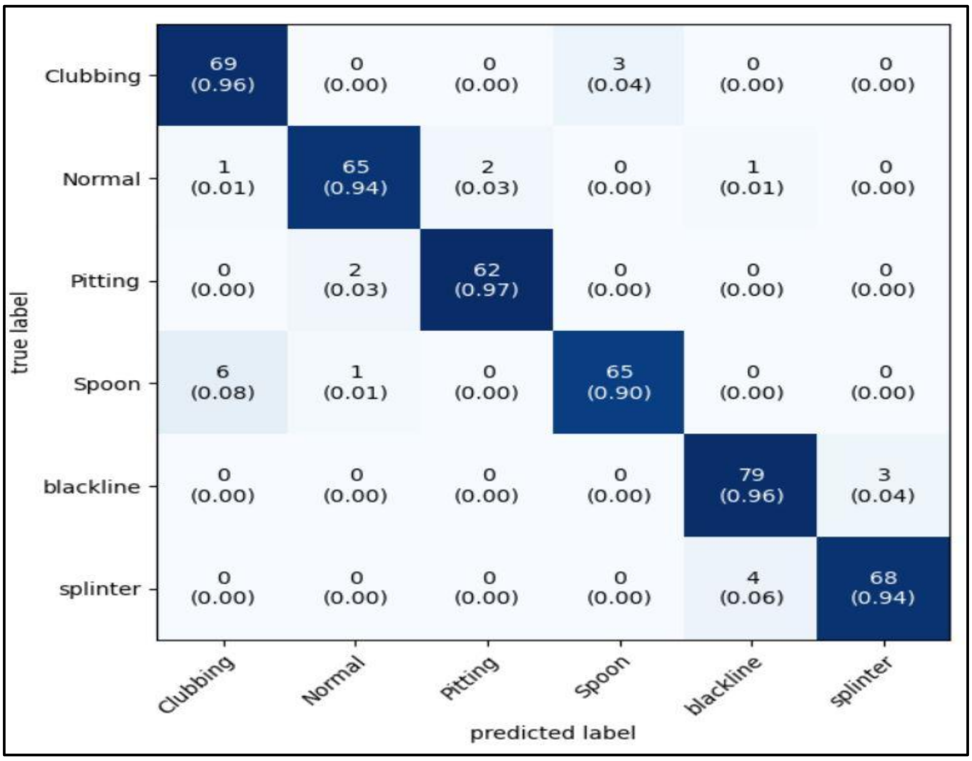


Figure 23. Confusion Matrix for ViT Nail Diagnosis Model

Classification Report:

	precision	recall	f1-score	support
Clubbing	0.91	0.96	0.93	72
Normal	0.96	0.94	0.95	69
Pitting	0.97	0.97	0.97	64
Spoon	0.96	0.90	0.93	72
blackline	0.94	0.96	0.95	82
splinter	0.96	0.94	0.95	72
accuracy			0.95	431
macro avg	0.95	0.95	0.95	431
weighted avg	0.95	0.95	0.95	431

Figure 24. Classification Report for ViT Nail Diagnosis Model

The classification report highlights the strong performance of the model across different categories. The Clubbing class achieved 91% precision and 96% recall by indicating consistent and accurate classification with minimal errors. The Normal class maintained a 96% precision and 94% recall which demonstrates balanced and reliable performance. The Pitting class excelled with 97% precision and recall while correctly identifying nearly all instances. Similarly, the Spoon class also performed well by achieving 96% precision and 90% recall with only a few missed cases. The Blackline class showed stability with 94% precision and 96% recall while the Splinter class maintained high accuracy at 96% precision and 94% recall.Overall, the model, trained with a 0.001 learning rate and the RMSprop optimizer delivered high accuracy and balanced performance and confirming its effectiveness in classifying nail conditions.

4.2.2 Experiment 2

The second experiment evaluated different optimizers to improve the performance of the model while maintaining the 0.001 learning rate from Experiment 1. The goal was to compare the effectiveness of Adam, RMSprop and Adadelta in achieving the highest accuracy. The results shown in the Table 7 indicate that RMSprop outperformed the others with a 96% accuracy while Adam and Adadelta both reached 95%. This confirms that RMSprop was the most effective optimizer for this configuration.

Table 7. Optimizer Comparison for ViT Nail Diagnosis Model

Optimizer	Accuracy
RMSprop	96%
Adam	95%
Adadelta	95%

The following classification report shown in Figure 26 was generated for the best performing configuration which used the RMSprop optimizer with a 0.001 learning rate. Furthermore, the model achieved 96% overall accuracy by marking a slight but meaningful improvement over the previous experiment.

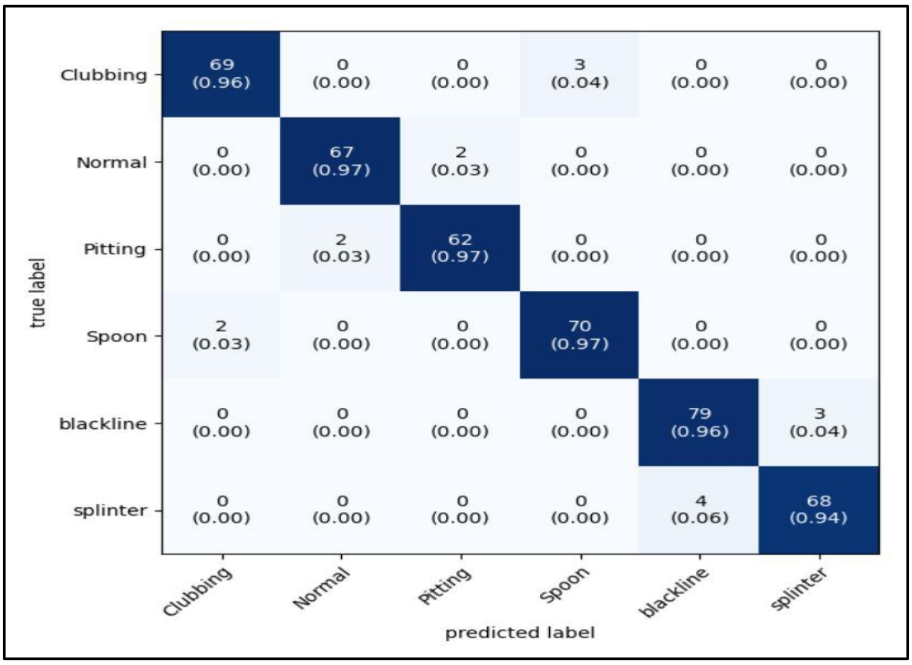


Figure 25. Confusion Matrix for Optimized ViT Nail Diagnosis Model

Classification Report:

	precision	recall	f1-score	support
Clubbing	0.97	0.96	0.97	72
Normal	0.97	0.97	0.97	69
Pitting	0.97	0.97	0.97	64
Spoon	0.96	0.97	0.97	72
blackline	0.95	0.96	0.96	82
splinter	0.96	0.94	0.95	72
accuracy			0.96	431
macro avg	0.96	0.96	0.96	431
weighted avg	0.96	0.96	0.96	431

Figure 26. Classification Report for Optimized ViT Nail Diagnosis Model

The classification report for this experiment revealed notable improvements across multiple categories. The Clubbing class maintained strong consistency with 97% precision and 96% recall, while the Normal class improved to 97% precision and recall, reducing misclassifications and enhancing stability. The Pitting class achieved 97% precision and recall, confirming near-perfect classification. The Spoon class showed a slight improvement with 96% precision and 97% recall, while the Blackline class remained stable at 95% precision and 96% recall.

The Splinter class retained 96% precision and 94% recall, reinforcing the model's reliability across all categories. These enhancements, particularly in the Pitting and Normal classes, suggest that RMSprop's optimization refined the learning process, leading to better classification accuracy and improved generalization. The confusion matrix, shown in Figure 25 further illustrates the model's predictive performance, highlighting its ability to correctly classify each category with minimal misclassifications.

4.2.3 Best Hyperparameters

After analyzing the results from both experiments, the optimal hyperparameters for the classification model were identified. The best learning rate was 0.001, and RMSprop emerged as the most effective optimizer, delivering the highest classification accuracy with balanced performance across all categories.

RMSprop's ability to dynamically adjust the learning rate, combined with the moderate 0.001 learning rate, contributed to stable and reliable classification results. The improvements in precision, recall, and overall accuracy indicate that this configuration enhances the model's capability to generalize across different categories. The diversity in the nail conditions dataset, shown in Figure 27 further demonstrates the model's ability to learn complex patterns across multiple nail abnormalities.

These optimized hyperparameters will serve as the foundation for further model refinements and optimizations, ensuring continuous improvement in diagnostic accuracy.

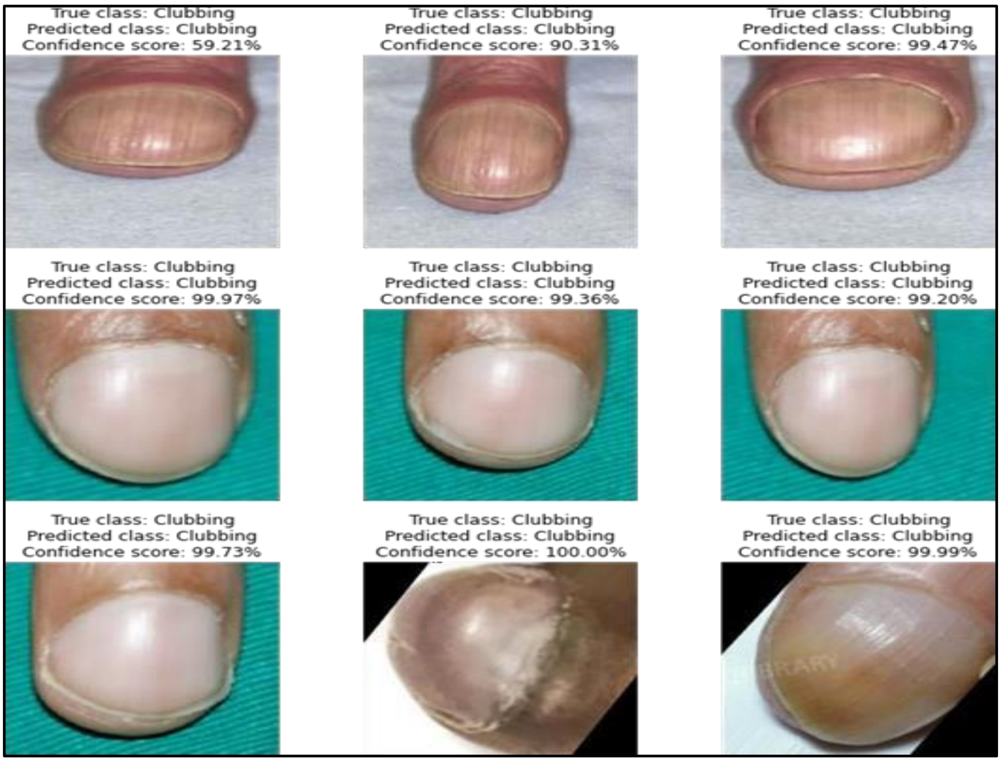


Figure 27. Sample Images from Nail Diagnosis Dataset

4.3 Evaluation of CNN Nail Diagnosis Model

4.3.1 Experiment 1

Table 8. Accuracy Comparison of CNN-Based Transfer Learning Models

Model	Accuracy
ResNet50	0.82
ResNet101	0.81
VGG16	0.85
VGG19	0.84
MobileNetV2	0.86
DenseNet121	0.87
InceptionResNetV2	0.88

Several CNN-based transfer learning models were trained and evaluated in this experiment, with their performance compared based on accuracy. The results, as shown in Table 8 indicate that ResNet50 achieved 0.82 accuracy, while ResNet101 performed slightly lower at 0.81. The VGG16 and VGG19 models demonstrated strong results, with 0.85 and 0.84 accuracy, respectively.

Among the tested models, MobileNetV2 and DenseNet121 both reached 0.86 accuracy, but the best-performing model was InceptionResNetV2, achieving an accuracy of 0.88. Based on these results, InceptionResNetV2 was selected as the most effective model for further refinement. Subsequently, two experiments were conducted to determine the optimal learning rate and optimizer to further enhance model performance and improve classification accuracy.

4.3.2 Experiment 2

In Experiment 2, architectural changes were made and selected the best three models from Experiment 1 based on their performance. The models that were chosen for further improvement were MobileNetV2, DenseNet121, and InceptionResNetV2. The corresponding accuracies for these models as shown in Table 9, were 0.88 for both MobileNetV2 and DenseNet121 while InceptionResNetV2 achieved an accuracy of 0.90. These top three models were further evaluated and improved in the subsequent experiment to enhance their performance.

Table 9. Accuracy of Top Three Models After Architectural Modifications

Model	*Accuracy*
MobileNetV2	0.88
DenseNet121	0.88
InceptionResNetV2	0.90

4.3.3 Experiment 3

In Experiment 3, architectural enhancements were applied to the models selected in Experiment 2 and their performance was re-evaluated. The results, as shown in Table 10 indicate that MobileNetV2 achieved an accuracy of 0.91 while DenseNet121 reached 0.90. InceptionResNetV2 outperformed the others with 0.92 accuracy by making it the top-performing model. Based on these results, InceptionResNetV2 was chosen for further experimentation in Experiments 4 and 5 to refine and optimize its performance.

Table 10. Accuracy Improvement After Architectural Enhancements

Model	**Accuracy**
MobileNetV2	0.91
DenseNet121	0.90
InceptionResNetV2	0.92

4.3.4 Experiment 4

Experiment 4 aimed to determine the optimal learning rate for the classification model using the RMSprop optimizer. Three learning rates which 0.1, 0.01, and 0.001 were tested to assess their impact on model performance. The results shown in the Table 11 indicate that a 0.1 learning rate achieved 91% accuracy while 0.01 reached 93% and 0.001 resulted in 92%. Since 0.01 yielded the highest accuracy, it was selected as the optimal learning rate for further experimentation.

Table 11. Learning Rate Optimization Results

Learning Rate	**Accuracy**
0.1	91%
0.01	93%
0.001	92%

A detailed classification report for further evaluation was generated for the best-performing learning rate which is 0.01. As shown in Figure 29, the model achieved an overall accuracy of 93% by confirming its strong classification capabilities.

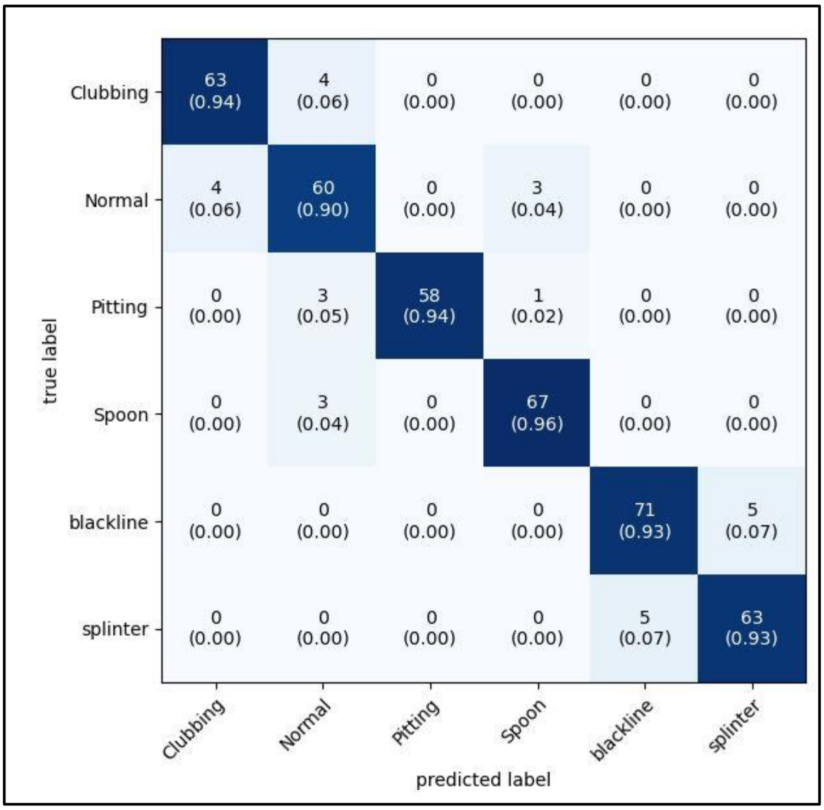
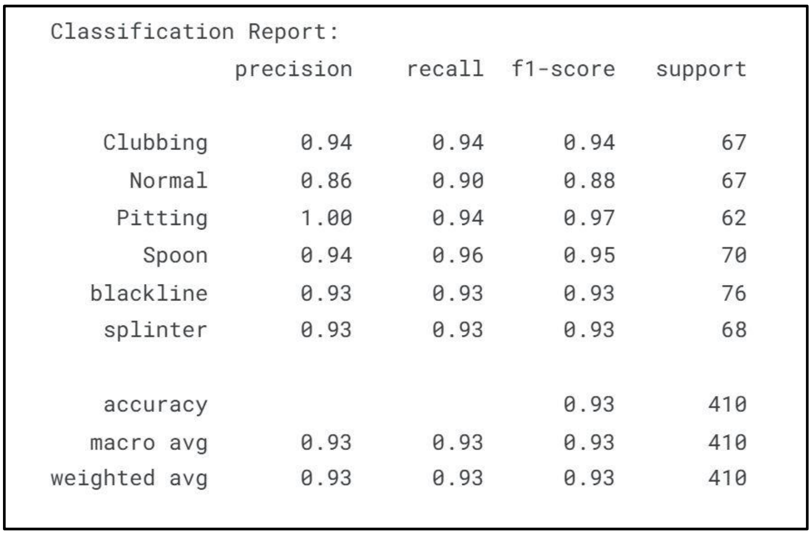


Figure 28. Confusion Matrix for Optimized Learning Rate (0.01)



Classification Report:

	precision	recall	f1-score	support
Clubbing	0.94	0.94	0.94	67
Normal	0.86	0.90	0.88	67
Pitting	1.00	0.94	0.97	62
Spoon	0.94	0.96	0.95	70
blackline	0.93	0.93	0.93	76
splinter	0.93	0.93	0.93	68
accuracy			0.93	410
macro avg	0.93	0.93	0.93	410
weighted avg	0.93	0.93	0.93	410

Figure 29. Classification Report for Optimized Learning Rate (0.01)

The classification report provided key insights into the model's performance across different categories. The Clubbing class achieved 94% precision and recall, demonstrating consistent and accurate classification with minimal errors. The Normal class had a slightly lower 86% precision and 90% recall, suggesting some misclassifications, likely due to overlapping features with similar conditions.

The Pitting class performed exceptionally well, with 100% precision and 94% recall, indicating that all predicted instances were correct, although a small percentage of actual cases were missed. Similarly, the Spoon class achieved 94% precision and 96% recall, correctly identifying most cases with only a few misclassifications. The Blackline and Splinter classes both demonstrated balanced performance, each achieving 93% precision and recall, indicating the model's stability in classifying these categories.

The confusion matrix, as shown Figure 28, further highlights the model's classification effectiveness across all categories. The high accuracy and well-balanced classification report confirm that training with a 0.01 learning rate and the RMSprop optimizer resulted in a highly effective model.

4.3.5 Experiment 5

The final experiment aimed to evaluate different optimizers to further enhance the performance. Based on Experiment 4, a 0.001 learning rate was selected due to its promising results. The objective was to compare the performance of Adam, RMSprop and Adadelta in achieving the highest accuracy. The results shown in the Table 12, indicate that the Adam optimizer outperformed the others with an accuracy of 94% while RMSprop and Adadelta both achieved 93%. This suggests that Adam's adaptive learning rate and momentum-based approach contributed to better model performance while making it the most suitable optimizer for this task.

Table 12.Optimizer Comparison for Final Experiment

Optimizer	Accuracy
Adam	94%
RMSprop	93%
Adadelta	93%

A detailed classification report was generated for the best-performing configuration, which used the Adam optimizer with a learning rate of 0.001. As shown in Figure 30, the model achieved an overall accuracy of 94%, demonstrating a slight but valuable improvement compared to the previous experiment.

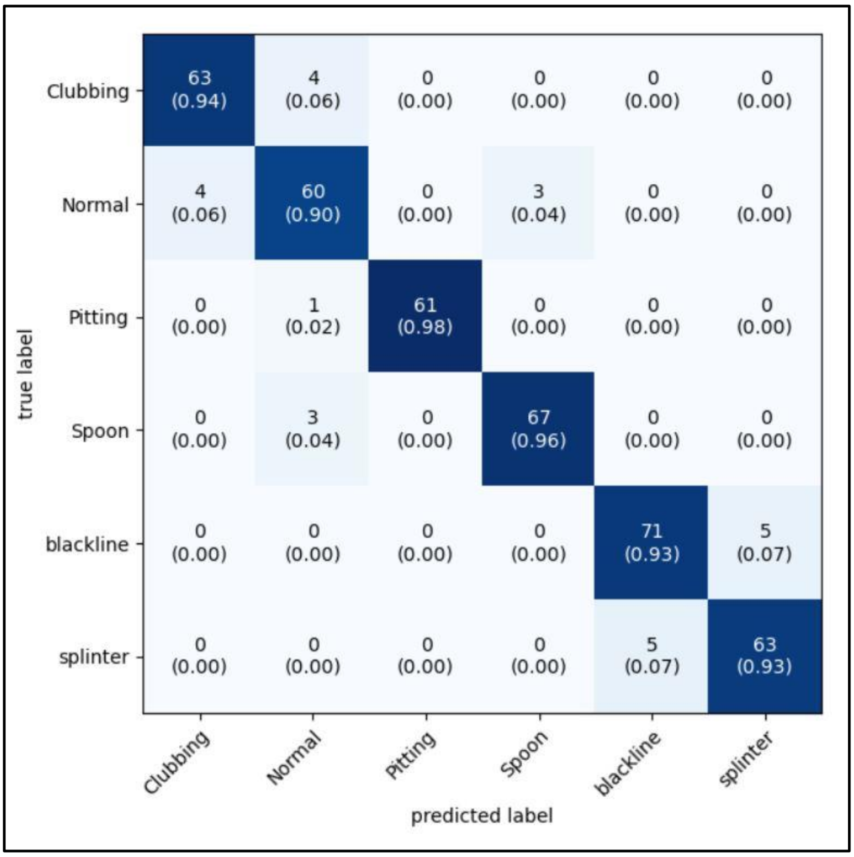


Figure 30. Confusion Matrix for Best-Performing Configuration (Adam, LR=0.001)

Classification Report:

	precision	recall	f1-score	support
Clubbing	0.94	0.94	0.94	67
Normal	0.88	0.90	0.89	67
Pitting	1.00	0.98	0.99	62
Spoon	0.96	0.96	0.96	70
blackline	0.93	0.93	0.93	76
splinter	0.93	0.93	0.93	68
accuracy			0.94	410
macro avg	0.94	0.94	0.94	410
weighted avg	0.94	0.94	0.94	410

Figure 31. Classification Report for Best-Performing Configuration (Adam, LR=0.001)

The classification report as shown in Figure 31 confirmed the improvements achieved using the Adam optimizer. The Clubbing class maintained 94% precision and recall by demonstrating consistent and reliable classification. The Normal class showed slight improvement with 88% precision and 90% recall while reducing misclassifications and improving stability compared to Experiment 1. The Pitting class saw the most significant improvement by reaching 100% precision and 98% recall which indicates nearly all cases were correctly classified with only a few being missed. The Spoon class also improved by achieving 96% precision and recall indicating better distinction of this category. The Blackline and Splinter classes looked stable at 93% precision and recall while showing no major deviations from previous results. Overall, the improvements, especially in the Pitting and Normal categories suggest that Adam's optimization method contributed to a more refined learning process which leads to better generalization and classification accuracy.

4.3.6 Best Hyper parameter

After analyzing the results from all experiments, the optimal hyperparameters for the classification model were identified. The best learning rate was 0.001, and Adam was chosen as the most effective optimizer. This combination delivered the highest classification accuracy while maintaining balanced performance across all categories.

Adam's adaptive learning rate and momentum-based approach, combined with a moderate learning rate, enabled the model to achieve stable and reliable classification results. As seen in

Figure 32, the dataset used for training and evaluation contained a diverse set of nail conditions, contributing to the model's ability to learn complex patterns effectively.

Figure 32. Sample Images from Optimized Nail Diagnosis Dataset

Improvements in precision, recall, and overall accuracy indicate that this configuration enhances the model’s ability to generalize effectively across different categories. Moving forward, these hyperparameters will serve as the foundation for further model enhancements and refinements.

4.4 Evaluation of Nail Abnormality Detection and Segmentation Model

The performance was evaluated based on standard metrics which are mean Average Precision (mAP), Precision and Recall. These metrics provide a comprehensive view of the model capability to accurately detect and segment abnormalities in the nail images.

Below is a comparison of YOLOv7 vs YOLOv8 models results as shown in Table 13.

Table 13. YOLOv7 and YOLOv8 models results

Standard Metric	YOLOv7	YOLOv8
mean Average Precision	82.5%	86.6%
Precision	90.3%	93.6%
Recall	79.1%	82.1%

YOLOv7 achieved a respectable mAP of 82.5% with high precision (90.3%) by indicating that the model is good at correctly identifying positive instances. However, the recall value of 79.1% suggests that there are some instances where the model may miss abnormalities, particularly in challenging or unclear cases.

YOLOv8 performed slightly better than YOLOv7 with an improved mAP of 86.6%. The precision of 93.6% indicates that YOLOv8 is highly accurate when it identifies abnormalities. The recall value of 82.1% shows a slight improvement over YOLOv7 which means YOLOv8 is better at identifying true positive instances, although there are improvements in terms of reducing false negatives.

Summarizing, the **YOLOv8 Instance Segmentation Model** outperformed the **YOLOv7 Instance Segmentation Model** in terms of both mAP and precision while achieving a slightly higher recall. The results indicate that YOLOv8 is better suited for detecting and segmenting nail abnormalities with higher accuracy and fewer missed cases by making it the preferred choice for this specific task.

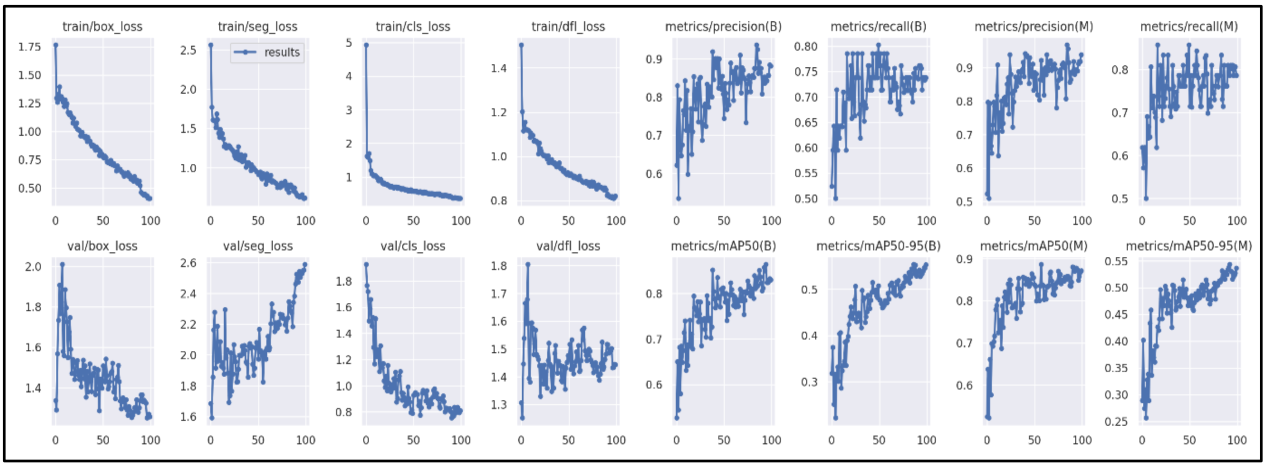


Figure 33. Training and Validation Metrics for YOLOv7 and YOLOv8 Models

As shown in Figure 33, the graphs display various metrics and losses over epochs for both training and validation. Below is a brief overview of each component:

I. **Box and Segmentation Loss (Training and Validation)**: These plots show the decrease in box and segmentation loss over epochs which indicates that the model is learning and improving its predictions over time. The validation loss curves are crucial to ensure that the model is not overfitting.

II. **Classification and Dice-Focal Loss (Training and Validation)**: These graphs similarly demonstrate a decrease in classification and dice-focal loss which suggests the model increasing accuracy in classifying objects and predicting bounding box masks.

III. **Precision and Recall (Bounding Boxes and Masks)**: The precision and recall plots for bounding boxes (B) and masks (M) show fluctuations but generally trend upwards which is positive as it indicates improving model performance in detecting the actual areas of interest in the images.

IV. **mAP Scores**: The mean Average Precision (mAP) scores for bounding boxes and masks at different IoU thresholds which indicate the accuracy of the model in differentiating between the classes and the exactness of the object localization. **mAP50 and mAP50-95** have been used for bounding boxes and masks. These scores show the precision of the model across a spectrum of IoU thresholds. The higher mAP at lower thresholds (mAP50) shows good model performance under less stringent conditions while the drop in mAP50-95 scores at higher thresholds suggests areas for improvement in precise localization.

This training process appears to be effective as indicated by the general downward trends in the loss functions and the overall increase in precision, recall, and mAP scores. To further interpret and possibly improve these results, the following were considered.

I. **Diagnosing Fluctuations**: The noticeable fluctuations, particularly in the validation metrics could suggest model sensitivity to certain data points or potential overfitting. Regularization techniques or more robust data augmentation might help stabilize these metrics.

II. **Further Experiments**: Experiment with different hyperparameters or additional epochs need to be done to see if the model can learn further without overfitting.

III. **Cross-Validation**: Implementing cross-validation could provide more insight into how your model performs across different subsets of your dataset by offering a more robust evaluation.

Figure 34 presents the instance segmentation results obtained from the trained model. The model effectively identifies and segments different nail abnormalities, such as black lines and splinters, with high confidence scores. The segmentation masks highlight the regions of interest, providing a visual representation of the model's ability to distinguish abnormalities. The high confidence scores (ranging from 69% to 86%) demonstrate that the model has successfully learned to classify and segment these conditions.

These qualitative results further validate the findings from the quantitative evaluation metrics discussed earlier. By comparing these results with the training and validation metrics (Figure 34), we can infer that the model generalizes well to unseen data while maintaining high precision and recall values. Future improvements may focus on refining boundary accuracy and enhancing the segmentation of more challenging nail abnormalities.

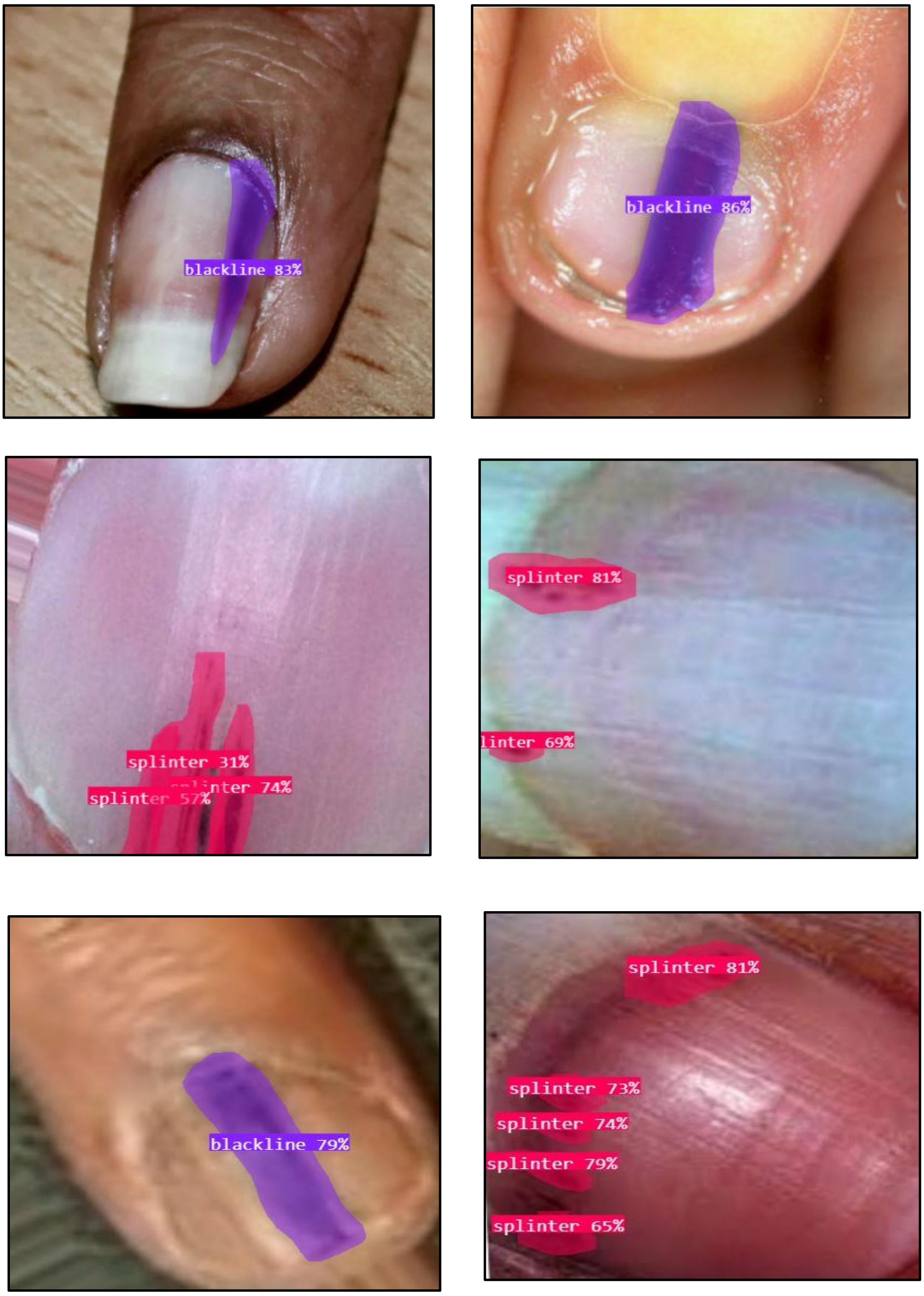


Figure 34. Instance Segmentation Results for Nail Abnormality Detection

4.5 Results of Attention Map Generation Using ViT

Below are sample outputs of the XAI ViT attention map for the ViT nail diagnosis model as shown in Figure 35.

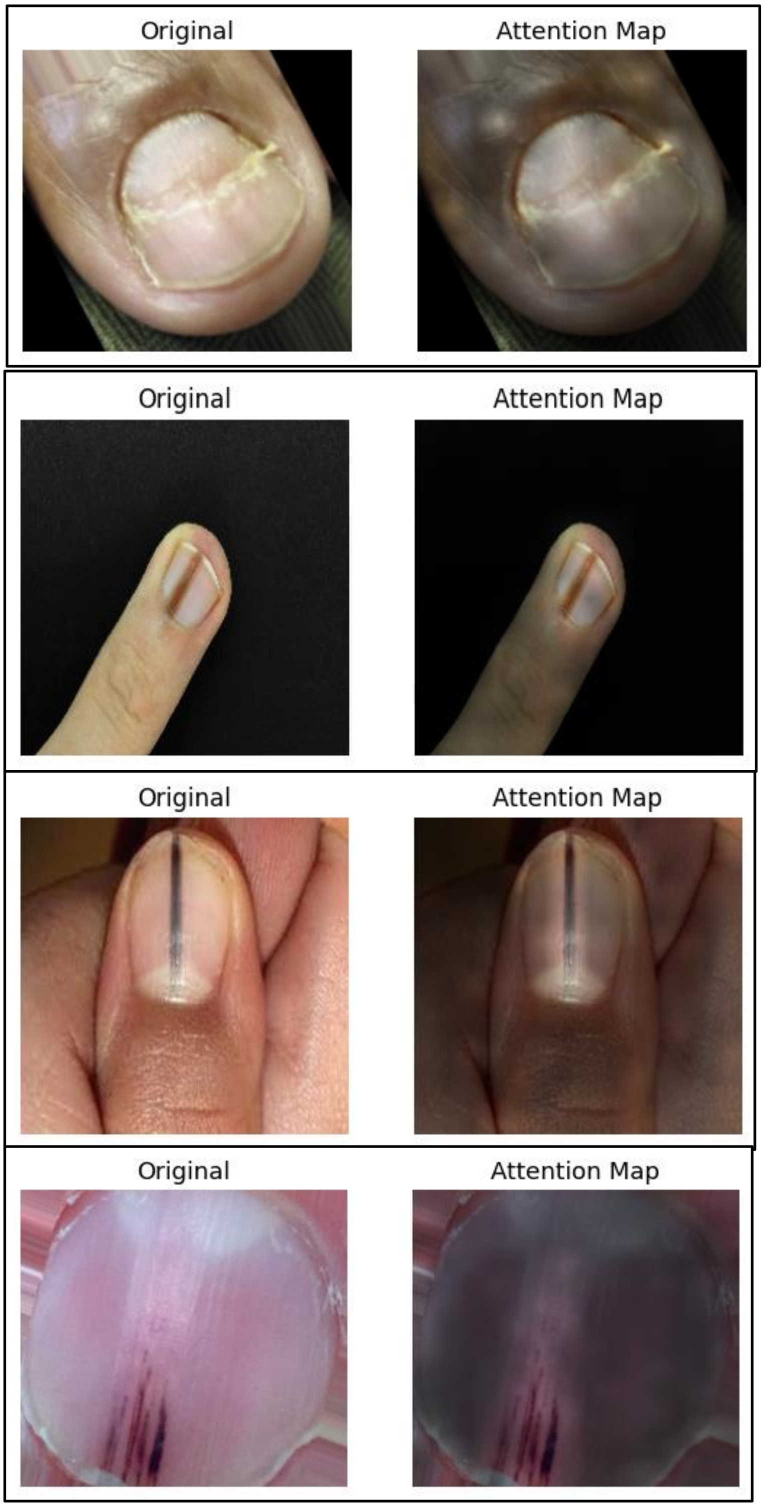


Figure 35. Attention Map for Nail Abnormality Detection for samples

4.6 Results of Convolutional Neural Networks (CNNs) and Saliency Maps

In this study, SIDU proved to be the most effective method for enhancing interpretability and producing highly detailed visualizations as shown in Figure 36. Although Ablation-CAM was implemented to mitigate some of the limitations of Grad-CAM, its heatmaps, while more focused and reliable, did not achieve the same level of clarity and sharpness as SIDU. The visualizations generated by SIDU were noticeably sharper and more detailed by making it the preferred choice in our comparative analysis. Overall, SIDU outperformed both Ablation-CAM and Grad-CAM in terms of precision and interpretability while demonstrating its superiority in this context.

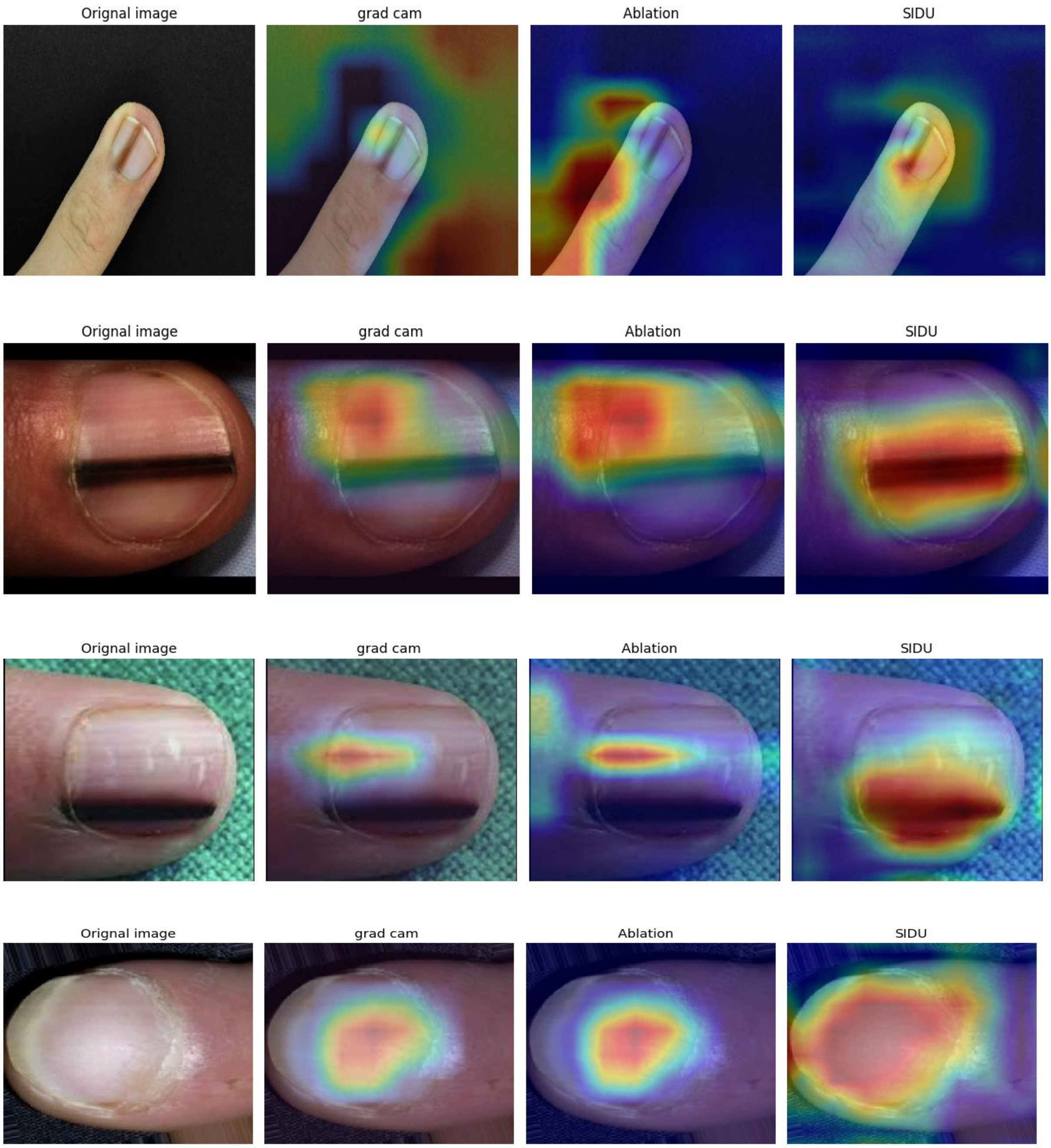


Figure 36. Interpretability and Visualization of Nail Abnormality Detection Using XAI Methods

In the context of Explainable Artificial Intelligence (XAI), the SIDU approach was employed to analyze the interpretability of the final layers in the model. Specifically, the last convolutional layer, the last batch normalization layer and the last activation layer were examined to determine which layer contributed most to explaining the model decisions. Among these, the last batch normalization layer provided the highest level of interpretability. This is likely due to its role in normalizing activations which stabilizes the outputs and makes them more consistent and interpretable, thus allowing for a clearer understanding of the learned features of the model and decision-making process. Figure 37 illustrates the comparative visualizations of different explainability methods, highlighting the effectiveness of SIDU in providing sharper and more interpretable feature maps.

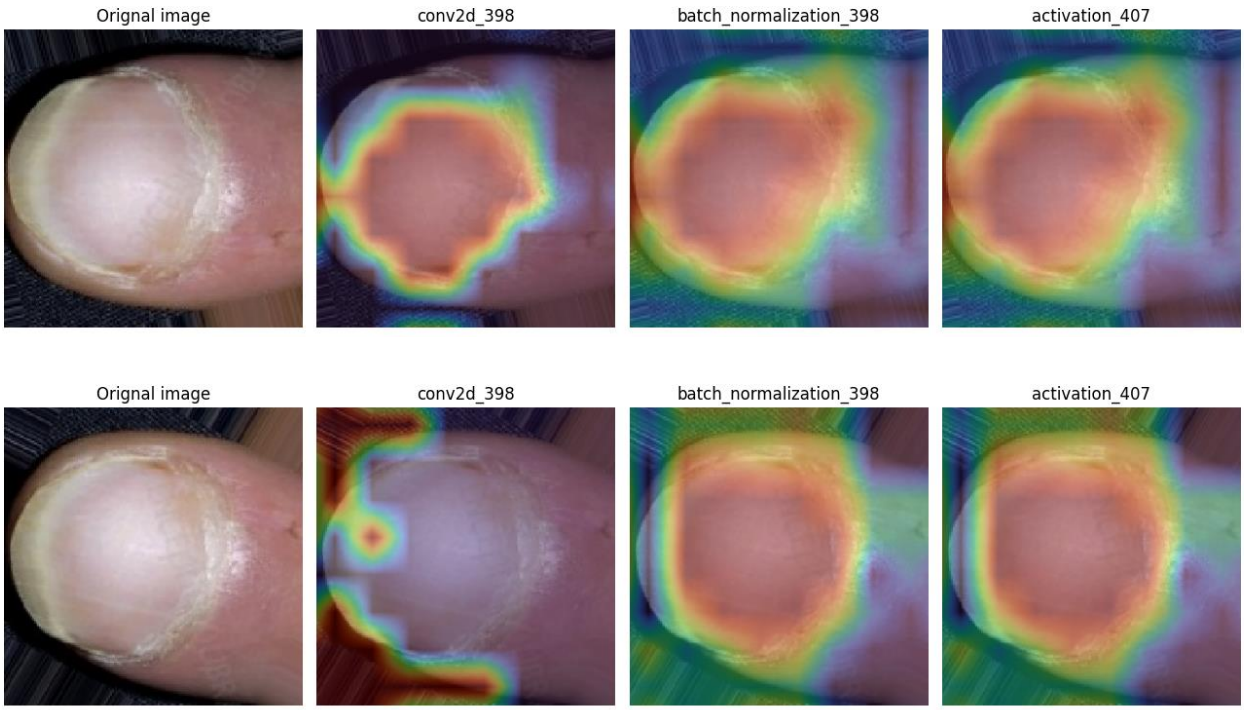


Figure 37. XAI-Based Model Interpretability Visualizations

The high interpretability of the final batch normalization (BatchNorm) layer, particularly in the SIDU approach, can be attributed to its fundamental role in the functioning of the model. Unlike the last convolutional or activation layers which primarily focus on local feature extraction or non-linear transformations, batch normalization standardizes the activations by adjusting their mean and variance. This normalization process not only stabilizes learning but also ensures that activations are distributed in a manner that enhances interpretability.

One of the key advantages of batch normalization is its ability to capture global patterns by operating across the entire mini-batch rather than focusing on localized details. This broader perspective allows it to encode a higher-level representation of learned features, thereby establishing a more consistent and interpretable relationship between the model's internal representations and its predictions. Additionally, batch normalization refines the outputs of preceding layers by reducing noise and improving stability, leading to more structured and simplified feature representations.

As a result, tracing the key factors that influence the decisions of the model becomes more straightforward when using the batch normalization layer compared to convolutional or activation layers. This makes it a more effective choice for improving interpretability in deep learning models, particularly when applied in methods like SIDU.

CHAPTER 5: CONCLUSION AND FUTURE WORK

5.1 Conclusion

The development and implementation of the **Nail-Scan Tool**, an AI-driven system for the detection and diagnosis of nail diseases, have demonstrated significant advancements in leveraging computer vision and deep learning techniques. The multi-model architecture, comprising the **Nail Filter Model**, **CNN Nail Diagnosis Model**, **ViT Nail Diagnosis Model**, **Nail Abnormality Detection and Segmentation Model** and the **Visual Saliency Algorithm** has proven to be a robust and effective solution for nail disease diagnosis. By combining the strengths of convolutional neural networks (CNNs) and vision transformers (ViTs), the system achieves high accuracy, precision, and recall across various nail conditions, including pitting, black lines, splinters, clubbing, and spooning.

The **Nail Filter Model** successfully preprocesses and filters out irrelevant or poor-quality images, ensuring that only valid nail images proceed to the diagnostic stages. This step significantly enhances the efficiency and accuracy of the downstream models. The **CNN Nail Diagnosis Model** and **ViT Nail Diagnosis Model** work in tandem to classify nail conditions with high accuracy with the ViT model achieving a remarkable 96% accuracy using the RMSprop optimizer. The **Dual-Stage Nail Diagnosis & Verification Algorithm** further refines the predictions by combining the outputs of the CNN and ViT models by ensuring reliable and confident diagnoses.

The **Nail Abnormality Detection and Segmentation Model**, utilizing YOLOv8 and YOLOv7 instance segmentation excelsin detecting and segmenting subtle nail abnormalities such as black lines and splinters. YOLOv8, in particular, outperformed YOLOv7 with a mean Average Precision (mAP) of 86.6%, precision of 93.6%, and recall of 82.1%. This model's ability to provide pixel-level segmentation is crucial for accurate diagnosis and treatment planning.

Finally, the **Visual Saliency Algorithm**, leveraging Explainable AI (XAI) techniques such as **SIDU**, **Grad-CAM** and **Ablation-CAM** provides transparent and interpretable visual explanations of the model's predictions. Among these, **SIDU** emerged as the most effective method, offering sharper and more detailed visualizations of the model's attention, particularly when analyzing the last batch normalization layer. This transparency fosters user trust and enhances the interpretability of the system's results.

5.2 Future Work

While the Nail-Scan Tool has achieved significant success, there are several areas for future improvement and expansion:

I. **Dataset Expansion and Diversity**:

- The current dataset, while robust, could benefit from further expansion to include more diverse nail conditions, ethnicities, and imaging conditions. This would improve the model's generalization capabilities and ensure its effectiveness across a broader population.
- Incorporating data from different imaging modalities (e.g., dermoscopy, 3D imaging) could further enhance the system's diagnostic accuracy.

II. **Model Optimization**:

- Continued optimization of the CNN and ViT models, including hyperparameter tuning and architectural refinements, could yield even higher accuracy and efficiency.
- Exploring newer transformer-based architectures (e.g., Swin Transformers) or hybrid models that combine CNNs and transformers in novel ways could further improve performance.

III. **Real-Time Deployment**:

- Developing a lightweight version of the system for real-time deployment on mobile devices or edge computing platforms would make the tool more accessible to healthcare providers and patients in remote or underserved areas.
- Optimizing the inference speed of the YOLOv8 and YOLOv7 models for real-time segmentation and detection would be crucial for practical applications.

IV. **Integration with Clinical Workflows**:

- Integrating the Nail-Scan Tool with electronic health records (EHR) and telemedicine platforms would streamline its adoption in clinical settings.

- Developing a user-friendly interface for dermatologists and healthcare providers to interact with the system and interpret its results would enhance its usability.

V. **Explainability and Trust**:

- Further research into XAI techniques, particularly those that provide more intuitive and clinically relevant explanations, would improve user trust and adoption.

- Exploring interactive visualization tools that allow users to explore the model's predictions and attention maps in greater detail could enhance the system's interpretability.

VI. **Multi-Modal Diagnosis**:

- Combining image analysis with other diagnostic data, such as patient history, lab results, or genetic information, could provide a more comprehensive diagnostic framework.
- Developing a multi-modal AI system that integrates visual, textual, and numerical data could further improve diagnostic accuracy and patient outcomes.

VII. **Validation and Clinical Trials**:

- Conducting large-scale clinical trials to validate the system's performance in real-world settings would be essential for regulatory approval and widespread adoption.
- Collaborating with dermatologists and healthcare providers to gather feedback and refine the system based on clinical needs would ensure its practical relevance.

VIII. **Ethical Considerations**:

- Addressing potential biases in the dataset and ensuring fairness across different demographic groups would be critical for ethical AI deployment.

- Developing guidelines for the responsible use of AI in dermatology, including privacy and data security considerations, would be essential.

The Nail-Scan Tool represents a significant step forward in the application of AI to dermatology and healthcare. By leveraging state-of-the-art computer vision techniques, the system provides accurate, efficient, and interpretable diagnoses of nail diseases, ultimately improving patient outcomes and reducing the burden on healthcare providers. Future work will focus on refining the system, expanding its capabilities, and ensuring its seamless integration into clinical workflows, paving the way for a new era of AI-driven dermatological care.

REFERENCES

7 fingernail problems not to ignore [WWW Document], 2023. . Mayo Clinic. URL https://www.mayoclinic.org/healthy-lifestyle/adult-health/in-depth/7-fingernail-problems-not-toignore/art-20546860

12 nail changes a dermatologist should examine [WWW Document], n.d.

Adadi, A., Berrada, M., 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Ansari, S., Vijay, S., 2023. Computer Aided Disease Detection using Finger Nail Image Processing Technique. International Journal of Advanced Research in Science, Communication and Technology 357–365. https://doi.org/10.48175/IJARSCT-13151

Arrieta, A.B., Díaz-Rodríguez, N., Ser, J.D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Cantone, M., Marrocco, C., Tortorella, F., Bria, A., 2023. Convolutional Networks and Transformers for Mammography Classification: An Experimental Study. Sensors 23, 1229. https://doi.org/10.3390/s23031229

Daglarli, E., 2020. Explainable Artificial Intelligence (xAI) Approaches and Deep Meta-Learning Models. https://doi.org/10.5772/intechopen.92172

Desai, S., Ramaswamy, H.G., 2020. Ablation-CAM: Visual Explanations for Deep Convolutional Network via Gradient-free Localization, in: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 972–980. https://doi.org/10.1109/WACV45572.2020.9093360

Dhanashree, K., Jayabal, P., Kumar, A., Logeswari, S., Priya, K., 2022. Fingernail analysis for early detection and diagnosis of diseases using machine learning techniques. International Journal of Nonlinear Analysis and Applications 13, 61–69. https://doi.org/10.22075/ijnaa.2022.6331

Gohel, P., Singh, P., Mohanty, M., 2021. Explainable AI: current status and future directions. https://doi.org/10.48550/arXiv.2107.07045

H. Pandit, D. M. Shah, 2013. A system for nail color analysis in healthcare, in: 2013 International Conference on Intelligent Systems and Signal Processing (ISSP). Presented at the 2013 International Conference on Intelligent Systems and Signal Processing (ISSP), pp. 221–223. https://doi.org/10.1109/ISSP.2013.6526906

Imouokhome, F.A., Ehimiyein, O.G., Chete, F.O., 2023. Diagnosis and interpretation of breast cancer using explainable artificial intelligence.

Indi, T., Patil, D., 2019. Nail Feature Analysis and Classification Techniques for Disease Detection. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING 7, 8. https://doi.org/10.26438/ijcse/v7i5.13761383

Kashefi, R., Barekatain, L., Sabokrou, M., Aghaeipoor, F., 2024. Explainability of Vision Transformers: a Comprehensive Review and New Perspectives. https://doi.org/10.2139/ssrn.5055345

Khan, W., 2014. A Survey: Image Segmentation Techniques. International Journal of Future Computer and Communication 3, 89–93. https://doi.org/10.7763/IJFCC.2014.V3.274

Lee, S., Yang, C.-S., 2018. Fingernail analysis management system using microscopy sensor and blockchain technology. International Journal of Distributed Sensor Networks 14, 155014771876704. https://doi.org/10.1177/1550147718767044

Lin, Y.-S., Lee, W.-C., Celik, Z.B., 2021. What Do You See?: Evaluation of Explainable Artificial Intelligence (XAI) Interpretability through Neural Backdoors. https://doi.org/10.1145/3447548.3467213

Maniyan, P., Shivakumar, B.L., 2018. Early Disease Detection Through Nail Image Processing Based on Ensemble Of KNN Classifier And Image Features.

Maniyan, P., Shivakumar, D.B.L., Computer, H.J., Science, 2018. Detection of Diseases using Nail Image Processing Based on Multiclass SVM Classifier Method.

Mankodiya, H., Jadav, D., Gupta, R., Tanwar, S., Hong, W.-C., Sharma, R., 2022. OD-XAI: Explainable AI-based Semantic Object Detection for Autonomous Vehicles. Applied Sciences 12. https://doi.org/10.3390/app12115310

Mente, R., 2017. A Review: Fingernail Images for Disease Detection. International Journal of Engineering and Computer Science 6. https://doi.org/10.18535/ijecs/v6i11.01

Meshram, S., 2016. An Automated Medical Support System based on Medical Palmistry and Nail Color Analysis. International Journal Of Engineering And Computer Science. https://doi.org/10.18535/ijecs/v5i6.36

Muddamsetty, S.M., Mohammad, N.S.J., Moeslund, T.B., 2020. SIDU: Similarity Difference And Uniqueness Method for Explainable AI, in: 2020 IEEE International Conference on Image Processing (ICIP). pp. 3269–3273. https://doi.org/10.1109/ICIP40778.2020.9190952

Nithya, D., Masil, A., Kurapati, R., Buggareddy, S.P., Divya, D., 2019. NAIL BASED DISEASE ANALYSIS AT EARLIER STAGE USING MEDIAN FILTER IN IMAGE PROCESSING.

Pandit, H.B., Bazar, N., Shah, D., 2013. The Model of nail color analysis - An application of Digital Image Processing.

Pawar, U., O'Shea, D., Rea, S., O'Reilly, R., 2020. Incorporating Explainable Artificial Intelligence (XAI) to aid the Understanding of Machine Learning in the Healthcare Domain.

Pertuz, S., Ortega, D., Suarez, É., Cancino, W., Africano, G., Rinta-Kiikka, I., Arponen, O., Paris, S., Lozano, A., 2023. Saliency of breast lesions in breast cancer detection using artificial intelligence. Scientific Reports 13, 20545. https://doi.org/10.1038/s41598-023-46921-3

Petsiuk, V., Das, A., Saenko, K., 2018. RISE: Randomized Input Sampling for Explanation of Black-box Models. https://doi.org/10.48550/arXiv.1806.07421

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., Rombach, R., 2023. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. https://doi.org/10.48550/arXiv.2307.01952

Raghavan, K., B, S., Veezhinathan, K., 2023. Attention guided grad-CAM : an improved explainable artificial intelligence model for infrared breast cancer detection. Multimedia Tools and Applications 83, 1–28. https://doi.org/10.1007/s11042-023-17776-7

Rani, D.A.M., Kaleeswari, M., Karpagam, M., 2018. March 2018 Diagnosing Human Diseases using Toe Nail.

Shivhare, I., Jogani, V., Purohit, J., Shrawne, S.C., 2023. Analysis of Explainable Artificial Intelligence Methods on Medical Image Classification, in: 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). pp. 1–5. https://doi.org/10.1109/ICAECT57570.2023.10118312

Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning Important Features Through Propagating Activation Differences. https://doi.org/10.48550/arXiv.1704.02685

Zaitoun, N.M., Aqel, M.J., 2015. Survey on Image Segmentation Techniques. Procedia Computer Science 65, 797–806. https://doi.org/10.1016/j.procs.2015.09.027

Zamzmi, G., Mouton, P., Salekin, S., Sun, Y., Goldgof, D., 2023. Explainable AI for Medical Data: Current Methods, Limitations, and Future Directions. ACM Computing Surveys 57. https://doi.org/10.1145/3637487