Predicting Anemia Using Non-Invasive Machine Learning Techniques

by

Assel Kenzhegariyeva

Submitted to the Department of Computer Science or Data Science
in partial fulfillment of the requirements for the degree of

Master of Science in Data Science

at the

NAZARBAYEV UNIVERSITY

May 2025

Author .....
Department of Computer Science or Data Science
May 8, 2025

Certified by .....
Adnan Yazici
Department Chair of Computer Science
Thesis Supervisor

Accepted by .....
Your Chairman
Dean, School of Engineering and Digital Sciences

Predicting Anemia Using Non-Invasive Machine Learning Techniques

by

Assel Kenzhegariyeva

Submitted to the Department of Computer Science or Data Science
on May 8, 2025, in partial fulfillment of the
requirements for the degree of
Master of Science in Data Science

Abstract

Anemia represents a significant public health problem, especially in young children, where timely identification is essential to prevent serious developmental and health complications. Conventional diagnostic approaches are based on invasive blood tests, which can cause distress in children and are frequently unavailable in low-resource environments.

This research investigates a non-invasive method of detecting anemia through the application of deep learning to images of the conjunctiva, palm, and fingernails, collected from children under five years of age in Ghana. Each modality exhibits distinct strengths and limitations: conjunctival images offer robust predictive features but carry an infection risk, fingernail images are small and challenging to analyze in young children, and palm images, while easy to capture, provide inferior contrast.

To address these challenges, a multimodal fusion model is proposed that uses CNN-based feature extraction, attention mechanisms, and weighted late fusion via XGBoost. Explainability techniques such as SHAP and Grad-CAM are incorporated to enhance transparency and interpretability. To improve generalization to real-world data, a semi-supervised learning approach is introduced, in which confident pseudo-labels from external datasets are merged with the labeled training data to train new models on the combined dataset.

Experimental results demonstrate that the fusion approach significantly outperforms standalone models, achieving 94.22% accuracy and 0.9918 AUC. On a manually collected real-world test set of 30 images, the semi-supervised model achieved 83.3% accuracy and 81.5% F1-score, outperforming the baseline model and improving sensitivity to anemic cases. This study presents a scalable, explainable, and field-ready solution for early anemia screening, suitable for mobile and cost-effective health diagnostics in resource-limited settings.

Thesis Supervisor: Adnan Yazici
Title: Department Chair of Computer Science

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Adnan Yazici, for his invaluable guidance and support. I extend my heartfelt appreciation to my family and friends for their unconditional love, encouragement, and companionship during the challenging phases of my academic journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Anemia is a prevalent global health issue impacting over 1.6 billion individuals, particularly in low and middle-income countries (LMICs) [11]. It is mainly attributed to iron deficiency, genetic disorders, and infectious diseases, which collectively lead to a diminished capacity for oxygen transport in the blood. The prevalence of undiagnosed anemia in children under five is particularly concerning, as it can result in long-term cognitive, physical, and developmental impairments [11]. Early detection is essential for facilitating timely intervention and enhancing health outcomes. Conventional diagnosis often depends on invasive blood tests, including complete blood count (CBC) and hemoglobin (Hb) measurements. However, these methods may be impractical in remote or low-resource settings due to factors such as cost, necessary infrastructure, and cultural resistance, particularly concerning infants or young children [17], [29].

A conventional low-cost method for anemia screening involves the visual evaluation of clinical pallor, specifically examining alterations in the conjunctiva, palms, and fingernails [26]. This method, while straightforward and non-invasive, demonstrates deficiencies in reliability and objectivity, particularly when executed by untrained individuals. Recent research has investigated the application of machine learning (ML) and computer vision techniques to automate and improve the diagnostic process using digital images [7], [4], [8].

A substantial body of research has utilized these visual regions as potential biomarkers for the detection of anemia. Researchers have proposed distinct pipelines for the

analysis of palm [7], conjunctiva [4], and fingernail [8] images. Each modality presents distinct benefits and constraints. Conjunctival images exhibit significant vascular detail; however, they may pose infection risks during close capture [6]. Fingernail images are often small and difficult to segment, especially in pediatric cases, whereas palm images are simpler to obtain but present lower contrast and less distinct features. The observed variations have led to the investigation of multi-modality systems that integrate data from all three image types to enhance diagnostic accuracy [22].

Various publicly accessible datasets, including the Ghana anemia image dataset, have significantly contributed to this advancement. The datasets consist of image samples categorized as anemic or non-anemic according to hemoglobin levels, forming the basis for training predictive models. Initial studies employed handcrafted features like color histograms, GLCM, and texture analysis [7]. In contrast, contemporary approaches utilize deep learning techniques, particularly convolutional neural networks (CNNs) and transfer learning [13].

This research investigates a multi-modal deep learning approach that combines conjunctiva, fingernail, and palm images to enhance the accuracy of anemia detection, addressing the limitations of single-modality methods. This architecture integrates deep convolutional neural network feature extraction, attention-based fusion, and explainable artificial intelligence methodologies. In addition, this study explores a semi-supervised learning strategy based on pseudo-labeling to improve generalization to real-world images. By generating confident pseudo-labels on raw external datasets and combining them with labeled data, the model is retrained from scratch on the merged dataset to reduce confirmation bias and adapt to distribution shifts [10].

The model is initially trained on labeled Ghana data and later retrained on a merged dataset containing pseudo-labeled external samples, enabling semi-supervised learning. The goal is to enhance performance on real-world images where labeled data is limited or unavailable.

Explainable AI (XAI) techniques, including SHAP (Shapley Additive Explanations) and Grad-CAM (Gradient-weighted Class Activation Mapping), will be employed to improve model interpretability, thereby rendering predictions more trans-

parent and comprehensible for healthcare professionals [14], [9], [15]. These tools are especially valuable in medical applications, where model transparency is essential for clinical adoption.

This study aims to develop a scalable, non-invasive anemia screening tool for deployment in low-resource healthcare settings by utilizing multi-modal fusion and explainable AI techniques in order to contribute to the advancement of AI-driven medical diagnostics.

To achieve this goal, the thesis is organized as follows: Chapter 2 reviews the related literature on non-invasive anemia detection and deep learning in medical imaging. Chapter 3 outlines the methodology, including dataset characteristics, preprocessing steps, and model design. Chapter 4 presents the experimental results and explains the model's performance and interpretability. Chapter 5 summarizes key findings and discusses future directions.

Chapter 2

Related works

Recent years have witnessed an increasing focus on non-invasive machine learning methods for anemia detection, utilizing visual indicators from body regions including the conjunctiva, palm, and fingernails. Researchers have investigated single-modality pipelines as well as multimodal fusion architectures. This section provides a comprehensive review of the current literature organized by input modality and methodology.

2.1 Single-Modality Approaches

Numerous studies have suggested models that rely on a singular visual modality, commonly utilizing images of the palm, fingernail, or conjunctiva. A comparative analysis was conducted in [6], evaluating all three modalities through deep learning techniques. The research indicated the efficacy of image-based screening systems, with accuracies ranging from 93% to 96%. In palm-only approaches, [5] employed color and shape descriptors alongside a CNN-based classifier, resulting in a classification accuracy of 99.92%. Fingernail images were analyzed in [28], employing CNN and EfficientNet models to process RGB features, achieving a hemoglobin estimation accuracy of 90.1%.

The conjunctiva serves as a significant modality, providing a clear visual representation of pallor. The authors in [25] utilized ResNet-based feature extraction on conjunctiva images from the Ghana dataset, resulting in an accuracy of 85%. The

Eyes-Defy-Anemia dataset was introduced and assessed in [19], utilizing Hybrid CNN models. In this context, [13] employed MobileNet and ResNet50 in a transfer learning framework, achieving an accuracy of 88%.

Alternative visual indicators have been investigated as well. In [30], researchers conducted an analysis of facial features through videos recorded in emergency departments, utilizing a ResNet + SVM architecture and achieving a classification accuracy of 81.58%. A further practical implementation is detailed in [21], in which the authors developed a smartphone-based system that integrates CNN feature extraction from conjunctiva images with patient demographic data. The hybrid model attained an accuracy of 96.99%, demonstrating the efficacy of combining structured and unstructured data. The NiADA system [12] utilized CNN and Vision Transformers for real-time detection of anemia through conjunctival images obtained with smartphones. The research highlighted the importance of accessibility and usability in low-resource environments, demonstrating robust predictive performance.

While single-modality approaches demonstrate potential, they are frequently limited by the constraints of their specific input type. For example, palm images exhibit low contrast, fingernail images are diminutive and more challenging to capture, and conjunctiva may present hygiene issues in certain clinical settings.

2.2 Multimodal Fusion Approaches

Recent studies have introduced fusion-based models to address the limitations of individual modalities.

In [22], conjunctiva images were combined with electronic health records (EHRs) through a hybrid CNN and Random Forest classifier. A novel Reverse Convolutional Block Attention Module (RCBAM) was proposed to improve attention-based feature learning. The model also utilized a hierarchical multi-scale attention mechanism to enhance discriminative power. This multimodal integration of visual and clinical data achieved a classification accuracy of 95%, demonstrating the effectiveness of combining spatial and channel-level attention techniques.

In [23], palm images were integrated with textual health data (from Kaggle) using a fusion pipeline based on AlexNet with spatial and multiple-channel attention mechanisms. The textual features were transformed into embeddings via a CNN layer and combined with image features at a late-fusion stage. The final model attained 95.8% accuracy and showed significant improvement through spatial attention, enhancing both interpretability and performance.

A major advancement is presented in [18], which introduced the Body-Part Anemia Network (BPANet) trained on conjunctiva, palm, and fingernail images. This model incorporates several innovations: (1) channel-spatial attention for automatic ROI selection, (2) fusion attention module that integrates image features with demographic data (age, gender), and (3) a dual-loss strategy to balance classification and regression tasks. The approach achieved an F1-score of 0.788 and demonstrated generalizability across two real-world datasets (EYES-DEFY-ANEMIA and NTUH), outperforming previous multimodal methods.

These studies collectively highlight how attention-based fusion, dual-loss optimization, and the inclusion of non-image modalities can substantially improve anemia detection in real-world settings.

2.3 Summary and Positioning

The analyzed studies demonstrate the swift advancement of non-invasive anemia detection technologies, transitioning from basic single-modality models utilizing handcrafted features to sophisticated multimodal fusion architectures augmented by deep learning. Although individual modalities have achieved high accuracies, fusion-based methods exhibit enhanced performance and robustness.

Nevertheless, few current systems demonstrate modality-agnostic capabilities, functioning effectively with only one or two image inputs. Moreover, the aspects of interpretability and real-time deployment are not addressed in numerous fusion systems. Additionally, none of the reviewed studies incorporate semi-supervised learning strategies to improve generalization on real-world unlabeled data. The use of confi-

dent pseudo-labeling for model retraining remains largely unexplored in this domain, despite its success in other medical imaging contexts.

This thesis proposes a flexible, attention-based, explainable, and semi-supervised multimodal model that integrates images of the palm, conjunctiva, and fingernails to address these challenges. Table 2.1 presents a comparative summary of related work.

Reference	**Dataset + Modality**	**Feature Extraction**	**Algorithm/Model**	**Accuracy / Metric**	**Year**
Asare et al. [6]	Ghana (Palm, Nail, Conjunctiva)	Color Features	CNN, Naïve Bayes, Decision Tree, k-NN, SVM	89.45–99.12%	2023
Asare et al. [5]	Ghana (Palm)	Color + Shape Analysis	CNN, k-NN, Decision Tree, Naïve Bayes, SVM	99.92%	2023
Viveha et al. [28]	Ghana (Fingernail)	RGB Extraction	CNN, EfficientNet, Ridge Regression	RMSE = 0.65, MAE = 0.624	2024
Singh et al. [25]	Ghana (Conjunctiva)	Color + Textural	ANN, SVM	85%	2023
Muljono et al. [19]	Eyes-Defy-Anemia (Conjunctiva)	Not mentioned	SVM, MobileNetV2	93%	2024
Dimauro et al. [13]	Eyes-Defy-Anemia (Conjunctiva)	HSI, HHR	SVM, KNN, MobileNet	88%	2023
Zhang et al. [30]	Self-collected (Face)	Facial Landmark Extraction	ResNet, DenseNet, EfficientNet, Inception	81.58%	2022
Pallavi et al. [21]	Self-collected (Conjunctiva + Demographics)	Segmentation (UNet), Feature Fusion	ResNet34	96.99%	2024
Semanti et al. [12]	Self-collected (Conjunctiva)	ViT + MobileNet	CNN + Multi-head Attention	Specificity 90%	2024
Ramzan1 et al. [22]	Ghana + EHR (Conjunctiva + Tabular)	ROI, Grad-CAM	RCBAM + MobileNet	95%	2024
Ramzan et al. [23]	Ghana + Kaggle (Palm + Text)	Color Features, Word Embeddings	AlexNet, Multiple Spatial Attention	95.80%	2024
Lin et al. [18]	Eyes-Defy + NTUH (Palm, Conjunctiva, Fingernail)	ROI via ResNet50, Dual Loss	BPANet, Fusion Attention	84.9%	2024

Table 2.1: Comparative overview of anemia detection studies by reference.

In contrast to these prior studies, our proposed approach introduces a modality-agnostic, explainable, and semi-supervised multi-modal fusion framework that allows flexible prediction from one, two, or three image inputs. This dynamic design, combined with SHAP- and Grad-CAM-based interpretability and late fusion via XG-

Boost, addresses both model transparency and real-world usability. Furthermore, it leverages confidence-filtered pseudo-labeling inspired by Cascante-Bonilla et al. [10] to retrain models on a merged dataset, mitigating confirmation bias and improving generalization under domain shift.

To further contextualize our contribution, Table 2.2 presents a focused comparison of existing studies that utilized the Ghana dataset. It contrasts the applied modality, model architecture, fusion strategy, segmentation approach, and explainability technique. Unlike prior works that relied on pre-cropped or single-modality inputs, our method combines multiple Ghana modalities, integrates explainability, and enhances real-world usability through segmentation and semi-supervised learning.

Ref.	**Modality**	**Model / Architecture**	**Fusion Method**	**Explainability**	**Agnostic Input Support**	**Accuracy**
[6]	Palm, Conjunctiva, Fingernail (Single modality)	CNN, SVM, k-NN, Naïve Bayes, Decision Tree	None	No	No	99.12% (Palm)
[5]	Palm	CNN, SVM, Naïve Bayes, k-NN, DT	None	No	No	99.92% (CNN)
[28]	Fingernail	EfficientNet, Ridge Regression	None	No	No	RMSE = 0.659
[25]	Conjunctiva	ANN, SVM	None	No	No	85.0%
[22]	Conjunctiva + EHR (Ghana)	RCBAM-based CNN + Random Forest	Feature-level fusion	Grad-CAM	No	95.0%
[23]	Palm + Text (Ghana + Kaggle)	AlexNet + Spatial Attention	Embedding Fusion	Yes	No	95.8%
Ours	Palm, Conjunctiva, Fingernail	ResNet50, DenseNet121 + XGBoost	Weighted Fusion (late)	SHAP, Grad-CAM	**Yes**	**94.22%**

Table 2.2: Comparison of anemia detection studies using the Ghana dataset.

Chapter 3

Methodology

This chapter presents the methodological framework utilized to create a non-invasive machine learning model for anemia detection through three visual modalities: conjunctiva, palm, and fingernail images. The process includes data preparation, model architecture, training, fusion strategies, and enhancements in interpretability.

3.1 Overview of the Proposed Architecture

The proposed system follows a modular, multi-stage machine learning pipeline designed to support flexible, image-based anemia screening. As shown in Figure 3-1, the pipeline consists of three CNN-based feature extractors, each optimized for a specific modality (palm, conjunctiva, and fingernail). These CNNs are trained using modality-specific attention mechanisms and loss functions to enhance representation learning. Extracted features from each modality are passed into independent XGBoost classifiers trained to predict anemia status. The final prediction is derived using a weighted fusion mechanism based on validation performance. The system supports agnostic input combinations, allowing predictions from one, two, or all three image modalities. Post-hoc interpretability is provided through SHAP and Grad-CAM, enabling both feature-level and spatial explanations. To improve real-world generalization, a semi-supervised learning strategy was applied using confident pseudo-labels derived from external unlabeled datasets.

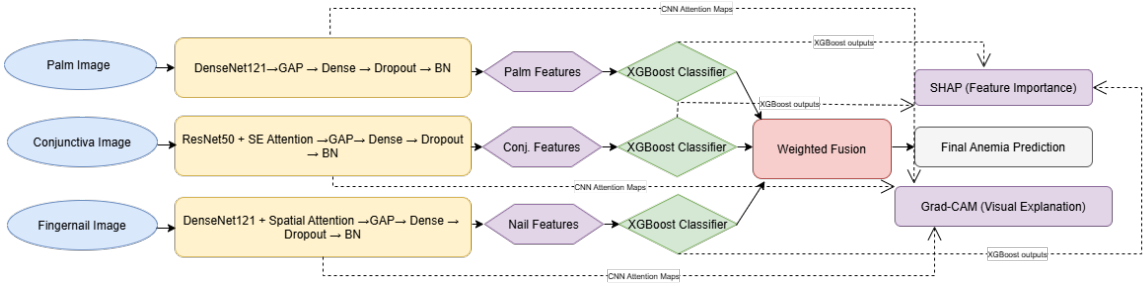


Figure 3-1: Overview of the proposed architecture

3.2 Dataset Preparation and Preprocessing

This section introduces the dataset origin, structure, and modality-specific preprocessing steps used to prepare the input for training.

This study used a publicly available dataset from Ghana, comprising annotated images of three anatomical regions: conjunctiva, palm, and fingernail [4, 8, 17]. Each image was labeled as anemic or non-anemic based on clinically validated hemoglobin levels. The dataset was already pre-segmented to focus on the region of interest (ROI) for each modality, simplifying the classification task. Images were stored in structured directories categorized by class and modality to ensure reproducibility.

Figures 3-2–3-4 show representative samples for each modality, highlighting the anatomical differences across classes.

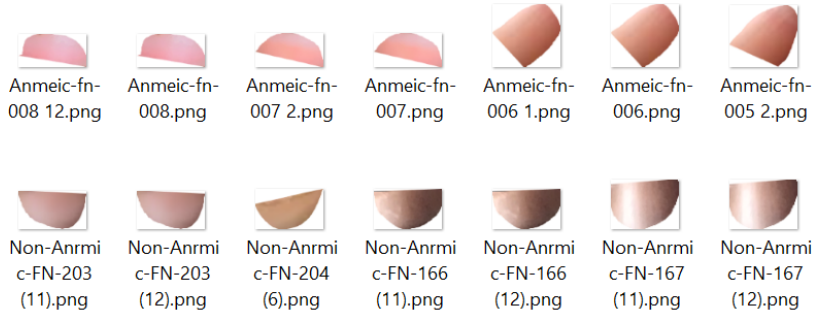


Figure 3-2: Sample images from the fingernail dataset.

All images were resized to 224×224 pixels with padding to preserve aspect ratio and avoid geometric distortion. Dynamic preprocessing was applied during data loading using a custom Keras pipeline.

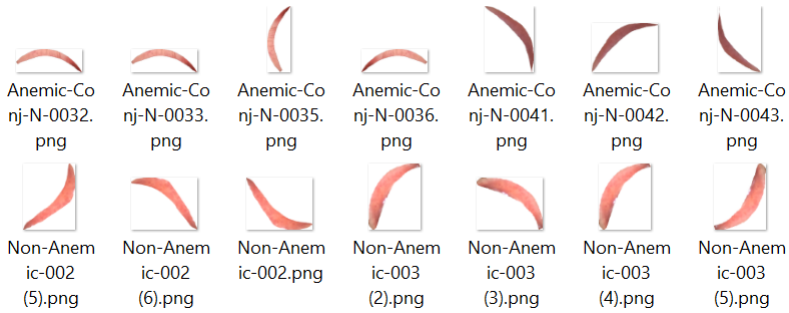


Figure 3-3: Sample images from the conjunctiva dataset.

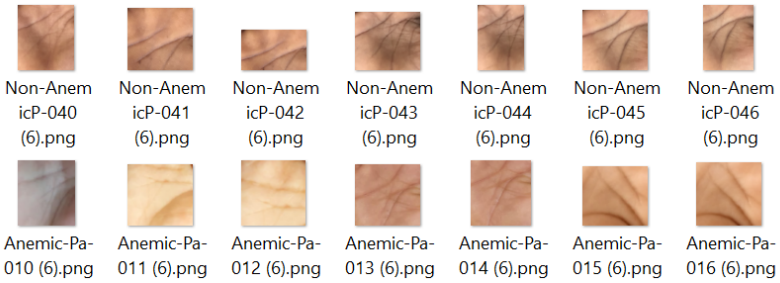


Figure 3-4: Sample images from the palm dataset.

To enhance contrast in poorly illuminated images, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to all input samples. This technique has been shown to improve visualization of blood vessel regions in conjunctiva images, which are critical for anemia detection [27].

A modality-specific ROI masking strategy was also used. For conjunctiva and fingernail images, grayscale thresholding followed by morphological dilation was used to extract biologically relevant areas. Palm images were used in full due to their minimal background interference.

To normalize brightness across samples, percentile-based intensity normalization was applied between the 2nd and 98th percentiles. This method has been successfully used in medical imaging, including CT scan segmentation, to reduce the effect of outliers and illumination variability [24].

After preprocessing, the dataset was split into training (70%), validation (15%), and test (15%) sets with class balance preserved. Table 3.1 summarizes the distribution per modality.

Note on Fusion Assumption: Since the palm, conjunctiva, and fingernail im-

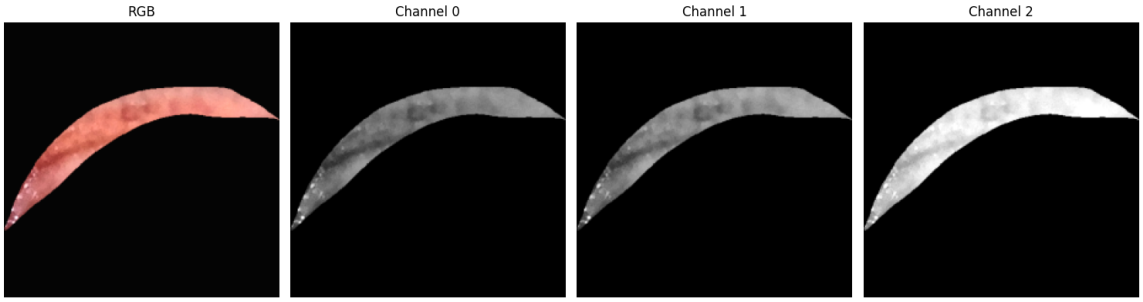


Figure 3-5: Preprocessed conjunctiva sample showing ROI mask and RGB channel separation.

Modality	Anemic (Train)	Non-Anemic (Train)	Anemic (Val)	Non-Anemic (Val)	Anemic (Test)	Non-Anemic (Test)
Conjunctiva	1796	1199	386	257	386	258
Palm	1792	1188	385	255	385	255
Fingernail	1795	1185	385	255	385	255

Table 3.1: Dataset split summary for each modality after preprocessing.

ages are not patient-aligned, a synthetic fusion input was generated by combining the i-th image from each modality during testing. While not representative of clinical deployment, this synthetic alignment allows controlled and balanced multimodal evaluation during experimentation. This method is commonly used in multi-source data fusion studies when subject-wise correspondence is unavailable.

3.3 Segmentation Pipeline in Practical Applications

To support real-world deployment, this section describes how raw images were segmented into relevant regions of interest (ROI) for conjunctiva, palm, and fingernail modalities. Although the Ghana dataset used for model training is already segmented and cleaned, real-world images—such as those captured by mobile devices—typically contain substantial background noise and variability. To simulate these conditions and prepare a deployment-ready pipeline, automated segmentation techniques were implemented and validated using external datasets.

Three publicly available datasets were used for this purpose, as summarized in Table 3.2. These datasets contain raw images of hands, eyes, and fingernails, serving as proxies for real-world conditions.

Dataset	Sample Size	Modality	Source (Citation)
Hand and Palm Images Dataset	11,076 images	Palm	Kaggle dataset [2]
Eye Conjunctiva Segmentation Dataset	547 images	Conjunctiva	Mendeley dataset [1]
Nail Disease Image Classification Dataset	2,265 images (healthy only)	Fingernail	Kaggle dataset [3]

Table 3.2: External datasets used for segmentation of palm, nail, and eye images.

For each modality, a dedicated segmentation model was developed or adapted. Initially, a subset of each dataset was manually annotated to define ROI boundaries. These annotations were used to train segmentation models using Roboflow's AutoML tools, which support both bounding box and polygon mask formats. The trained models were then deployed via API to segment the remaining images. After inference, the predicted ROI was extracted and resized to 224×224 pixels with padding to preserve aspect ratio and avoid geometric distortion. This resizing step ensures compatibility with the CNN classifier input dimensions and supports robust downstream processing.

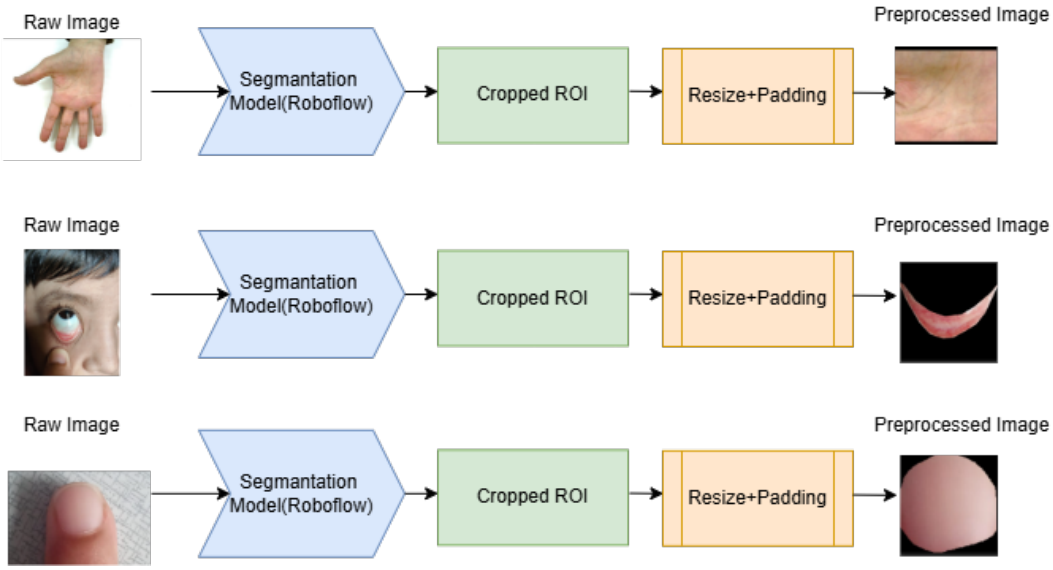


Figure 3-6: Segmentation and preprocessing workflow for palm, conjunctiva, and fingernail modalities.

Figure 3-6 illustrates the end-to-end workflow used to process raw images. This

includes input acquisition, segmentation prediction, ROI cropping, padding, and final image resizing.

This segmentation pipeline ensures that the proposed anemia screening system can be extended to operate on unsegmented real-world images, enabling eventual integration into mobile-based clinical screening applications.

3.4 Architecture and Training of the CNN Model

This section details the specific architectures, training configurations, and loss functions used to train the CNN classifiers for each modality.

Separate convolutional neural networks were developed for each modality, leveraging transfer learning from ImageNet-pretrained backbones. DenseNet121 was used for palm and fingernail modalities, while ResNet50 was selected for the conjunctiva modality due to its superior performance in previous related studies. Each network was modified to include a global average pooling layer, followed by a dense layer with 256 ReLU-activated units. Dropout and batch normalization layers were incorporated for regularization.

Figure 3-7 illustrates the complete CNN pipelines for each modality, including the input image, the backbone model, attention blocks (if used), and the extracted feature vector.

To enhance modality-specific representation, attention mechanisms were selectively applied. The conjunctiva model integrated a Squeeze-and-Excitation (SE) block to emphasize the most informative channels [16], while the fingernail model employed spatial attention using a learnable 1×1 convolution followed by a sigmoid activation [31]. The palm model, in contrast, achieved competitive results without additional attention layers, relying on the representational strength of DenseNet121 and global average pooling.

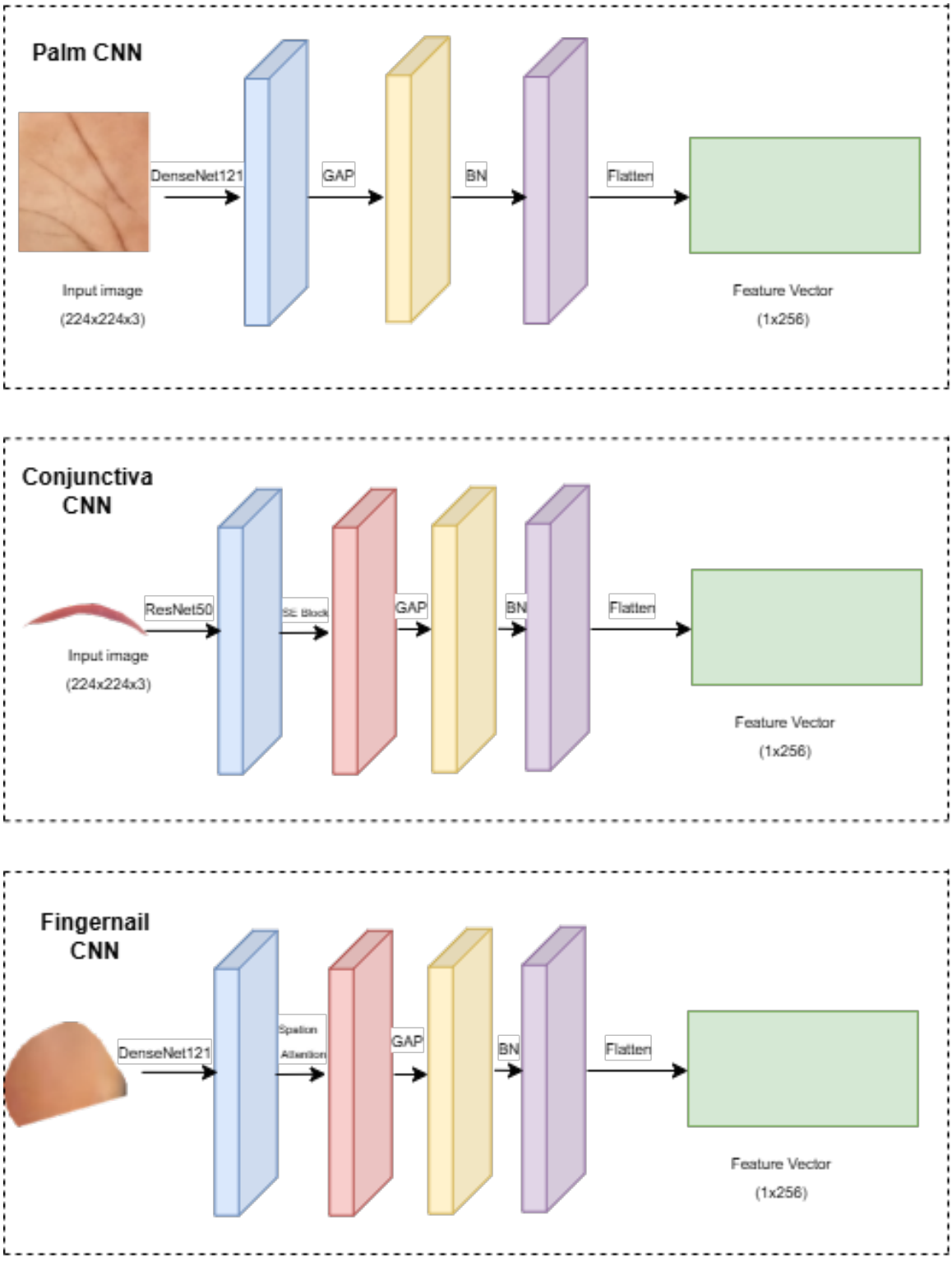


Figure 3-7: CNN architectures per modality.

Parameter	Conjunctiva	Palm	Fingernail	Value (Shared)	Notes
Base Model	ResNet50	DenseNet121	DenseNet121	–	Transfer learning from ImageNet
Attention	SE Block	None	Spatial Attention	–	Custom per modality
Loss Function	Focal Loss	Binary Cross-Entropy	Binary Cross-Entropy	–	Class imbalance handling
Optimizer	Adam	Adam	Adam	–	Adaptive gradient descent
Learning Rate	5e-5	1e-4	7e-5	ReduceLROnPlateau	Scheduled decay
Dropout	0.3	0.5	0.5	–	Regularization
Epochs	20	20	20	Shared	With early stopping
Batch Size	32	32	32	Shared	Based on memory constraints

Table 3.3: Training configuration and hyperparameters used per modality.

As summarized in Table 3.3, all CNNs were trained as binary classifiers using sigmoid activation. Class imbalance was addressed by computing class weights dynamically for each training fold. Early stopping was employed based on validation AUC to prevent overfitting, and ReduceLROnPlateau was used to adjust learning rates dynamically.

The conjunctiva model used Focal Loss to mitigate the impact of low signal-to-noise ratio and data imbalance, while the palm and fingernail models used binary cross-entropy loss. Training was performed for up to 20 epochs. The feature vector extracted from the final Batch Normalization layer of each CNN was saved for downstream fusion using XGBoost classifiers, as described in the following section.

3.5 Integration of Feature Fusion and Multi-Modal Approaches

Following feature extraction, the fixed-length vectors from each CNN model were passed into independent XGBoost classifiers. The feature vectors were first standardized using the `StandardScaler` function from the `sklearn.preprocessing` mod-

ule to ensure zero-mean, unit-variance distributions for each modality. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) from the `imblearn.over_sampling package` was applied, generating synthetic minority class samples.

Each modality-specific standardized and balanced feature set was then used to train an `XGBClassifier` from the `xgboost` package. The XGBoost models were configured with 200 estimators, a maximum depth of 6, a learning rate of 0.08, and GPU acceleration using the `tree_method="hist"` and `device="cuda"` settings. XGBoost was selected for its robustness in handling imbalanced, tabular data and its ability to model non-linear decision boundaries efficiently. In preliminary testing, it outperformed dense neural network heads when trained on fixed-length feature vectors.

During inference, for each modality, the trained XGBoost models outputted probability scores for the anemic class using the `predict_proba()` method. Specifically, the second column of the predicted probability array, corresponding to the positive (anemic) class, was extracted for further fusion.

To integrate predictions from multiple modalities, a weighted averaging strategy was employed. The final classification probability was calculated as:

$$P_{\text{final}} = w_1 P_{\text{conjunctiva}} + w_2 P_{\text{palm}} + w_3 P_{\text{fingernail}} \tag{3.1}$$

where $P_{\text{conjunctiva}}$, P_{palm}, and $P_{\text{fingernail}}$ represent the predicted probabilities for the anemic class from each modality-specific XGBoost model. The weights w_1, w_2, and w_3 were derived from the validation accuracies achieved during the training of each modality and normalized such that $w_1 + w_2 + w_3 = 1$. This weighting approach ensures that modalities demonstrating stronger validation performance contribute more significantly to the final decision, thereby enhancing overall robustness.

This modular and flexible design allows the system to operate agnostically, making predictions with one, two, or three modalities depending on the available inputs. The weighted fusion method thus supports real-world deployment scenarios where not all image types may be captured.

The fusion ensemble was evaluated through five-fold cross-validation, using metrics such as accuracy, AUC, and F1-score. Detailed results, including ablation comparisons among alternative fusion strategies (early fusion and multi-modal CNN fusion), are presented in Chapter 4. These experiments justify the selection of weighted fusion as the optimal approach.

3.6 Explainability and Agnostic Prediction

This final methodology section explains how interpretability tools were used to understand and validate model decisions. SHAP and Grad-CAM were incorporated into the evaluation pipeline to improve model interpretability. SHAP was employed to evaluate the contribution of each CNN-derived feature to the final prediction made by the XGBoost classifier, providing insight into model reasoning across modalities [22], [23]. Summary statistics and bar plots were created to illustrate the impact of various features on predictions across different modalities. Grad-CAM was utilized to visualize the spatial regions focused on by each CNN during prediction[21]. Heatmaps were produced from the final convolutional layer and evaluated in relation to the ground truth ROI masks. Metrics including Intersection over Union (IoU), Confidence Drop, and Attention Shift were calculated and are detailed in Chapter 4 (Results and Evaluation), as they stem from post-training analysis aimed at quantifying the alignment between model attention and biological relevance. The system was ultimately evaluated under agnostic conditions, employing various combinations of available modalities. The model exhibited robust performance in all scenarios, including single-modality input, thereby confirming its adaptability to incomplete or variable input in practical applications. This modular and interpretable architecture facilitates high accuracy and flexibility in deployment, rendering it appropriate for low-resource and mobile-based screening contexts.

3.7 Semi-Supervised Learning Pipeline

To improve the model's generalization on real-world inputs and address the scarcity of annotated medical images, a semi-supervised self-training framework was implemented [20]. This strategy leveraged a small set of labeled samples from the Ghana dataset alongside a larger pool of unlabeled images from external datasets (Mendeley and Kaggle). The method follows a classical pseudo-labeling workflow with enhancements through modality-specific confidence filtering.

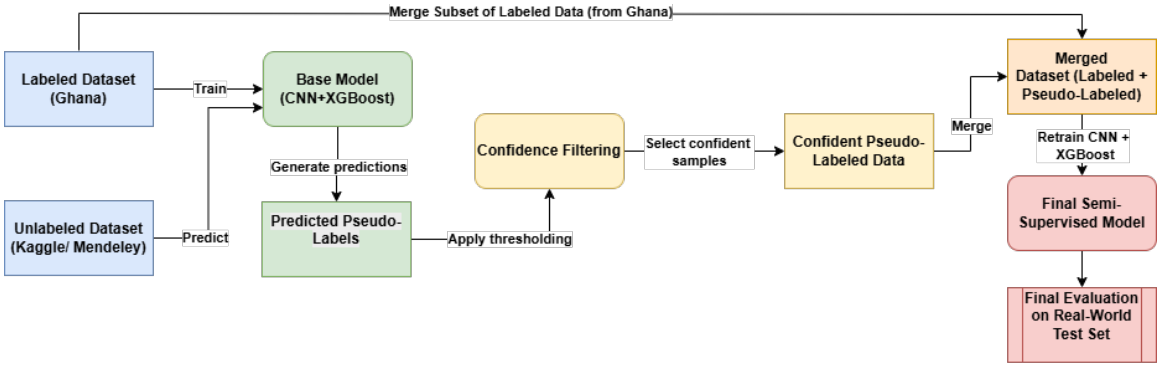


Figure 3-8: Semi-supervised self-training pipeline implemented for real-world refinement.

As illustrated in Figure 3-8, the pipeline is structured into five key steps:

1. Supervised Training on Labeled Data

CNN feature extractors and XGBoost classifiers were initially trained on the labeled Ghana dataset for all three modalities—palm, conjunctiva, and fingernail. These models formed the base system, achieving high validation performance, and were used to infer predictions on new, unseen data.

2. Pseudo-Label Generation on Unlabeled Data

External images were segmented and preprocessed using the segmentation pipeline (Section 3.3) and passed through the trained models. This produced probability-based predictions, which served as pseudo-labels for the unlabeled data.

3. Confidence Filtering

To reduce the risk of noisy labels, predictions were filtered using empirically defined confidence thresholds based on labeled test set statistics (Table 4.8 in Chapter 4): 0.9 for palm and fingernail, and 0.7 for conjunctiva. Only samples with predicted confidence above these thresholds were retained, following the principle of curriculum learning, where simpler (more confident) samples are learned first [10].

4. Merging Labeled and Pseudo-Labeled Data

The high-confidence pseudo-labeled samples were combined with a 50% subset of the original labeled Ghana dataset to form a merged training set. This balanced mix was used to increase data diversity and adapt the model to a broader distribution, while still grounding learning in clinically verified labels.

5. Retraining and Evaluation

New CNNs were trained from scratch using the merged dataset. Feature vectors were re-extracted and used to retrain XGBoost classifiers. Fusion weights were also recalculated based on the updated validation performance. The resulting semi-supervised model was finally evaluated on a manually collected real-world test set to assess its robustness under domain shift conditions.

This strategy demonstrated that confident pseudo-labeling and partial retraining can enhance model generalization in real-world settings while minimizing annotation effort.

3.8 Evaluation Metrics

To quantitatively evaluate the model performance, standard classification metrics were used:

The performance of the proposed model was evaluated using five commonly used classification metrics: Accuracy, Precision, Recall, F1-Score, and Area Under the

ROC Curve (AUC). Their mathematical definitions are provided below:

- **Accuracy**: Measures the proportion of correctly predicted samples among all samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.2}$$

- **Precision**: Measures the proportion of positive predictions that were correct.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.3}$$

- **Recall (Sensitivity)**: Measures the proportion of actual positives that were correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.4}$$

- **F1-score**: Harmonic mean of Precision and Recall.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.5}$$

- **ROC-AUC**: Represents the area under the Receiver Operating Characteristic curve; used to evaluate classification performance independent of threshold.

These metrics are reported in Chapter 4 across single-modality and multi-modality models under various configurations.

Chapter 4

Results and Evaluation

This chapter presents the experimental results of the proposed non-invasive anemia detection system. The evaluation covers model performance across three individual modalities (palm, conjunctiva, and fingernail), multi-modal fusion, interpretability analysis using SHAP and Grad-CAM, and robustness under agnostic input scenarios. All experiments were conducted using the preprocessed and balanced dataset described in Chapter 3.

4.1 Training Performance per Modality

Each CNN model was trained for 20 epochs using early stopping and learning rate scheduling. Figures 4-1, 4-2, and 4-3 illustrate training curves for accuracy, AUC, and loss across epochs for the palm, conjunctiva, and fingernail models respectively.

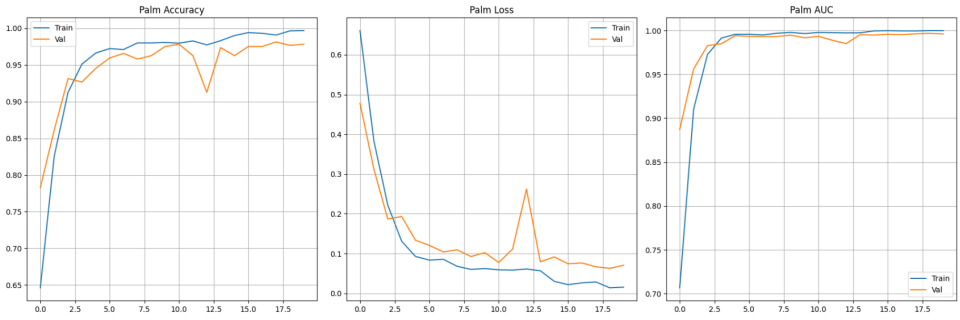


Figure 4-1: Palm CNN training curves (accuracy, AUC, loss)

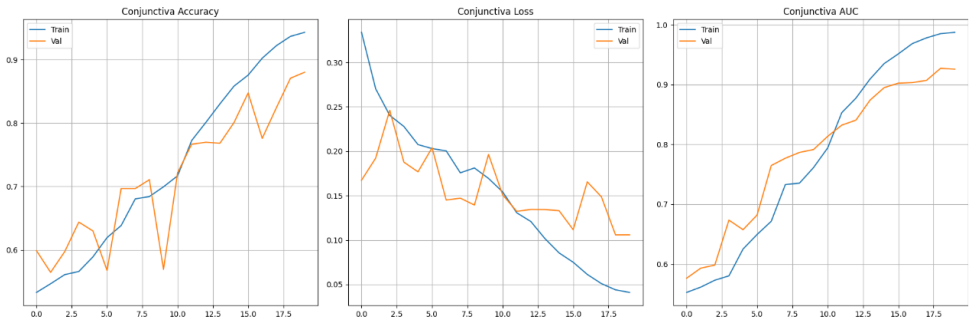


Figure 4-2: Conjunctiva CNN training curves (accuracy, AUC, loss)

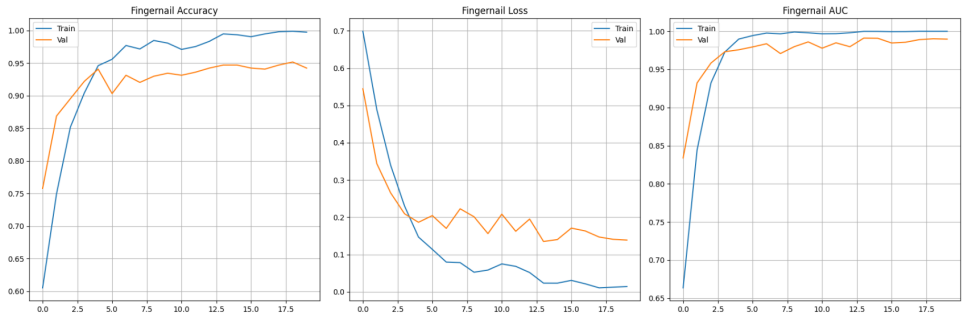


Figure 4-3: Fingernail CNN training curves (accuracy, AUC, loss)

4.2 Evaluation Metrics per Modality

The trained CNN models were evaluated on their respective test sets using common classification metrics. Table 4.1 summarizes the results. These scores reflect each anatomical region's independent ability to support non-invasive anemia detection.

Modality	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**ROC-AUC**
Palm	0.97	0.97	0.98	0.98	0.997
Conjunctiva	0.88	0.91	0.85	0.88	0.927
Fingernail	0.94	0.96	0.94	0.95	0.986

Table 4.1: Evaluation metrics per modality on test set.

As shown in Figure 4-4, the palm model demonstrated the highest classification reliability with only 31 misclassified cases, while the conjunctiva model had a higher false negative rate. This aligns with previous analysis of model interpretability and texture clarity per modality.

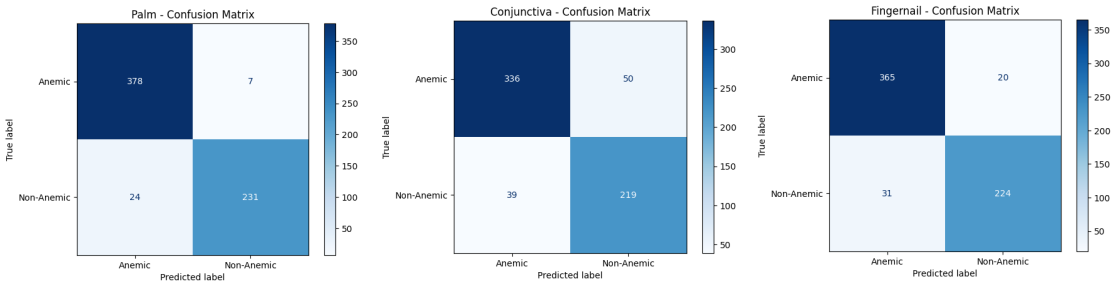


Figure 4-4: Confusion matrices for Palm, Conjunctiva, and Fingernail CNN classifiers.

4.3 Multi-Modal Fusion Performance

A feature-level weighted fusion strategy using XGBoost classifiers was evaluated under 5-fold cross-validation. Metrics include classification accuracy, AUC, and F1-score for both classes across folds.

Fold	Accuracy	AUC	F1 (Anemic)	F1 (Non-Anemic)
1	0.9406	0.9518	0.9112	0.9544
2	0.9453	0.9583	0.9247	0.9559
3	0.9420	0.9537	0.9165	0.9513
4	0.9448	0.9561	0.9210	0.9551
5	0.9383	0.9492	0.9098	0.9485

Table 4.2: 5-Fold cross-validation results for weighted fusion using XGBoost (aligned with test set performance).

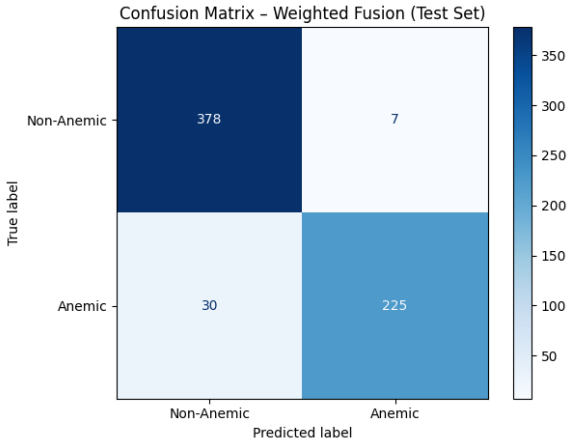


Figure 4-5: Confusion matrix of the final weighted fusion model on the test set.

Figure 4-5 presents the confusion matrix of the final weighted fusion model tested on the Ghana test set. The model achieved high specificity (378/385 non-anemic correctly classified) and strong sensitivity (225/255 anemic correctly classified), resulting in a total test accuracy of 94.22%. These results confirm the model's robustness and practical suitability for deployment in low-resource environments, where reliable non-invasive anemia screening is essential

4.4 Explainability with SHAP

SHAP (SHapley Additive Explanations) was used to interpret the output of the XGBoost classifiers. For each modality, the top 10 most influential features were extracted and visualized.

The plots Figure 4-6 identify which CNN-derived features most influenced the XGBoost classifier, supporting interpretability and clinical insight. SHAP interpretations reveal that features derived from finger creases in palm images, color variation in conjunctiva, and nail curvature influenced the classifier's decisions.

These visual explanations match domain knowledge and highlight the trustworthiness of the model.While SHAP assigns importance to high-dimensional CNN embeddings, the model's behavior can be better understood when combined with visual evidence from Grad-CAM. For palm images, high SHAP values likely correspond to vascular patterns and skin fold regions — areas typically associated with clinical pallor. In fingernail images, important features appear to be influenced by nail bed contour and brightness, aligning with expected discoloration patterns. The conjunctiva modality, however, showed less consistent SHAP attribution, possibly due to segmentation uncertainty or weaker texture signals. These connections strengthen the trust in the model's reliance on relevant anatomical features.

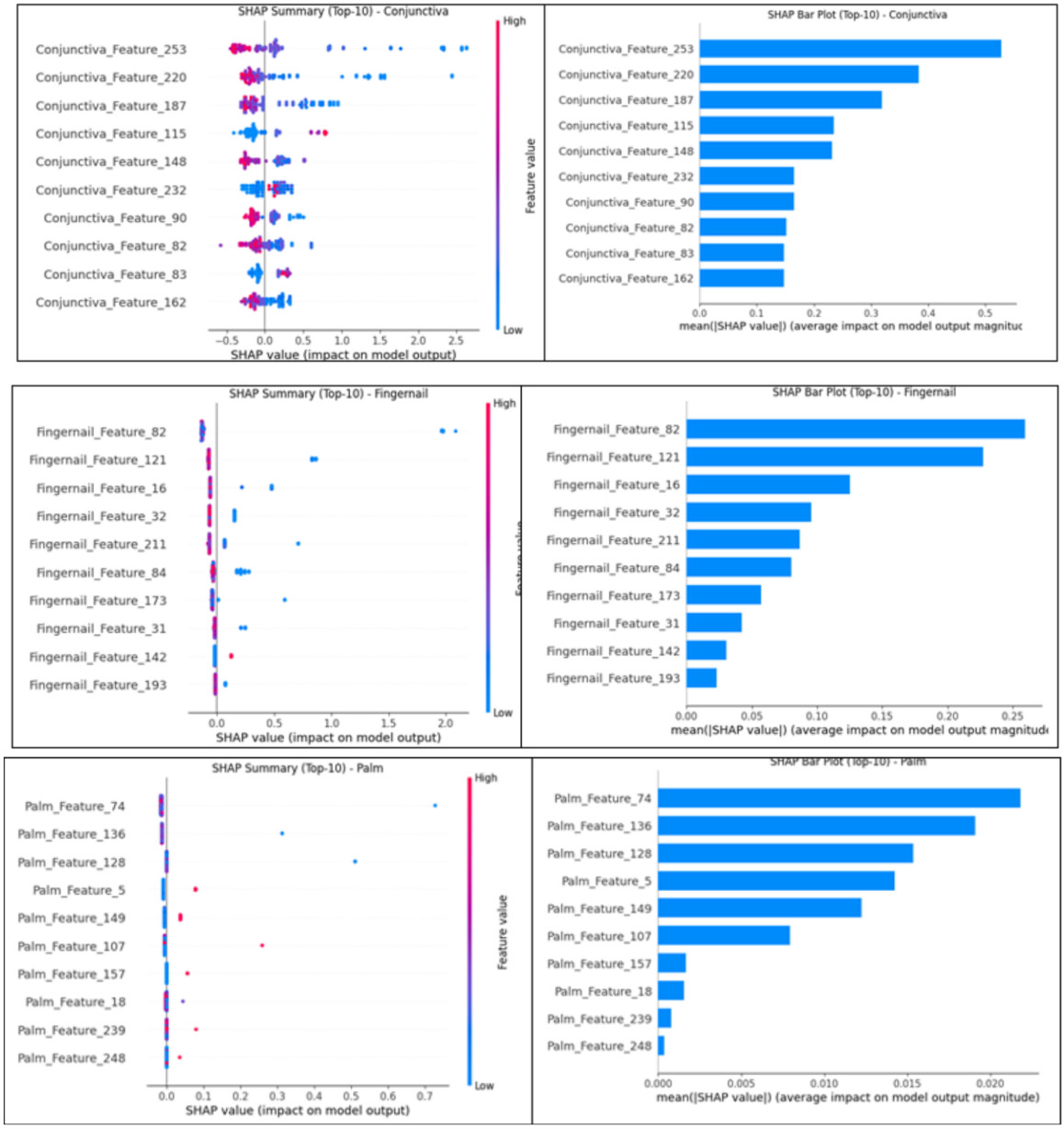


Figure 4-6: SHAP visual explanations for conjunctiva, fingernail, and palm modalities.

4.5 Explainability with Grad-CAM

To enhance the interpretability of the CNN predictions, Grad-CAM (Gradient-weighted Class Activation Mapping) was employed across all three modalities—palm, conjunctiva, and fingernail. This technique generates heatmaps that reveal which spatial regions of an input image most influenced the model's decision. It does so by computing the gradients of the target class with respect to the final convolutional layer of the CNN, effectively highlighting important areas.

Figure 4-7 presents example Grad-CAM visualizations for each modality. Each row shows the original image, the heatmap, and the overlaid attention map along with the predicted class probability. These examples demonstrate that CNNs attend to clinically relevant regions like nail beds and palmar creases during decision-making.

Palm images exhibited strong and diffuse activations across vascular regions, with heatmaps focused on textured skin folds—regions commonly used in clinical assessments of pallor. The prediction was made with 99.0% confidence, showing strong model certainty.

Fingernail images demonstrated confident focus over the entire nail bed, a key area for assessing anemia-related discoloration. The model showed 100% confidence for this non-anemic case, and the attention was well aligned with the expected anatomical region.

Conjunctiva images, however, showed more localized and uncertain attention, with low confidence (34.7%) in predicting an anemic case. The heatmap was narrow and focused on a small segment, which may indicate difficulty in generalizing conjunctival features due to low contrast or segmentation limitations.

To assess the alignment between model attention and biological relevance, quantitative evaluation was conducted using three interpretability metrics:

- **Intersection over Union (IoU)**: Measures the overlap between the Grad-CAM heatmap and the manually annotated ROI mask.
- **Confidence Drop**: Difference in model confidence when the heatmap region is masked out. A larger drop indicates higher dependence on the highlighted

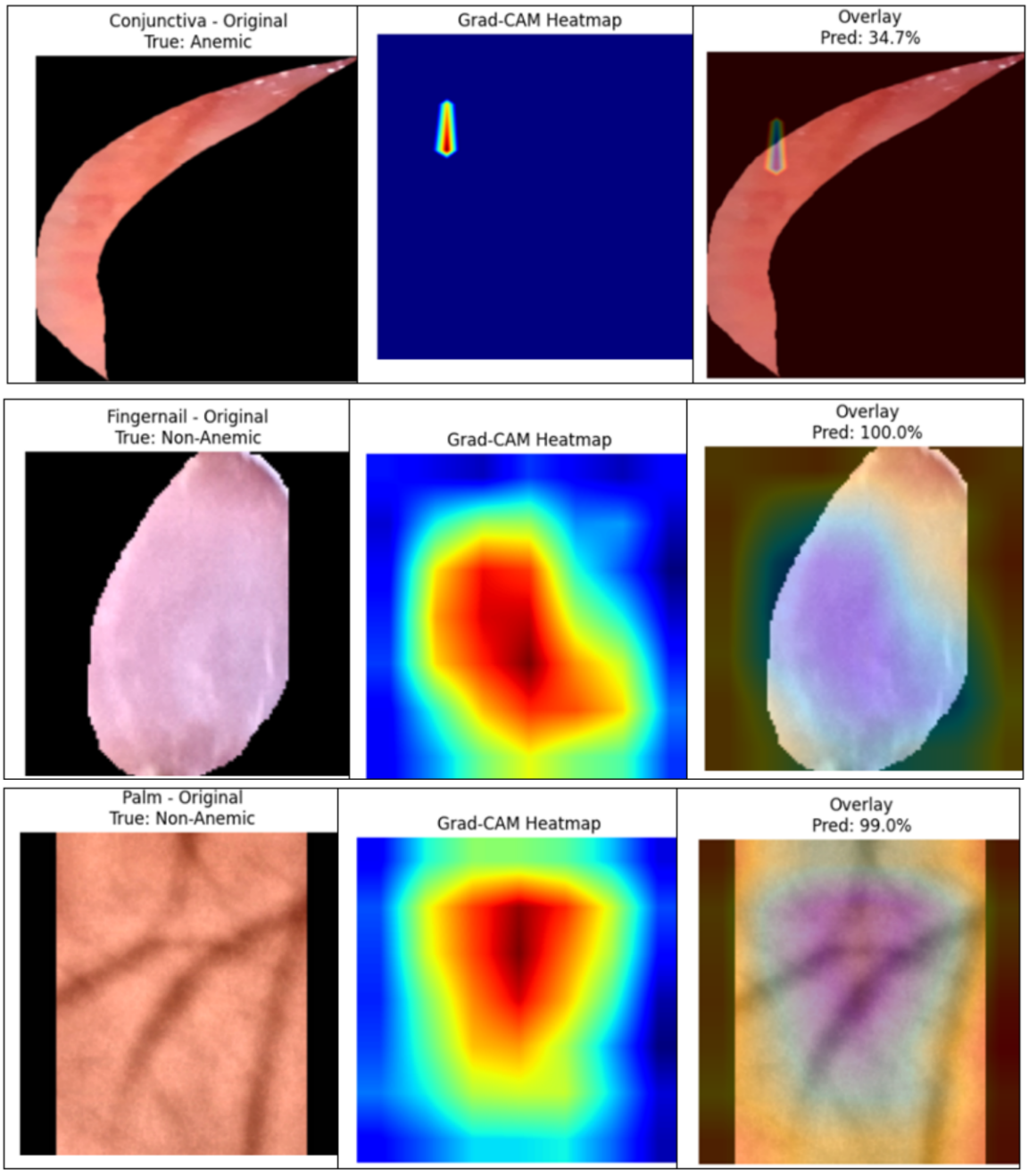


Figure 4-7: Grad-CAM visualizations for 3 modalities.

region.

- **Attention Shift**: Average pixel distance between the Grad-CAM heatmap centroid and the center of the ROI mask.

Modality	IoU	Confidence Drop	Attention Shift (px)
Palm	0.1871	**-0.0045**	32.22
Conjunctiva	**0.0118**	0.0045	N/A
Fingernail	0.0250	0.0031	45.00

Table 4.3: Grad-CAM interpretability metrics per modality (average over 10 test samples)

The palm modality achieved the highest IoU score, indicating strong spatial agreement between Grad-CAM and the biologically relevant ROI. Additionally, the near-zero confidence drop for palm (-0.0045) suggests model robustness, as predictions remained stable even when the attention region was masked.Despite effective attention alignment in palm and fingernail modalities, the conjunctiva model showed weak localization performance, as indicated by a very low IoU (0.0118). This suggests the CNN struggled to focus on medically meaningful regions in the eye, possibly due to low contrast, inconsistent ROI extraction, or lack of texture diversity in the conjunctiva dataset. This highlights the need for stronger segmentation models or higher-quality conjunctival image data in future work.

Overall, these results indicate that the CNN models are learning to focus on modality-specific regions that are consistent with medical expectations. The palm model, in particular, exhibited the most interpretable and stable attention behavior. Grad-CAM thus reinforces the validity of the model decisions and provides essential transparency for clinical adoption.

4.6 Ablation Studies

To better understand the contribution of each design decision, this chapter presents three targeted ablation studies. These experiments focus on (1) fusion strategy design,(2) CNN backbone selection and Agnostic Inference Evaluation (3) as they were

central to the iterative development and optimization of the proposed anemia detection model.

4.6.1 Fusion Strategy

Three fusion strategies were tested to determine the most effective method for combining modality-specific predictions. As shown in Table 4.4, weighted fusion using XGBoost outperformed both early concatenation and multi-modal CNN fusion. Each configuration used the same CNN feature extractors, altering only the method of combining outputs.

Method	Fusion Type	Accuracy	AUC	F1 (Anemic)	F1 (Non-Anemic)
Baseline	Early Fusion (Concat)	89.75%	0.9410	0.87	0.91
Ablation 1	Multi-Modal CNN Fusion	91.83%	0.9574	0.90	0.93
Proposed	**Weighted Fusion (XGBoost)**	**94.22%**	**0.9918**	**0.92**	**0.95**

Table 4.4: Ablation study of fusion strategies: Comparison of performance using different fusion methods.

Weighted fusion demonstrated significantly better generalization due to its modular architecture and ability to handle modality-specific classifier strengths. It also enabled interpretability through SHAP and robustness under agnostic input conditions.

4.6.2 CNN Backbone Selection

Each modality was evaluated using multiple CNN architectures to determine the best-performing model per anatomical region. DenseNet121 consistently yielded superior results for palm and fingernail images due to its deeper, feature-reuse-friendly architecture. For conjunctiva, ResNet50 with a Squeeze-and-Excitation (SE) block performed better than alternatives. Table 4.5 summarizes the backbone comparison by modality.

Modality	CNN Backbone	Accuracy	AUC	Notes
Palm	DenseNet121	95.47%	0.9868	Best overall performance among palm CNNs
Palm	EfficientNetB0	93.28%	0.97	Slightly weaker than DenseNet121
Conjunctiva	ResNet50 + SE	88.12%	0.9187	SE attention improved representation
Conjunctiva	ResNet50 (no SE)	85.11%	0.91	Lower generalization, weak region focus
Fingernail	DenseNet121 + Spatial Attention	90.16%	0.9798	Spatial cues improved small ROI feature extraction
Fingernail	EfficientNetB0	87.19%	0.975	Weaker performance on noisy background

Table 4.5: Comparison of CNN backbones per modality with accuracy, AUC, and architectural observations.

4.6.3 Agnostic Inference Evaluation

To test the model's robustness in real-world deployment scenarios, inference was performed with different combinations of available modalities. Table 4.6 summarizes the accuracy and AUC values under each modality combination.

Modalities	Accuracy	ROC-AUC
Palm only	0.9301	0.9852
Conjunctiva only	0.8120	0.8903
Fingernail only	0.8250	0.9617
Palm + Conjunctiva	0.9485	0.9874
Palm + Fingernail	0.9441	0.9896
Conjunctiva + Fingernail	0.8653	0.9522
All (3 modalities)	0.9422	0.9918

Table 4.6: Agnostic modality combinations and their classification performance

The model demonstrated high performance even with one or two modalities missing, confirming its agnostic design capability and real-world deployment readiness.

4.7 Segmentation Model Training and Evaluation

To support real-world deployment, segmentation models were trained using Roboflow for each modality using YOLOv8 object detection pipelines. Performance was evaluated using mean Average Precision (mAP), precision, and recall metrics.

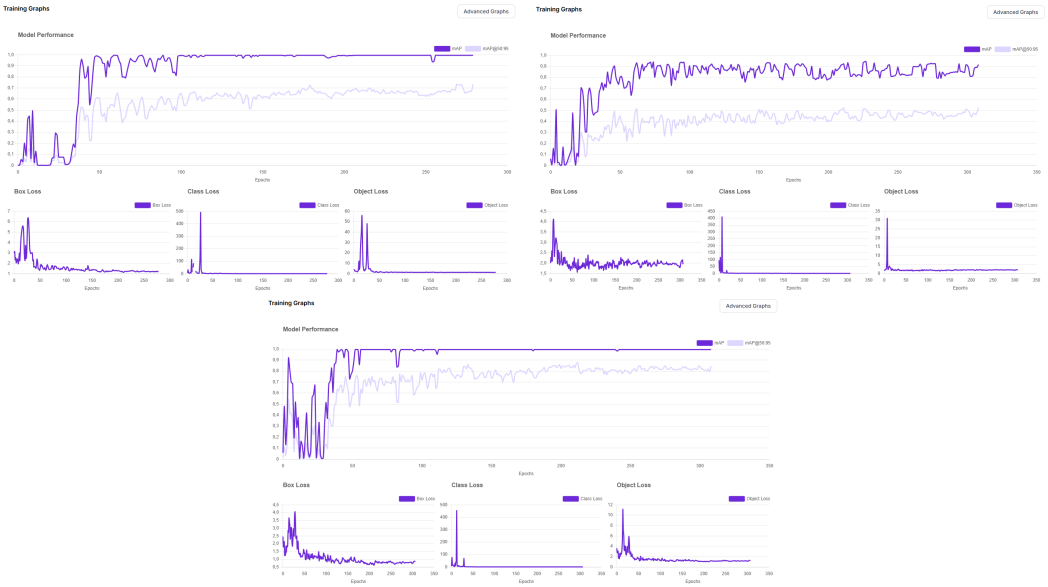


Figure 4-8: Training performance (mAP and loss curves) for palm, conjunctiva, and fingernail segmentation models.

Modality	**mAP@50**	**Precision**	**Recall**
Palm	99.5%	100%	99.9%
Conjunctiva	91.4%	86.6%	86.4%
Fingernail	99.5%	99.4%	100%

Table 4.7: Roboflow segmentation performance metrics: mAP@50, precision, and recall per modality.

Table 4.7 summarizes the segmentation model performance across all three modalities, reporting mAP@50, precision, and recall as computed by the Roboflow training pipeline. The palm segmentation model achieved the highest performance with mAP@50 of 99.5%, precision of 100%, and recall of 99.9%. The fingernail model also performed strongly (mAP@50 = 99.5%), while the conjunctiva segmentation was comparatively weaker with mAP@50 of 91.4% and lower recall. These findings explain the reduced localization confidence in the conjunctiva CNN model observed in Grad-CAM analysis.

4.8 Semi-Supervised Learning Results

To improve model generalization and utilize additional unlabeled data, a semi-supervised learning pipeline was implemented using confidence-filtered pseudo-labeling. This section presents the outcomes of each major step described in the methodology.

4.8.1 Confidence Estimation on Labeled Test Set

To define thresholds for pseudo-labeling, we computed average predicted probabilities (confidence scores) from CNN classifiers on the labeled Ghana test set. For each modality, we calculated the mean confidence for both classes and their standard deviation. These results are shown in Table 4.8.

Modality	Avg Confidence	Anemic Avg	Non-Anemic Avg	Std Dev
Palm	0.9622	0.9561	0.9715	0.0899
Conjunctiva	0.7497	0.7264	0.7846	0.1473
Fingernail	0.9709	0.9751	0.9646	0.0823

Table 4.8: Confidence statistics of CNN predictions for each modality on the Ghana test set

Figure 4-9 visualizes these confidence distributions, reinforcing that palm and fingernail models produce high-confidence predictions, while conjunctiva exhibits more variability.

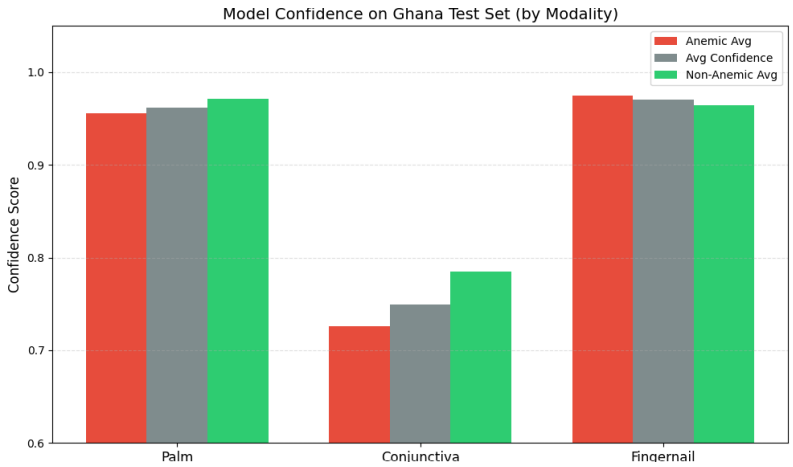


Figure 4-9: Average model confidence on labeled test data, by modality and class.

Based on these findings, modality-specific confidence thresholds were defined to filter pseudo-labels: 0.9 for palm and fingernail, and 0.7 for conjunctiva. These thresholds were selected to maximize precision and minimize noisy pseudo-label propagation.

4.8.2 Pseudo-Label Generation and Merging

Using pretrained CNNs, unlabeled external images were passed through the model, and predictions above the defined confidence thresholds were retained. The number of accepted pseudo-labels per modality was:

Modality	**Total Unlabeled Images**	**Pseudo-Labeled (Retained)**	**Confidence Threshold**
Palm	499	246	≥ 0.9
Fingernail	220	167	≥ 0.9
Conjunctiva	149	30	≥ 0.7

Table 4.9: Summary of pseudo-labeled samples retained after confidence filtering.

As summarized in Table 4.9, the number of retained pseudo-labeled samples varied by modality, depending on the selected confidence thresholds.

The confident pseudo-labeled samples were merged with 50% of the original labeled data to create a hybrid training set. This enriched dataset was used for retraining each modality-specific CNN from scratch.

4.8.3 Retraining CNNs on Merged Dataset

After retraining on the merged dataset, all models showed improved validation AUC, particularly for conjunctiva, which previously had the weakest performance. Ta-

Modality	**Best Val AUC**	**Val Accuracy**	**Final Loss**	**Epoch**
Palm	0.9738	91.65%	0.2040	16
Conjunctiva	0.8953	83.10%	0.1527	20
Fingernail	0.9619	92.02%	0.2450	15

Table 4.10: CNN validation performance after semi-supervised retraining.

ble 4.10 summarizes these results.

4.9 Real-World Testing on Unseen Dataset

To evaluate the model's ability to generalize beyond the training distribution, both the Main Model (trained only on the Ghana dataset) and the Semi-Supervised Model (retrained using pseudo-labeled real-world images) were tested on a manually collected dataset of 30 real-world samples. This dataset contains 16 non-anemic and 14 anemic cases, simulating a realistic distribution for practical screening applications.

4.9.1 Evaluation Metrics and Confusion Matrices

Table 4.11 presents key evaluation metrics for both models. While both models achieved relatively strong performance, the Semi-Supervised Model demonstrated improved recall and F1-score, indicating enhanced sensitivity to anemic cases—an essential trait for minimizing false negatives in clinical settings. The Main Model showed higher precision but was more prone to missing anemic predictions.

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**
Main Model (Ghana)	80.0%	90.0%	64.3%	75.0%	0.84
Semi-Supervised Model	83.3%	84.6%	78.6%	81.5%	0.89

Table 4.11: Performance metrics on manually collected real-world dataset (30 samples).

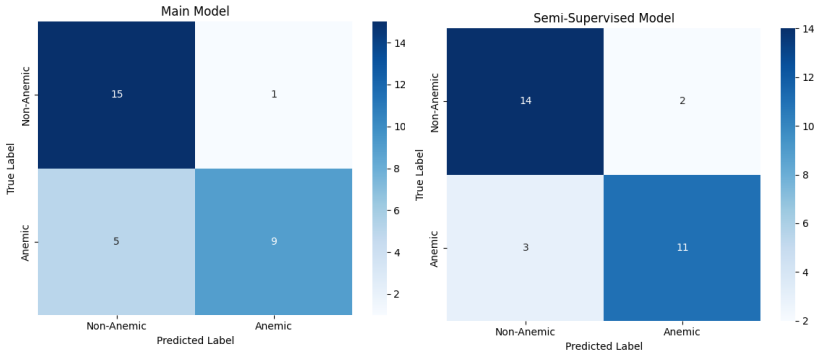


Figure 4-10: Confusion matrices: Left – Main Model; Right – Semi-Supervised Model.

Figure 4-10 shows the confusion matrices. The Semi-Supervised Model reduced false negatives compared to the baseline model, while maintaining balanced precision and recall. This supports the benefit of incorporating domain-shifted pseudo-labeled samples during retraining.

4.9.2 Qualitative Example

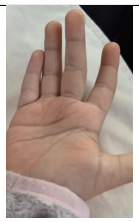
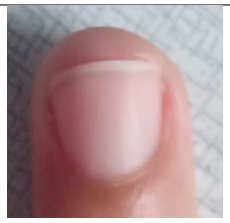
Modality	Raw Image	ROI Image	Prediction
Palm			Anemic
Conjunctiva			Anemic
Fingernail			Anemic
Final Fusion	–	–	**Anemic**

Table 4.12: Prediction on a real-world image: All modalities agreed with the ground truth.

A representative real-world test image was selected per modality. These were segmented, preprocessed, and passed through the models. The final fusion output matched the ground truth label, as shown in Table 4.12.

4.9.3 Discussion

This real-world evaluation demonstrates the effectiveness of the semi-supervised model in reducing false negatives and improving overall performance under domain shift. Compared to the baseline model trained only on the Ghana dataset, the semi-supervised

model achieved higher accuracy (83.3% vs. 80.0%), recall (78.6% vs. 64.3%), and F1-score (81.5% vs. 75.0%). These improvements are critical in clinical screening, where identifying anemic cases accurately is a top priority.

The observed performance gains can be attributed to the inclusion of pseudo-labeled real-world samples, which expanded the diversity of training data and helped the model adapt to new conditions. This supports the conclusion that confident pseudo-labeling combined with retraining enhances model robustness, especially in low-resource environments where manually labeled data is scarce.

Overall, the proposed semi-supervised framework demonstrates practical readiness for deployment in real-world healthcare applications, showing resilience to domain shifts and improved generalization beyond curated datasets.

Chapter 5

Conclusion

This thesis presented a non-invasive, explainable, and multimodal deep learning framework for early anemia detection using images of the palm, conjunctiva, and fingernail. The proposed system combines CNN-based feature extraction with modality-specific attention mechanisms and a weighted late fusion strategy using XGBoost. To enhance model transparency, explainable AI techniques such as SHAP and Grad-CAM were integrated, and agnostic input handling was introduced to support inference with any subset of the three modalities.

Experimental results confirmed the effectiveness of the approach. Among single-modality CNN classifiers, the palm model achieved the highest validation performance, with a test accuracy of 97.0% and AUC of 0.997. The proposed weighted fusion method outperformed both early and attention-based fusion strategies, reaching 94.22% accuracy and 0.9918 AUC on the balanced Ghana test set. SHAP and Grad-CAM analyses validated the model's focus on clinically relevant regions, especially for palm and fingernail inputs. Additionally, agnostic evaluation demonstrated that robust predictions were possible even when one or more modalities were missing, achieving over 93% accuracy in two-modality configurations.

A major contribution of this study is the implementation of a semi-supervised self-training strategy that leverages confident pseudo-labeled images from external datasets. This approach enabled retraining of CNNs on a merged dataset, improving the model's robustness under domain shift. On a manually collected real-world test

set of 30 samples, the semi-supervised model achieved 83.3% accuracy and 81.5% F1-score, outperforming the baseline model (80.0% accuracy, 75.0% F1-score) by improving sensitivity to anemic cases while maintaining strong precision.

The three key contributions of this research are: (1) a modality-agnostic, late-fusion pipeline for anemia prediction across palm, conjunctiva, and fingernail images; (2) the integration of SHAP and Grad-CAM to enable interpretable and clinically meaningful predictions; (3) a confidence-filtered semi-supervised learning framework for enhancing model generalization on unlabeled real-world data.

One limitation of this work is the synthetic nature of the multimodal fusion—since palm, conjunctiva, and fingernail images were not collected from the same individuals, fusion was performed by combining the i-th image from each modality. Although this preserves statistical balance for evaluation, future work should explore person-aligned datasets to better reflect deployment conditions.

Ultimately, this research demonstrates a scalable and explainable AI-based solution for non-invasive anemia screening, suitable for deployment in low-resource settings and adaptable for mobile health platforms. Future directions include expanding to larger, demographically diverse datasets, investigating active learning strategies for efficient annotation, and validating performance in clinical field trials.

Bibliography

[1] Eye conjunctiva segmentation dataset. https://data.mendeley.com/datasets/yxwjgcndg2/1.

[2] Hand and palm images dataset. https://www.kaggle.com/datasets/shyambhu/hands-and-palm-images-dataset.

[3] Nail disease image classification dataset. https://www.kaggle.com/datasets/josephrasanjana/nail-disease-image-classification-dataset.

[4] Peter Appiahene, Justice Williams Asare, and Emmanuel Donkoh. Application of machine learning in detecting iron deficiency anemia using conjunctiva image dataset from ghana. Mendeley Data, V1, 2022.

[5] Peter Appiahene, Justice Williams Asare, and Emmanuel T. Donkoh. Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. *BioData Mining*, 16:2, 2023.

[6] Justice Asare, Peter Appiahene, Emmanuel Donkoh, and Giovanni Dimauro. Iron deficiency anemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images. Preprint, 2023.

[7] Justice Williams Asare, Peter Appiahene, and Emmanuel Donkoh. Anemia detection using palpable palm image datasets from ghana. Mendeley Data, V1, 2022.

[8] Justice Williams Asare, Peter Appiahene, and Emmanuel Donkoh. Detection of anemia using colour of the fingernails image datasets from ghana. Mendeley Data, V1, 2022.

[9] Paola Barra, Attilio Della Greca, Ilaria Amaro, Augusto Tortora, and Mariacarla Staffa. A comparative analysis of xai techniques for medical imaging: Challenges and opportunities. In *Proceedings of the 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 6782–6788, 2024.

[10] Paola Cascante-Bonilla, Fei Tan, Yezhou Qi, and Vicente Ordonez. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. *arXiv preprint arXiv:2001.06001*, 2021. Accessed: 2024-05-06.

[11] C.M. Chaparro and P.S. Suchdev. Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries. *Annals of the New York Academy of Sciences*, 1450(1):15–31, 2019.

[12] S. Das, F. Ahamed, A. Das, et al. Niada (non-invasive anemia detection app), a smartphone-based application with artificial intelligence to measure blood hemoglobin in real-time: A clinical validation. *Cureus*, 16(7):e65442, 2024.

[13] Giovanni Dimauro, Maria Elena Griseta, Mauro Giuseppe Camporeale, et al. An intelligent non-invasive system for automated diagnosis of anemia exploiting a novel dataset. *Artificial Intelligence in Medicine*, 136:102477, 2023.

[14] Mohammad Ennab and Hamid Mcheick. Advancing ai interpretability in medical imaging: A comparative analysis of pixel-level interpretability and grad-cam models. *Machine Learning and Knowledge Extraction*, 7(1):12, 2025.

[15] Yan Hu and Ahmad Chaddad. Shap-integrated convolutional diagnostic networks for feature-selective medical analysis, 2025.

[16] M. I. Iqbal and D. Avianto. Squeeze-and-excitation networks and attention mechanism in automatic detection of coffee leaf diseases based on images. *Journal of Soft Computing Exploration*, 5(4):320–331, 2024.

[17] Tuba Karagül Yıldız, Nilüfer Yurtay, and Birgül Öneç. Classifying anemia types using artificial learning methods. *Engineering Science and Technology, an International Journal*, 24(1):50–70, 2021.

[18] E.T. Lin, S.C. Lu, A.S. Liu, et al. Deep learning-based model for non-invasive hemoglobin estimation via body parts images: A retrospective analysis and a prospective emergency department study. *Journal of Digital Imaging*, 38:775–792, 2025.

[19] S.A.W. Muljono, H.A. Wulandari, M. Azies, et al. Breaking boundaries in diagnosis: Non-invasive anemia detection empowered by ai. *IEEE Access*, 12:9292–9307, 2024.

[20] Avital Oliver, Augustus Odena, Colin A Raffel, Ekin D Cubuk, and Ian Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. In *NeurIPS*, 2018.

[21] B. Pallavi, R. Shukla, R. Kumar, et al. A deep learning-based system for detecting anemia from eye conjunctiva images taken from a smartphone. *IETE Technical Review*, 41(3):274–286, 2023.

[22] M. Ramzan, M.U. Saeed, and G. Ali. Enhancing anemia detection through multimodal data fusion: a non-invasive approach using ehrs and conjunctiva images. *Discover Artificial Intelligence*, 4:100, 2024.

[23] M. Ramzan, J. Sheng, M.U. Saeed, et al. Revolutionizing anemia detection: integrative machine learning models and advanced attention mechanisms. *Visual Computing for Industry, Biomedicine, and Art*, 7:18, 2024.

[24] Punita Rani. Ct-scan segmentation with percentile normalization. https://www.punitarani.com/research/ct-scan-segmentation, 2024. Accessed: 2025-05-06.

[25] J. Singh, K. Saxena, V. Yadav, and L. Sisodia. Machine learning methods in non-invasive detection of iron deficiency. In *11th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)*, pages 1–5, 2024.

[26] R.J. Stoltzfus, A. Edward-Raj, M.L. Dreyfuss, et al. Clinical pallor is useful to detect severe anemia in populations where anemia is prevalent and severe. *Journal of Nutrition*, 129(9):1675–1681, 1999.

[27] Jessie James P. Suarez, Elaine Bhel L. Lagman, Kirstine Kate M. Malabanan, Matthew B. Mendoza, and Raeniel T. Saavedra. Enhancing blood vessel images using contrast limited adaptive histogram equalization and adaptive gaussian thresholding for anemia detection in palpebral eye conjunctiva. In *Proceedings of the ICIAI 2024*, 2024.

[28] C. Viveha, Vani Rajasekar, S. Sowmiya, et al. Point of care noninvasive screening tool for early detection of anemia using smartphone. In *IEEE International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pages 1–5, 2024.

[29] R. Vohra, A. Hussain, A.K. Dudyala, J. Pahareeya, and W. Khan. Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting. *PLoS ONE*, 17(7):e0269685, 2022.

[30] A. Zhang, J. Lou, Z. Pan, et al. Prediction of anemia using facial images and deep learning technology in the emergency department. *Frontiers in Public Health*, 10:964385, 2022.

[31] X. Zhang, C. Liu, D. Yang, T. Song, Y. Ye, K. Li, and Y. Song. Rfaconv: Innovating spatial attention and standard convolutional operation. *arXiv preprint arXiv:2304.03198*, 2023.