

# NutriVision: A Deep Learning-Based Framework for Nutritional Deficiency Detection and Dietary Recommendation using Eye and Nail Images

Aysha Nuzha Nazeer, Angelin Maria Jose, Najiya Fathima K T, Nithya K Unni

Department of Computer Science and Engineering, Vimal Jyothi Engineering College, APJ Abdul Kalam Technological University, India

MS. Sona P

Assistant Professor, Department of Computer Science and Engineering, Vimal Jyothi Engineering College, India

**ABSTRACT** Nutritional deficiencies remain a significant global health concern, particularly in resource constrained environments where access to laboratory-based diagnostics is limited. Conventional clinical assessment methods are often invasive, time-consuming, and unsuitable for large-scale early screening. This study proposes NutriVision, a deep learning-based framework for non-invasive nutritional deficiency screening using multimodal visual inputs derived from eye and nail images. A curated dataset of eye and nail images exhibiting real-world variability, including illumination changes, occlusions, and quality inconsistencies, is utilized to train and evaluate the proposed system. The framework employs transfer learning with the InceptionV3 architecture to extract discriminative visual features. Experimental results indicate that the nail-based model achieves validation accuracy in the range of 95–98%, while the eye-based model achieves slightly lower performance due to the subtle nature of color-based features. Analysis of training dynamics and confusion matrices suggests that nail images provide more stable and discriminative representations compared to eye images, which are more sensitive to illumination and inter-class similarity. A web-based implementation enables real-time prediction and integrates a rule-based recommendation module for dietary guidance. The proposed system is intended as a preliminary screening and decision-support tool rather than a clinical diagnostic system. While the results demonstrate the feasibility of multimodal visual analysis for nutritional assessment, further validation using clinically annotated datasets is required for real-world deployment.

**INDEX TERMS** Nutrient deficiency detection, Computer vision, Convolutional neural networks, Non-invasive diagnosis, Dietary recommendation systems.

## I. INTRODUCTION

**N**UTRITIONAL deficiencies remain a significant global health concern, particularly in low-resource settings where access to routine clinical diagnostics is limited. Deficiencies in essential micronutrients, such as iron and vitamins, are associated with a wide range of adverse health outcomes, including anemia, impaired immunity, and cognitive dysfunction. A key limitation in current healthcare practice is the delayed identification of such deficiencies, as early-stage manifestations are often subtle and typically require laboratory-based confirmation.

Conventional diagnostic approaches rely primarily on biochemical analysis of blood samples, which, although clinically reliable, are invasive, time-consuming, and often im-

practical for large-scale or frequent screening. These limitations highlight the need for alternative approaches that are non-invasive, cost-effective, and accessible for early-stage assessment.

Recent advances in deep learning and computer vision have enabled automated extraction of meaningful patterns from visual data across various healthcare applications. In particular, peripheral visual indicators observable in regions such as the conjunctiva of the eye and the nail bed have been associated with underlying nutritional conditions. While prior studies have explored image-based detection of specific conditions such as anemia using single-modality inputs, these approaches are often constrained by limited dataset diversity, controlled acquisition settings, or lack of robustness in real-

world scenarios.

Motivated by the limitations of conventional diagnostic approaches, this work introduces *NutriVision*, a deep learning-based framework for non-invasive nutritional deficiency screening using multimodal visual inputs derived from eye and nail images. The proposed framework leverages transfer learning with the InceptionV3 architecture, a deep convolutional neural network specifically designed to capture multi-scale spatial features through parallel convolutional operations. This property is particularly beneficial for the current problem, as nutritional deficiencies manifest as subtle variations in color, texture, and structural patterns that require both local and global feature representation for effective classification.

A custom-curated dataset comprising over 500 images is utilized, incorporating real-world variations such as occlusions (e.g., eyeglasses, partially closed eyes, and cosmetic alterations such as nail polish), illumination inconsistencies, and variations in image quality. These factors introduce significant variability in input data, making robust preprocessing essential for reliable model performance.

To address this, a structured preprocessing and data augmentation pipeline is employed. Image resizing ensures uniform input dimensions compatible with the network architecture, while normalization standardizes pixel intensity distributions, improving training stability. Geometric transformations such as rotation, flipping, and scaling are applied to artificially increase dataset diversity, enabling the model to learn invariant features and improving generalization under real-world conditions.

Furthermore, multiple deep learning architectures, including a Sequential CNN and MobileNet, are systematically evaluated under identical experimental conditions to ensure a fair and unbiased comparison. The Sequential CNN serves as a baseline model to assess fundamental learning capability, while MobileNet, a lightweight architecture optimized for computational efficiency, evaluates performance under resource-constrained settings. Among these, InceptionV3 demonstrates superior performance due to its ability to extract hierarchical and multi-scale features, making it more suitable for capturing the complex visual patterns associated with nutritional deficiencies.

Experimental evaluation indicates that the proposed framework achieves validation accuracy in the range of 95–98% for nail image classification, while the eye image model achieves a slightly lower validation accuracy of approximately 96–97%, but comparable performance, with minor overfitting observed in later training epochs, as indicated by a gradual divergence between training and validation accuracy curves, while overall generalization performance remains stable. These results highlight the effectiveness of deep feature extraction in multimodal visual analysis while also reflecting the inherent variability present in eye-based features.

It is important to note that the class labels used in this study represent visually distinguishable categories associated

with nutritional conditions and do not constitute clinically confirmed diagnoses. Accordingly, the proposed system is designed as a preliminary screening and decision-support tool rather than a replacement for professional medical evaluation.

The proposed framework is integrated into a scalable web-based system to enable real-time inference and user accessibility. By leveraging multimodal visual cues and deep learning techniques, this work demonstrates the feasibility of developing low-cost, non-invasive tools for early nutritional health awareness, particularly in resource-constrained environments.

## II. RELATED WORK

Recent advancements in artificial intelligence have significantly enhanced the development of non-invasive diagnostic systems for healthcare, particularly in anemia detection and nutritional assessment. Existing research can be broadly categorized into three directions: (i) classical machine learning-based diagnosis, (ii) deep learning-based image analysis, and (iii) AI-driven dietary assessment and recommendation systems.

Early approaches to anemia detection primarily relied on classical machine learning algorithms such as Support Vector Machines (SVM), decision trees, and k-Nearest Neighbors (k-NN). These methods require handcrafted feature extraction, where domain-specific features such as color intensity, texture, and statistical descriptors are manually derived from medical images before classification. While such approaches are computationally efficient, they are inherently limited in their ability to capture complex and hierarchical patterns in visual data, leading to reduced generalization in real-world scenarios [2]. Furthermore, their performance is highly sensitive to feature engineering quality, making them less robust when dealing with variations in lighting conditions and image acquisition settings.

To address these limitations, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have been widely adopted for medical image analysis. CNNs are designed to automatically learn hierarchical feature representations directly from raw pixel data through convolutional operations. These operations enable the extraction of low-level features such as edges and color gradients in early layers, and high-level semantic features in deeper layers. In the context of anemia detection, CNNs allow the model to capture subtle visual cues such as conjunctival pallor and nail discoloration without requiring manual feature engineering. In this project, CNN-based architectures are employed to enable automated feature extraction, improving both robustness and classification performance.

Several studies have demonstrated the effectiveness of CNN-based approaches in non-invasive anemia detection. Mohammed *et al.* [1] proposed a multi-branch CNN architecture integrated with optimization and explainable AI techniques, enabling improved classification performance and interpretability. Similarly, Muljono *et al.* [4] combined

MobileNetV2 with SVM, where MobileNetV2 serves as a feature extractor and SVM performs classification. This hybrid approach leverages the feature extraction capability of deep learning and the decision boundaries of classical classifiers, achieving high accuracy. However, such hybrid methods introduce additional complexity and may not fully exploit end-to-end learning capabilities.

Different CNN architectures exhibit distinct trade-offs between computational efficiency and feature representation capability. MobileNetV2 is a lightweight architecture designed for efficiency, employing depthwise separable convolutions. This technique decomposes standard convolution into two separate operations: depthwise convolution (spatial filtering) and pointwise convolution (channel-wise combination), significantly reducing computational cost and model parameters. As a result, MobileNetV2 is well-suited for resource-constrained environments and real-time applications. In this project, MobileNetV2 is used as a baseline model to evaluate performance efficiency and provide a comparative reference [4].

However, lightweight models often struggle to capture complex multi-scale patterns present in medical images. To overcome this limitation, deeper architectures such as InceptionV3 are utilized. InceptionV3 introduces multi-scale feature extraction through parallel convolutional filters of varying sizes within a single module. This allows the model to simultaneously capture fine-grained local features and broader contextual information, which is essential for detecting subtle variations in eye and nail images. In this project, InceptionV3 is adopted due to its superior ability to model complex visual patterns and improve classification stability, particularly under varying image conditions [3].

In addition to model architecture, data preprocessing and augmentation are critical components in deep learning pipelines. Data preprocessing refers to standardizing input data through operations such as image resizing and normalization. Resizing ensures that all images conform to the input size required by the model architecture, while normalization scales pixel values to a consistent range, improving numerical stability and accelerating convergence during training. In this project, preprocessing is essential to handle variability in image resolution, lighting, and acquisition conditions.

Data augmentation is employed to artificially expand the training dataset by applying transformations such as rotation, flipping, and geometric distortions. This technique introduces variability into the dataset, enabling the model to learn invariant features and reducing overfitting, where the model memorizes training data instead of generalizing to unseen samples. Given the limited size of medical image datasets, augmentation plays a crucial role in improving model robustness and generalization performance in this project [4].

Beyond diagnostic systems, significant research has focused on dietary assessment and nutritional recommendation. Image-based dietary assessment systems utilize deep learning models to recognize food items and estimate nutritional intake [5]. More advanced approaches incorporate multi-

modal learning, combining text, images, and large language models to enhance dietary analysis [6]. Evolutionary and knowledge-based recommendation systems further enable personalized nutrition planning by integrating user preferences and expert knowledge [7]. Additionally, systems such as NURECON leverage biological and microbial data for advanced nutrition recommendation [8], while deep learning-based frameworks combine CNNs, recurrent networks, and natural language processing for comprehensive nutritional analysis [9].

Recent work has also explored automated dietary assessment using speech processing and natural language processing (NLP), where speech-to-text systems convert verbal dietary descriptions into structured nutritional data [10]. Furthermore, knowledge graph-based approaches model relationships between nutrients and diseases, enabling intelligent, query-based dietary recommendations [11].

Despite these advancements, existing research largely treats anemia detection and dietary recommendation as independent problems. Diagnostic systems focus primarily on classification accuracy without providing actionable health interventions, while dietary systems rely on structured or self-reported inputs rather than physiological indicators. Moreover, most studies adopt single-modality approaches, limiting their applicability in real-world healthcare scenarios.

In contrast, the proposed framework integrates multi-modal image-based nutritional deficiency detection (eye and nail analysis) with a personalized dietary recommendation system within a unified architecture. By combining advanced deep learning techniques such as InceptionV3 for robust feature extraction with user-centric nutritional guidance, the proposed approach addresses the limitations of existing methods and advances toward a more comprehensive, practical, and preventive healthcare solution.

### III. PROPOSED SYSTEM

#### A. SYSTEM OVERVIEW

The proposed *NutriVision* framework is an end-to-end artificial intelligence system for non-invasive nutritional deficiency detection using visual biomarkers extracted from eye and nail images. The system further integrates a dietary recommendation and conversational interface to provide actionable health insights.

The architecture of the system is illustrated in Fig. 1. The pipeline consists of three major stages: (i) data acquisition and preprocessing, (ii) deep learning-based deficiency detection, and (iii) personalized recommendation and interaction modules.

Given an input image  $I$ , the model learns a mapping:

$$f_{\theta} : I \rightarrow y \quad (1)$$

where  $y$  represents the predicted nutritional deficiency class and  $\theta$  denotes the learned model parameters.

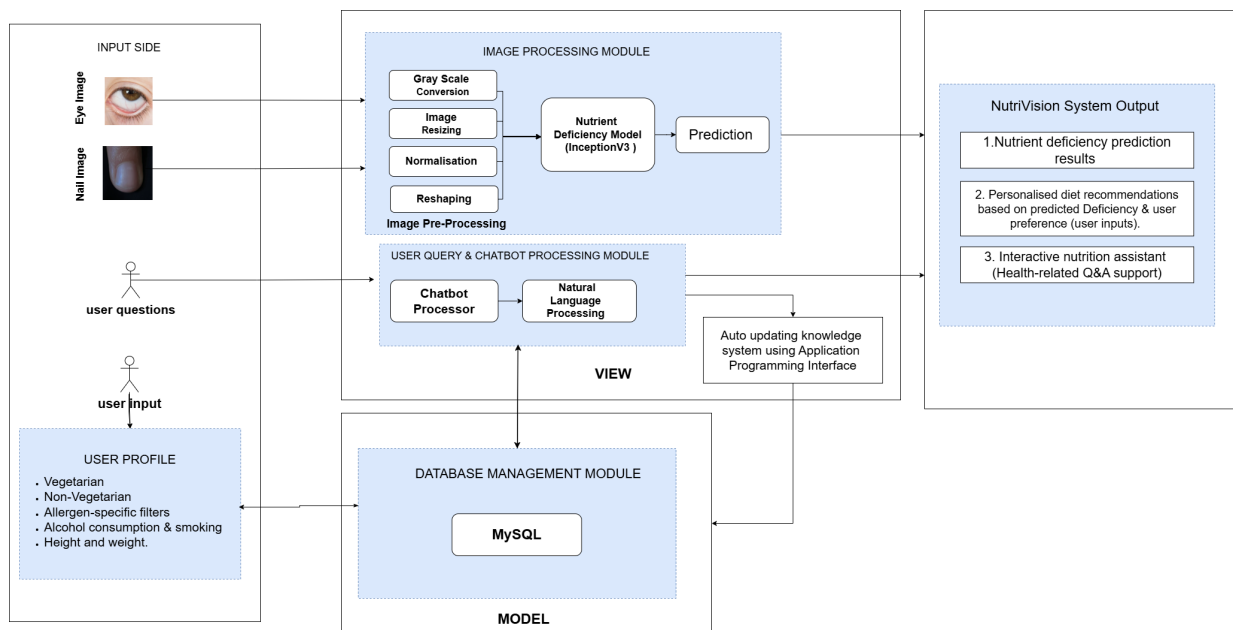


FIGURE 1. NutriVision Architecture Diagram

### B. DATA ACQUISITION AND PREPROCESSING

The proposed system utilizes eye and nail image datasets described in Section X. These datasets are acquired from publicly available sources and contain significant variability in imaging conditions, requiring robust preprocessing to ensure consistent model performance.

#### 1) Preprocessing

To standardize input data, all images are resized to a fixed resolution (e.g.,  $224 \times 224$ ) to match the input requirements of the InceptionV3 architecture. Pixel normalization is applied as:

$$I_{norm} = \frac{I}{255} \quad (2)$$

where  $I$  represents the input image and  $I_{norm}$  denotes the normalized output. This transformation scales pixel values to the range  $[0, 1]$ , improving numerical stability and accelerating convergence during training.

RGB color information is preserved, as color variations are critical for identifying nutritional deficiencies in both eye and nail images.

#### 2) Data Augmentation

To improve generalization and address dataset variability, data augmentation is applied during training. Augmentation operations include rotation, horizontal flipping, scaling, and brightness adjustments:

$$I' = T(I) \quad (3)$$

where  $T(\cdot)$  represents a transformation function. The augmented image  $I'$  retains the semantic content of the original image while introducing variability.

This process enables the model to learn invariant representations under different imaging conditions, thereby reducing overfitting and improving robustness.

### C. DEEP LEARNING-BASED DEFICIENCY DETECTION

The deficiency detection module employs convolutional neural networks (CNNs) for automated feature extraction and classification of visual biomarkers from eye and nail images.

A convolution operation is defined as:

$$Y_{i,j} = \sum_m \sum_n I_{i+m,j+n} \cdot K_{m,n} \quad (4)$$

where  $I$  denotes the input image and  $K$  represents the convolution kernel. The output  $Y_{i,j}$  corresponds to the feature response at spatial location  $(i, j)$ . The indices  $m$  and  $n$  iterate over the spatial dimensions of the kernel, enabling localized feature extraction through a weighted aggregation of neighboring pixel values.

In the proposed framework, convolutional layers are responsible for capturing spatial patterns relevant to nutritional deficiency detection. In particular, low-level features such as color intensity variations are critical for identifying conjunctival pallor in eye images, while texture and structural patterns are essential for detecting abnormalities in nail images.

To effectively model these variations, the InceptionV3 architecture employs multi-scale convolutional processing, where filters of different sizes (e.g.,  $1 \times 1$  and  $3 \times 3$ ) operate in parallel within the same layer. This design enables

simultaneous extraction of fine-grained details and broader contextual features, improving the model's ability to detect subtle yet discriminative visual cues under varying imaging conditions.

As the network depth increases, the learned representations transition from simple patterns to higher-level semantic features, allowing the model to differentiate between visually similar deficiency categories. This hierarchical feature learning is particularly important in this study, where distinctions between classes are often characterized by minor variations in color and texture.

### 1) Eye-Based Analysis

Eye images are analyzed with a focus on the conjunctival region, where color variations serve as key indicators of nutritional deficiencies. In particular, conjunctival pallor is associated with reduced hemoglobin levels and deficiencies such as iron and vitamin B12.

Following preprocessing, the input image is processed by the convolutional neural network, which extracts features related to color intensity, spatial distribution, and localized gradients. Early convolutional layers capture low-level color differences, while deeper layers learn higher-level representations corresponding to variations in redness and brightness within the conjunctival region.

These hierarchical features enable the model to distinguish subtle differences between normal and deficient cases, even under varying illumination and imaging conditions. The preservation of RGB color information is critical in this modality, as grayscale representations would eliminate essential diagnostic cues.

### 2) Nail-Based Analysis

Nail images are analyzed to identify both color-based and structural abnormalities associated with nutritional deficiencies. Visual indicators such as discoloration, ridging, brittleness, and surface irregularities are commonly linked to deficiencies including iron, protein, and zinc.

The CNN extracts discriminative features from nail images by combining color and texture analysis. Convolutional layers capture spatial patterns corresponding to surface irregularities, while deeper layers encode structural variations and textural consistency across the nail region.

Texture feature extraction is particularly important in this modality, as many nail-related deficiencies manifest through changes in surface structure rather than color alone. By learning both color and texture representations, the model achieves more robust classification performance under diverse imaging conditions.

### 3) Model Architecture (InceptionV3)

To address the challenges associated with limited dataset size and high variability in visual features, transfer learning is employed using the InceptionV3 architecture. Transfer learning enables the reuse of feature representations learned

from large-scale datasets, allowing effective training even with limited domain-specific samples.

InceptionV3 enhances feature extraction through multi-scale convolutional processing, where filters of different sizes operate in parallel within the same layer:

$$F = [f_{1 \times 1}(I), f_{3 \times 3}(I), f_{5 \times 5}(I)] \quad (5)$$

where  $I$  represents the input image and  $f_{k \times k}(\cdot)$  denotes a convolution operation with a kernel of size  $k \times k$ . The output feature map  $F$  is formed by concatenating the responses from multiple convolutional filters, each capturing patterns at different spatial scales.

The  $1 \times 1$  convolution captures channel-wise correlations and performs dimensionality reduction, improving computational efficiency. The  $3 \times 3$  and  $5 \times 5$  convolutions capture spatial features at different receptive fields, enabling the model to detect both fine-grained local details and broader contextual patterns.

This multi-scale feature extraction is particularly important in this study, as nutritional deficiencies manifest through subtle variations in color intensity (e.g., conjunctival pallor in eye images) and texture irregularities (e.g., ridging and discoloration in nail images). By combining features at multiple scales, the model improves its ability to distinguish visually similar classes under varying imaging conditions.

Following feature extraction, the learned representations are passed through fully connected layers, and the final classification is performed using the softmax function:

$$P(y = k | x) = \frac{e^{z_k}}{\sum_i e^{z_i}} \quad (6)$$

where  $z_k$  represents the logit corresponding to class  $k$ , and  $P(y = k | x)$  denotes the predicted probability of the input belonging to class  $k$ . The softmax function converts raw model outputs into a normalized probability distribution across all classes.

This probabilistic formulation enables the model to assign confidence scores to each deficiency category, facilitating reliable classification of nutritional conditions based on visual features extracted from eye and nail images.

## D. PERSONALIZED DIETARY RECOMMENDATION FRAMEWORK

Following deficiency classification, the system generates dietary recommendations tailored to the predicted nutritional condition. This is implemented through a rule-based mapping mechanism:

$$R : y \rightarrow D \quad (7)$$

where  $y$  denotes the predicted deficiency class and  $D$  represents a set of recommended dietary items.

This mapping is constructed using domain knowledge that associates each deficiency with relevant nutrients and corresponding food sources. For example, iron deficiency is

mapped to iron-rich foods, while vitamin-related deficiencies are mapped to appropriate dietary sources.

The mapping function  $R$  operates as a deterministic lookup mechanism, ensuring that recommendations are directly aligned with the predicted condition. This approach is particularly suitable in this context, as it provides interpretable and reliable outputs without requiring large-scale dietary datasets for model training.

#### E. CONVERSATIONAL AI INTEGRATION

To enhance usability and provide interactive guidance, a conversational AI module is integrated into the system. The chatbot processes user queries and generates responses based on model predictions and user-specific information.

The response generation process can be expressed as:

$$r = g(q, y, U) \quad (8)$$

where  $q$  represents the user query,  $y$  is the predicted deficiency class, and  $U$  denotes user-specific attributes such as dietary preferences, allergies, and medical conditions.

The function  $g(\cdot)$  combines contextual information from the prediction and user profile to produce personalized and context-aware responses. In practice, this involves retrieving appropriate dietary recommendations, filtering them based on user constraints, and presenting them in natural language.

This integration improves user engagement and bridges the gap between automated prediction and practical dietary guidance, enabling the system to function as a decision-support tool rather than a standalone classifier.

#### F. SYSTEM DEPLOYMENT

The proposed framework is deployed as a web-based application using the Django framework. The backend handles image preprocessing, model inference, and integration between system components, while a MySQL database is used for storing user data and prediction results.

The system supports real-time inference, allowing users to upload images and receive predictions and recommendations with minimal latency. The modular architecture facilitates scalability and enables future integration of additional features.

### IV. DATASET DESCRIPTION

The dataset used in this study consists of eye and nail images collected from publicly available sources, including Kaggle datasets. It is designed to support supervised multi-class classification of nutritional deficiencies based on visual biomarkers.

The complete dataset contains a total of 7,325 images, comprising 4,877 eye images and 2,448 nail images. The eye dataset focuses on the conjunctival region, where color variations are indicative of deficiencies such as Vitamin A and Vitamin B complex. The nail dataset captures both color and structural features, including ridging, discoloration, and

surface irregularities associated with deficiencies such as iron, protein, and zinc.

The dataset exhibits substantial variability due to real-world acquisition conditions, including changes in illumination, background clutter, image resolution, and occlusions. Non-diagnostic samples, such as images containing eyewear artifacts or cosmetic interference (e.g., nail polish), are excluded during preprocessing to ensure that the model focuses on clinically relevant features.

The dataset is partitioned into mutually exclusive training and validation subsets:

$$\mathcal{D} = \mathcal{D}_{train} \cup \mathcal{D}_{val} \quad (9)$$

where  $\mathcal{D}_{train}$  is used for model learning and  $\mathcal{D}_{val}$  is used for performance evaluation. This separation ensures unbiased assessment of model generalization.

Despite its size, the dataset presents challenges such as class imbalance and the absence of clinically validated annotations, as labels are inferred from visual characteristics. These limitations are considered in the experimental evaluation.

TABLE 1. Class Distribution of the Filtered Eye Image Dataset

Class	Number of Images
<b>Valid Diagnostic Classes</b>	
Normal	126
Vitamin A Deficiency	117
Vitamin B Complex Deficiency	120
Vitamin B12 Deficiency	119
Vitamin B3 Deficiency	120
<b>Total Valid Samples</b>	<b>602</b>
<b>Excluded Non-Diagnostic Samples</b>	
Closed Eyes	2360
Eyeglasses / Sunglasses	1419
Invalid Samples	496
<b>Total Excluded Samples</b>	<b>4275</b>
<b>Total Eye Dataset</b>	<b>4877</b>

The eye dataset contains a relatively limited number of valid diagnostic samples (602 images) compared to the total dataset size, primarily due to the exclusion of non-diagnostic images such as closed-eye and occluded samples. The valid classes exhibit a near-uniform distribution across deficiency categories; however, the overall sample size remains small. This constraint increases the risk of overfitting and necessitates the use of data augmentation and transfer learning to ensure stable feature learning and generalization.

The nail dataset provides a larger set of valid diagnostic samples (1,970 images), enabling more effective learning of texture and structural features. However, significant class imbalance is observed, particularly for zinc deficiency, which contains substantially fewer samples compared to

TABLE 2. Class Distribution of the Filtered Nail Image Dataset

Class	Number of Images
<b>Valid Diagnostic Classes</b>	
Healthy	248
Iron Deficiency	250
Protein Deficiency	264
Vitamin B12 Deficiency	300
Vitamin B7 Deficiency	276
Vitamin C Deficiency	300
Vitamin D Deficiency	282
Zinc Deficiency	50
<b>Total Valid Samples</b>	<b>1970</b>
<b>Excluded Non-Diagnostic Samples</b>	
Nail Polish	352
Invalid Samples	126
<b>Total Excluded Samples</b>	<b>478</b>
<b>Total Nail Dataset</b>	<b>2448</b>

other classes. This imbalance introduces potential bias toward majority classes and is explicitly addressed during training through data augmentation and validation strategies to improve performance on underrepresented categories.

## V. EXPERIMENTAL SETUP

### A. EXPERIMENTAL DESIGN

The experimental setup is designed to evaluate the performance of the proposed NutriVision framework under controlled and reproducible conditions. The primary objective is to assess the model's capability to learn discriminative features from eye and nail images while maintaining robustness to variations in illumination, orientation, and image quality.

All experiments are conducted using a consistent training protocol to ensure comparability between modalities.

### B. DATA PARTITIONING STRATEGY

The dataset is partitioned into training and validation subsets using an 80:20 split, implemented through the ImageDataGenerator framework. The partitioning is defined as:

$$\mathcal{D} = \mathcal{D}_{train} \cup \mathcal{D}_{val}, \quad \mathcal{D}_{train} \cap \mathcal{D}_{val} = \emptyset \quad (10)$$

where  $\mathcal{D}_{train}$  denotes the training set and  $\mathcal{D}_{val}$  represents the validation set.

The validation subset is used to monitor model performance during training and to guide hyperparameter tuning, including learning rate adjustments and early stopping.

Due to dataset constraints and the use of directory-based data loading, a dedicated hold-out test set is not explicitly defined in this study. Instead, validation performance is used as a proxy for generalization behavior. While this approach provides useful insights into model learning, it may not fully reflect performance on completely unseen data. To mitigate

potential bias, early stopping and validation monitoring are employed to prevent overfitting during training.

Future work will incorporate a separate test set and cross-validation strategies to provide a more comprehensive evaluation.

### C. INPUT CONFIGURATION

All input images are resized to a fixed spatial resolution of  $224 \times 224$  pixels to match the input requirements of the InceptionV3 architecture. This ensures consistency in tensor dimensions and enables efficient batch processing.

Pixel normalization is applied as:

$$I_{norm} = \frac{I}{255} \quad (11)$$

where  $I$  represents the original image and  $I_{norm}$  denotes the normalized image.

This operation rescales pixel intensities from the range  $[0, 255]$  to  $[0, 1]$ . By constraining input values to a smaller numerical range, normalization stabilizes gradient updates during backpropagation and accelerates convergence of the optimization process. In the context of nutritional deficiency detection, where subtle color variations (e.g., conjunctival pallor or nail discoloration) are critical, normalization ensures that intensity differences are preserved while avoiding scale dominance in feature learning.

### D. DATA AUGMENTATION STRATEGY

To improve generalization and mitigate overfitting due to limited dataset size, online data augmentation is applied during training. The transformation process is defined as:

$$I' = T(I) \quad (12)$$

where  $T(\cdot)$  represents a stochastic transformation function and  $I'$  denotes the augmented image.

The transformation function  $T(\cdot)$  generates multiple variations of an input image while preserving its semantic content. This enables the model to learn invariant representations under real-world variations.

The applied augmentation operations include:

- **Rotation ( $\pm 30^\circ$ ):** accounts for variations in camera orientation.
- **Width and height shifts (up to 20%):** simulate positional misalignment during image capture.
- **Zoom (up to 20%):** captures scale variations in image acquisition.
- **Horizontal flipping:** increases data diversity for approximately symmetric structures.
- **Brightness variation ( $[0.8, 1.2]$ ):** models illumination changes by scaling pixel intensity.

These transformations are applied online during training using the ImageDataGenerator framework, ensuring that each training epoch observes a dynamically augmented dataset. This strategy reduces overfitting and improves the model's ability to generalize to unseen samples under varying imaging conditions.

### E. MODEL CONFIGURATION

The proposed model is based on the InceptionV3 architecture with pre-trained ImageNet weights, enabling transfer learning from large-scale visual representations. The original classification head is replaced with a task-specific architecture consisting of:

- Global average pooling (GAP)
- Fully connected layer (1024 units, ReLU activation)
- Dropout layer (rate = 0.5)
- Softmax output layer

Global average pooling is used instead of fully connected flattening to reduce the number of trainable parameters and mitigate overfitting. The dense layer enables non-linear feature transformation, while dropout regularization (rate = 0.5) prevents co-adaptation of neurons. The softmax layer converts output logits into class probabilities for multi-class classification.

To adapt the model to domain-specific features, fine-tuning is performed by unfreezing the last 50 layers of the network. This allows higher-level feature representations to adjust to nutritional deficiency patterns while preserving lower-level general features.

### F. TRAINING PROCEDURE

The model is trained using the Adam optimizer with a learning rate of  $1 \times 10^{-4}$ , which provides adaptive parameter updates and stable convergence.

The optimization objective is defined using categorical cross-entropy:

$$\mathcal{L} = - \sum_k y_k \log(\hat{y}_k) \quad (13)$$

The loss function measures the discrepancy between the true class distribution  $y_k$  and the predicted probability  $\hat{y}_k$ . It assigns higher penalties when the model assigns low probability to the correct class, thereby guiding the model toward more accurate predictions.

Training is performed using mini-batch gradient descent with the following configuration:

- Batch size: 32
- Number of epochs: 5

To improve generalization and prevent overfitting, the following strategies are employed:

- Early stopping based on validation loss
- Learning rate reduction on plateau
- Data augmentation applied during training

### G. REGULARIZATION TECHNIQUES

To mitigate overfitting and improve generalization, multiple regularization strategies are employed:

- Dropout (rate = 0.5)
- Early stopping based on validation loss
- Learning rate reduction on plateau

Dropout randomly deactivates a fraction of neurons during training, preventing the model from relying on specific feature activations and encouraging more robust feature learning. Early stopping monitors validation loss and halts training when performance begins to degrade, thereby preventing overfitting. Learning rate reduction dynamically decreases the learning rate when validation performance stagnates, enabling finer convergence in later training stages.

### H. PREDICTION AND DECISION RULE

The predicted class is obtained as:

$$\hat{y} = \arg \max(\mathbf{p}) \quad (14)$$

where  $\mathbf{p}$  represents the predicted probability distribution over classes. The  $\arg \max$  operation selects the class with the highest predicted probability. This corresponds to the most confident prediction of the model for a given input sample.

### I. EVALUATION METRICS

Model performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis. These metrics provide both overall and class-wise performance assessment, particularly in the presence of class imbalance.

### J. IMPLEMENTATION DETAILS

The system is implemented using TensorFlow and Keras. Data preprocessing and augmentation are performed using the ImageDataGenerator framework. Model training is conducted on a GPU-enabled environment.

## VI. EXPERIMENTAL RESULTS AND DISCUSSION

### A. PERFORMANCE EVALUATION METRICS

The performance of the proposed framework is evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis. These metrics provide both overall and class-wise assessment, which is essential in multi-class settings with class imbalance.

Accuracy reflects overall prediction correctness, while precision and recall capture class-specific performance. The F1-score provides a balanced measure of these metrics. The confusion matrix enables detailed analysis of misclassification patterns, particularly among visually similar deficiency categories.

These metrics collectively ensure a robust and interpretable evaluation of the proposed framework under real-world conditions.

### B. EYE-BASED DEFICIENCY DETECTION

The eye-based model achieves a validation accuracy of approximately 95–96%, as shown in Fig. 2.

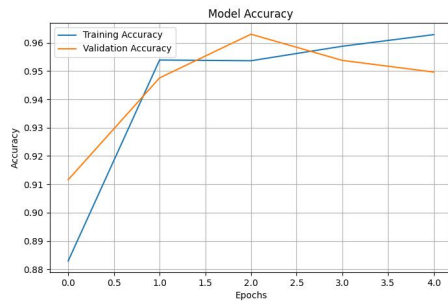


FIGURE 2. Training and validation accuracy for the eye-based model.

The corresponding loss curve (Fig. 3) indicates convergence during early epochs, followed by slight divergence, suggesting mild overfitting.

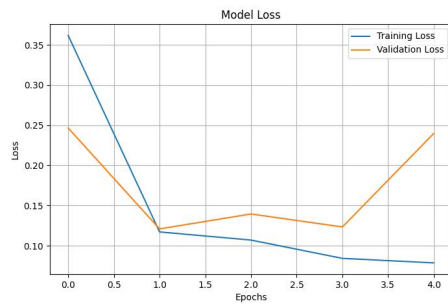


FIGURE 3. Training and validation loss for the eye-based model.

The confusion matrix (Fig. 4) reveals noticeable off-diagonal elements, indicating misclassification among visually similar deficiency categories.

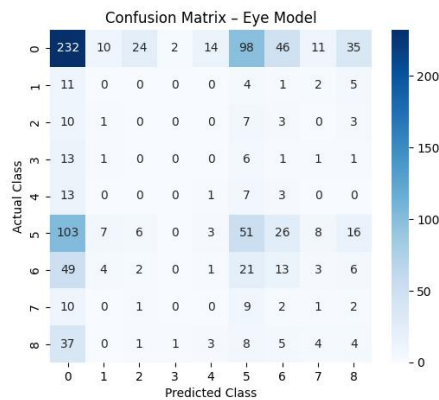


FIGURE 4. Confusion matrix for the eye-based model.

The classification decision is obtained using the softmax function:

$$P(y = k | x) = \frac{e^{z_k}}{\sum_i e^{z_i}} \quad (15)$$

The model outputs a set of logits  $z_k$  for each class. The softmax function converts these logits into probabilities, where each value represents the likelihood of the input image belonging to a specific deficiency class. The predicted class corresponds to the highest probability value.

Eye-based detection relies primarily on subtle color variations, such as conjunctival pallor. These features are extracted by convolutional layers as low-intensity gradients and color distributions. However, due to their low contrast and sensitivity to illumination conditions, the resulting feature representations may overlap across classes.

As a result, the softmax probabilities for visually similar classes become comparable, leading to ambiguous predictions and increased misclassification. This explains the presence of off-diagonal elements in the confusion matrix.

Furthermore, the divergence observed in validation loss indicates that the model begins to specialize on training-specific patterns, reducing its ability to generalize to unseen eye images. This behavior is consistent with the limited distinctiveness of color-based features compared to texture-based representations.

### C. NAIL-BASED DEFICIENCY DETECTION

The nail-based model achieves a validation accuracy in the range of 97–98%, as illustrated in Fig. 5.

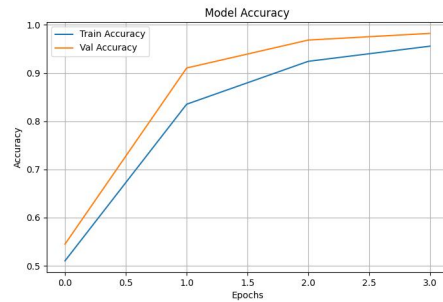


FIGURE 5. Training and validation accuracy for the nail-based model.

The corresponding loss curve (Fig. 6) demonstrates consistent convergence during early epochs, followed by minor fluctuations in validation loss, indicating controlled overfitting behavior.

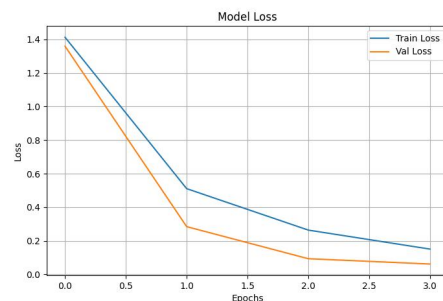


FIGURE 6. Training and validation loss for the nail-based model.

The confusion matrix (Fig. 7) exhibits strong diagonal dominance, indicating high true positive rates across most classes, with minimal inter-class confusion.

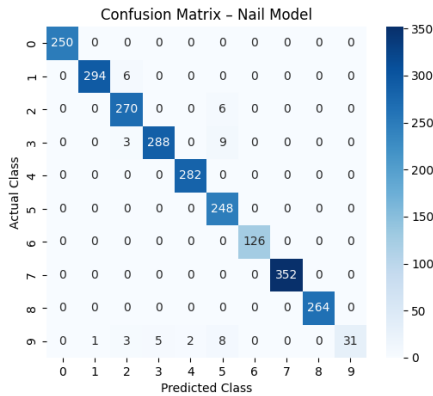


FIGURE 7. Confusion matrix for the nail-based model.

To further quantify classification performance, standard evaluation metrics are defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

Precision evaluates the correctness of positive predictions, while recall measures the model's ability to identify all relevant instances. The F1-score provides a harmonic balance between precision and recall, making it suitable for multi-class classification scenarios.

The superior performance of the nail-based model can be attributed to the availability of both color and structural features. Convolutional neural networks extract hierarchical representations from input images, which can be expressed as:

$$\mathbf{f} = \phi(I) \quad (19)$$

where  $I$  denotes the input nail image and  $\phi(\cdot)$  represents the nonlinear feature extraction function learned by the network. These features capture both chromatic variations and spatial texture patterns, improving class separability in the learned feature space.

The final classification is obtained through:

$$\hat{y} = \arg \max_k P(y = k | \mathbf{f}) \quad (20)$$

where  $P(y = k | \mathbf{f})$  is computed using the softmax function over the extracted features.

Compared to eye-based inputs, nail images provide more stable visual cues with higher structural consistency, leading to improved discriminative performance across classes.

#### D. EFFECT OF DATA AUGMENTATION AND TRANSFER LEARNING

Data augmentation is applied to improve the robustness of the model by generating transformed samples:

$$I' = T(I) \quad (21)$$

where  $T(\cdot)$  represents stochastic transformations such as rotation, scaling, and brightness variation.

This transformation exposes the model to multiple variations of the same input image, reducing sensitivity to changes in illumination, orientation, and scale. As a result, the model learns more invariant feature representations, which contributes to improved generalization performance, particularly in the presence of real-world variability.

Transfer learning is employed using a pre-trained InceptionV3 backbone, allowing the model to leverage feature representations learned from large-scale datasets. This enables effective extraction of low-level and high-level visual features even with a limited domain-specific dataset.

The benefit of transfer learning is more pronounced in the eye-based model, where visual differences between classes are subtle and primarily based on color variations. Pre-trained convolutional layers capture generic edge and color features, which are refined during fine-tuning to distinguish deficiency-specific patterns.

#### E. MODEL COMPARISON

To validate the effectiveness of the proposed architecture, the performance of InceptionV3 is compared with alternative deep learning models, including a Sequential CNN and MobileNet.

The Sequential CNN represents a conventional convolutional architecture with stacked convolutional and pooling layers, while MobileNet is a lightweight model that utilizes depthwise separable convolutions to reduce computational complexity.

The Sequential CNN demonstrates lower classification accuracy due to its limited depth and reduced capacity to capture complex visual patterns. In contrast, MobileNet achieves moderate performance with improved efficiency but lacks the representational richness required for fine-grained medical image classification.

InceptionV3 outperforms both models by leveraging multi-scale feature extraction, where parallel convolutional filters of different sizes capture both local and global features. This capability is particularly important for detecting subtle variations in color and texture present in eye and nail images.

The comparative results indicate that InceptionV3 provides superior performance in terms of accuracy and stability, making it more suitable for the proposed nutritional deficiency detection task.

#### F. AI CHATBOT EVALUATION

The chatbot module is evaluated in terms of its ability to generate contextually relevant and interpretable dietary

recommendations based on model predictions. Unlike the classification component, the chatbot is assessed qualitatively due to the absence of standardized ground truth labels for conversational outputs.

The recommendation process is defined as:

$$R = f(y) \quad (22)$$

where  $y$  represents the predicted nutritional deficiency class and  $R$  denotes the generated set of dietary recommendations.

The mapping function  $f(\cdot)$  associates each predicted deficiency with a predefined set of nutritional guidelines. This ensures that recommendations are directly aligned with the model output and remain consistent across similar predictions.

The chatbot outputs are assessed based on the following qualitative criteria:

- **Relevance:** The extent to which the recommended dietary actions correspond to the predicted deficiency.
- **Consistency:** The stability of recommendations across repeated queries for the same prediction.
- **Interpretability:** The clarity and usability of responses for non-expert users.

The chatbot generates deficiency-specific dietary recommendations that are aligned with the predicted class and user-specific constraints defined in the user profile, such as dietary preferences and known allergies. The responses remain stable across repeated interactions due to the deterministic mapping between prediction outputs and recommendation rules. Furthermore, the generated responses are presented in a simplified and user-friendly format, enhancing interpretability and usability.

The evaluation is qualitative and does not include quantitative performance metrics or user studies. The effectiveness of the chatbot is inherently dependent on the accuracy of the underlying classification model and the completeness of user-provided information. Consequently, the chatbot is treated as a supportive component rather than a primary analytical contribution.

### G. DISCUSSION

The experimental results indicate that the proposed framework achieves reliable performance for non-invasive nutritional screening, with the nail-based model consistently outperforming the eye-based model in terms of accuracy and stability.

This performance difference can be explained by the nature of feature representations learned by the model. Nail images provide both structural and chromatic information, enabling the extraction of more discriminative features. In contrast, eye-based detection relies primarily on subtle color variations, which exhibit higher intra-class similarity and are more sensitive to illumination changes.

Formally, the classification process can be expressed as:

$$\hat{y} = \arg \max_k P(y = k | \mathbf{f}) \quad (23)$$

where  $\mathbf{f} = \phi(I)$  denotes the feature representation extracted from input image  $I$ , and  $P(y = k | \mathbf{f})$  represents the predicted probability for class  $k$ .

**Explanation:** The function  $\phi(\cdot)$  maps the input image to a feature space, where class separability determines classification performance. Nail images produce more distinct feature distributions in this space, leading to higher confidence predictions. In contrast, eye images yield overlapping feature representations, resulting in increased misclassification.

These findings suggest that while eye-based analysis provides complementary information, nail-based features offer more robust and discriminative cues for deficiency detection.

Overall, the proposed system demonstrates potential for preliminary nutritional assessment in non-clinical settings. However, its practical deployment requires further validation on larger and more diverse datasets, along with improvements in robustness under varying imaging conditions.

### H. QUALITATIVE SYSTEM OUTPUT ANALYSIS

In addition to quantitative evaluation, the practical applicability of the proposed framework is examined through real-time system outputs.

Fig. 8 illustrates a representative prediction from the nail-based module. The system processes the input image and generates a classification result along with associated dietary guidance.

Fig. 9 presents an example of eye-based analysis, where the system produces a deficiency prediction and corresponding recommendation.

The system output can be formally represented as:

$$O = (\hat{y}, R) \quad (24)$$

where  $\hat{y}$  denotes the predicted deficiency class and  $R$  represents the corresponding dietary recommendation.

The output combines classification results with actionable guidance, forming an integrated decision-support response. This design enables the system to move beyond prediction and provide user-oriented interpretation.

These qualitative results demonstrate the end-to-end functionality of the framework, including feature extraction, classification, and recommendation generation. The integration of prediction outputs with user-facing guidance highlights the practical applicability of the system while maintaining a clear separation between analytical and interface components.

### VII. LIMITATIONS

Despite the encouraging performance of the proposed framework, several limitations must be acknowledged to ensure an accurate interpretation of the results.

The dataset used in this study exhibits imbalance across different deficiency categories. This may bias the model toward frequently occurring classes and reduce its sensitivity

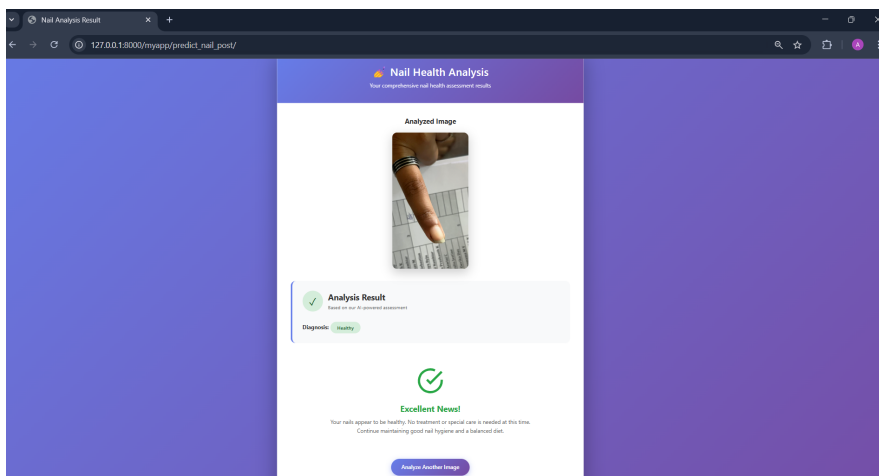


FIGURE 8. Sample output of nail-based deficiency detection and recommendation interface.

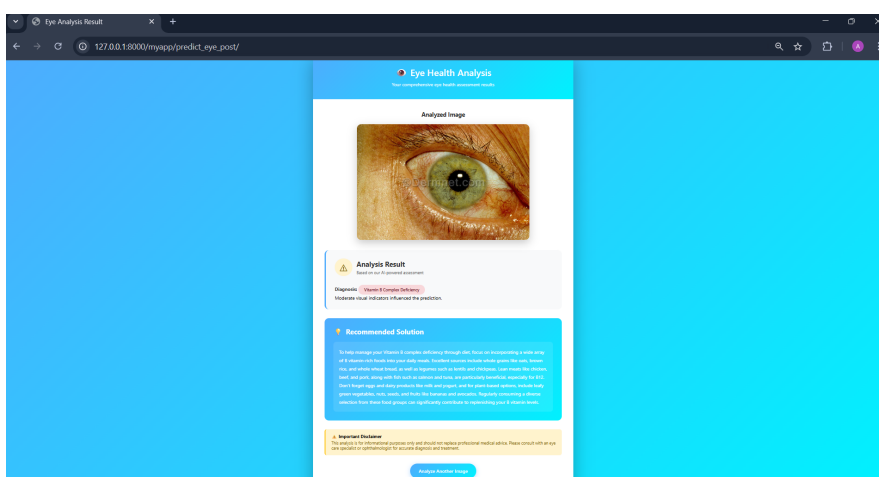


FIGURE 9. Sample output of eye-based deficiency detection with personalized dietary recommendation.

to underrepresented conditions. As a result, classification performance may vary across categories.

The reliability of visual feature extraction is affected by the nature of the input modality. Eye-based analysis relies on subtle color variations, which are inherently sensitive to illumination changes and exhibit higher similarity across classes. This limits the discriminative capability of the model for certain deficiencies.

The system remains sensitive to variations in image acquisition conditions, including lighting, resolution, and occlusions such as eyewear or cosmetic alterations. Although preprocessing and augmentation improve robustness, they cannot fully eliminate real-world variability.

In addition, the framework relies exclusively on visual biomarkers, which may not fully reflect underlying physiological conditions. Consequently, the system is intended for preliminary screening and should not be interpreted as a substitute for clinical diagnosis.

Finally, the dietary recommendation module is based on predefined mappings and does not incorporate adaptive learning. Its effectiveness depends on both the accuracy of the predicted class and the completeness of user-provided information, limiting its ability to handle complex or evolving user needs.

## VIII. CONCLUSION

This work presents a deep learning-based framework for non-invasive nutritional deficiency detection using eye and nail images. The experimental results demonstrate that nail-based analysis provides more stable and discriminative visual features compared to eye-based analysis, leading to improved classification performance.

The proposed system integrates image-based prediction with a recommendation mechanism, enabling the generation of interpretable and user-oriented outputs. By linking classification results to dietary guidance, the framework extends beyond conventional image analysis and supports practical

decision-making.

The findings indicate that the system can serve as a supportive tool for preliminary nutritional assessment in non-clinical settings. However, its effectiveness is influenced by dataset characteristics, input quality, and the limitations of visual biomarkers. Further validation is required before considering real-world or clinical deployment.

## IX. FUTURE WORK

Future work will focus on improving the robustness and generalization capability of the proposed framework. Expanding the dataset with a larger and more diverse population will help address class imbalance and improve performance across different demographic and environmental conditions. Incorporating clinically validated data will further strengthen the reliability of the system.

Enhancements in feature representation are also necessary to better capture subtle variations, particularly in eye-based analysis where visual differences are less pronounced. Improving feature separability can reduce ambiguity and enhance classification accuracy.

The recommendation module can be extended to incorporate adaptive mechanisms that learn from user interaction and feedback, enabling more personalized and dynamic dietary guidance. This would improve long-term usability and user engagement.

Future extensions may also explore the integration of additional non-invasive indicators, such as skin or facial features, to provide a more comprehensive assessment of nutritional status. Finally, large-scale deployment and user studies will be essential to evaluate system performance, usability, and real-world impact.

## DECLARATIONS

### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Author Contributions

Aysha Nuzha Nazeer: Conceptualization, Formal Analysis, Writing – Original Draft.

Angelin Maria Jose: Methodology, Data Curation.

Najiya Fathima K T: Software, Data Curation.

Nithya K Unni: Software.

All authors: Validation, Investigation, Writing – Review & Editing.

Sona P.: Supervision.

All authors have read and approved the final manuscript.

## Data Availability

The datasets used and/or analyzed during the current study are publicly available from Kaggle and other open-access repositories. Further details are available from the corresponding author on reasonable request.

## Research Involving Human and/or Animals

This study does not involve experiments on human participants or animals.

## Informed Consent

Not applicable.

## REFERENCES

- [1] K. K. Mohammed, N. Dahmani, R. Ahmed, A. Darwish, and A. E. Hassaniien, "An explainable AI and optimized multi-branch convolutional neural network model for eye anemia diagnosis," *IEEE Access*, vol. 13, pp. 71840–71858, 2025, doi: 10.1109/ACCESS.2025.3560689.
- [2] M. S. Farooq et al., "Developing a transparent anaemia prediction model empowered with explainable artificial intelligence," *IEEE Access*, vol. 13, pp. 1307–1321, 2025, doi: 10.1109/ACCESS.2024.3522080.
- [3] J. R. Navarro-Cabrera et al., "Machine vision model using nail images for non-invasive detection of iron deficiency anemia in university students," *Frontiers in Big Data*, vol. 8, Art. no. 1557600, 2025, doi: 10.3389/fdata.2025.1557600.
- [4] Muljono et al., "Breaking boundaries in diagnosis: Non-invasive anemia detection empowered by AI," *IEEE Access*, vol. 12, pp. 9292–9310, 2024, doi: 10.1109/ACCESS.2024.3353788.
- [5] F. S. Konstantakopoulos, E. I. Georga, and D. I. Fotiadis, "An automated image-based dietary assessment system for Mediterranean foods," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 4, pp. 45–54, 2023, doi: 10.1109/OJEMB.2023.3266135.
- [6] F. P.-W. Lo et al., "Dietary assessment with multimodal ChatGPT: A systematic analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 12, pp. 7577–7589, 2024, doi: 10.1109/JBHI.2024.3417280.
- [7] B. Ortiz-Viso et al., "Evolutionary approach for building, exploring and recommending complex items with application in nutritional interventions," *IEEE Access*, vol. 11, pp. 65891–65907, 2023, doi: 10.1109/ACCESS.2023.3290918.
- [8] Z.-Q. Hu et al., "NURECON: A novel online system for determining nutrition requirements based on microbial composition," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 21, no. 2, pp. 254–267, 2024, doi: 10.1109/TCBB.2024.3349572.
- [9] P. Rojanaphan, "Automated nutrient deficiency detection and recommendation systems using deep learning in nutrition science," *International Journal of Scientific Research and Management*, vol. 12, no. 11, pp. 1746–1763, 2024, doi: 10.18535/ijerm/v12i11.ec09.
- [10] C. T. Dodd et al., "Automated processing of speech recordings for dietary assessment: Evaluation in the LLMIC context," *IEEE Access*, vol. 13, pp. 59911–59925, 2025, doi: 10.1109/ACCESS.2025.3555998.
- [11] C. Fu et al., "KG4NH: A comprehensive knowledge graph for question answering in dietary nutrition and human health," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 3, pp. 1793–1807, 2025, doi: 10.1109/JBHI.2023.3338356.
- [12] S. Rajpurkar, J. Irvin, K. Zhu, et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," *IEEE Trans. Med. Imaging*, vol. 38, no. 5, pp. 1159–1167, 2019.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE CVPR*, pp. 770–778, 2016.
- [14] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *NeurIPS*, 2012.
- [15] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Proc. ICML*, 2019.
- [16] G. Litjens, T. Kooi, B. Bejnordi, et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

- [17] J. Esteva, A. Kuprel, R. Novoa, et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *MICCAI*, 2015.
- [19] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," *Wiley Interdisciplinary Reviews: Data Mining*, 2019.
- [20] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," *NeurIPS*, 2017.
- [21] H. Tizhoosh and F. Pantanowitz, "Artificial intelligence and digital pathology: challenges and opportunities," *Journal of Pathology Informatics*, 2018.
- [22] Z. Obermeyer and E. Emanuel, "Predicting the future — big data, machine learning, and clinical medicine," *NEJM*, 2016.
- [23] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in Bioinformatics*, 2018.