A Comparative Analysis of Interpretable Deep Learning Models for Nutrient Analysis in Vulnerable Populations

Zvinodashe Revesai and Okuthe P. Kogeda(✉)

School of Mathematics, Statistics and Computer Science, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville Campus, Durban 3209, Republic of South Africa
kogedao@ukzn.ac.za

Abstract. Nutrient deficiencies affect over 2 billion people globally, with vulnerable populations disproportionately impacted. While artificial intelligence offers promising solutions for nutrient analysis, the lack of transparency in deep learning models limits their adoption in healthcare settings where interpretability is essential. This study presents a comparative analysis of interpretable deep learning models for nutrient analysis in food images, focusing on the trade-offs between model performance, interpretability, and computational efficiency. Using the Food101 dataset augmented with nutritional information, we implement and evaluate three distinct approaches: a ResNet50 architecture with Grad-CAM, a Vision Transformer with attention maps, and a DenseNet with LIME integration. Our hybrid approach, combining DenseNet with LIME, achieves 94% accuracy in nutrient detection whilst providing comprehensive interpretability, though at increased computational cost. The ResNet50 model demonstrates efficient performance with 92% accuracy and minimal computational overhead, whilst the Vision Transformer offers unique insights through attention mechanisms despite lower accuracy at 89%. Experimental results show varying effectiveness across nutrient types, with macronutrient detection achieving 95% accuracy and mineral detection proved more challenging at 87%. These findings provide valuable insights for selecting appropriate architecture based on specific application requirements, particularly where the balance between accuracy, interpretability, and computational efficiency is crucial.

Keywords: Interpretable Artificial Intelligence · Nutritional Analysis · Deep Learning Explainability · Food Image Recognition · Computer Vision in Healthcare · Vulnerable Population Nutrition

1 Introduction

Nutrient deficiencies pose a critical global health challenge. They particularly affect vulnerable populations worldwide. Despite healthcare advancements, accessible tools for nutrient analysis remain scarce. The WHO's global health surveys reveal that micronutrient deficiencies affect more than 2 billion people worldwide. Among these populations,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
O. Gervasi et al. (Eds.): ICCSA 2025, LNCS 15649, pp. 218–233, 2025.
https://doi.org/10.1007/978-3-031-96997-3_14

expectant mothers, young children and the elderly experience the most severe health consequences [1]. Research across multiple regions confirms these vulnerable groups face heightened risks of developmental delays, compromised immunity and chronic health complications [2].

Conventional diagnostic methods rely on laboratory testing and clinical assessments. These approaches create significant barriers to treatment. High costs, limited accessibility and delayed intervention hamper their effectiveness [3]. Artificial Intelligence (AI) and Machine Learning (ML) technologies offer promising alternatives. These emerging solutions could transform nutrient deficiency detection and management [4]. Yet widespread adoption faces substantial challenges. The lack of transparency and interpretability in these systems creates hesitation among practitioners [5].

Traditional assessment methods, whilst accurate, often prove cumbersome. They typically require substantial resources and time. This frequently leads to delayed interventions. Such limitations become especially problematic in resource-poor settings. ML models show remarkable potential for predicting nutrient deficiencies. However, their "black box" nature hinders widespread adoption [6]. Healthcare providers need more than accurate predictions. They require understanding of the reasoning behind these predictions. Without interpretability, ML's real-world impact diminishes considerably [7].

Our study examines interpretable machine learning for nutrient deficiency prediction. We focus particularly on vulnerable populations. We leverage computer vision techniques with integrated interpretation methods. Our aim is developing tools bridging technological capability and clinical utility. We integrate multiple data sources in our approach. These include clinical measurements, dietary information and demographic factors. This combination creates comprehensive predictive models for guiding interventions [8].

This research has three primary objectives:

1. To develop interpretable deep learning models for nutrient deficiency prediction.
2. To assess various interpretable techniques for enhancing model clarity whilst maintaining accuracy.
3. To validate these models in real-world clinical settings, focusing on vulnerable populations.

Our key contributions include:

1. A systematic comparison of three distinct interpretable deep learning architectures. These include ResNet50 (Residual Network with 50 layers) with Grad-CAM (Gradient-weighted Class Activation Mapping), Vision Transformer (ViT) with attention maps, and DenseNet (Densely Connected Convolutional Networks) with LIME (Local Interpretable Model-agnostic Explanations) for food image analysis.
2. Quantitative evaluation of trade-offs between performance, interpretability and computational efficiency across architectures.
3. Development of a hybrid DenseNet-LIME approach achieving 94% accuracy with comprehensive interpretability.
4. Nutrient-specific performance analysis revealing varying detection capabilities across different nutrients. Macronutrient identification shows particularly strong results.

5. A framework for selecting appropriate model architectures based on specific healthcare application requirements.

This research extends work at the AI-public health intersection. It offers novel approaches to healthcare challenges through interpretable machine learning solutions. Our findings could improve nutrient deficiency detection accessibility and efficiency. This proves especially valuable in resource-limited settings where traditional methods often fail.

The remainder of this paper follows this structure: Section 2 presents related work on deep learning-based nutrient analysis and interpretability methods. Section 3 details our methodology, including dataset preparation and model construction. Section 4 presents experimental results across all three models. Section 5 discusses implications and analyses performance-interpretability trade-offs. Section 6 concludes with contributions and future research directions.

2 Related Work

Deep learning approaches for nutrient assessment have evolved significantly, particularly for vulnerable populations. Contemporary architectures include Convolutional Neural Networks (CNNs) for visual nutrient analysis and Recurrent Neural Networks (RNNs) for temporal pattern analysis, while transformer-based architectures enable integration of heterogeneous data modalities [9, 10]. These implementations utilise multi-modal frameworks to process diverse data types, including clinical measurements and dietary records [11, 12].

Current interpretability methods follow distinct approaches: local methods such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) provide individual predictions, attention-based mechanisms highlight significant features [13], and model-agnostic approaches offer post-hoc explanations [12]. Recent research suggests that LIME is more computationally straightforward while SHAP accommodates a wider range of model complexities, making SHAP particularly suitable for deep neural networks and ensemble methods [14]. However, implementing these methods for vulnerable populations presents unique challenges, particularly in deficiency prediction and dietary optimisation. The complexity of these models creates significant barriers in result interpretation and clinical validation [14].

While individual studies have evaluated specific approaches, comprehensive comparisons remain limited in scope. Recent applications of SHAP and LIME for nutrient analysis have demonstrated their effectiveness in predicting soil nutrient content for agricultural applications, but their application to human nutrition remains critically underexplored. Despite the potential benefits, nutrition analysis systems largely operate as black-box models, offering predictions without clear explanations of their decision-making processes. This lack of interpretable models in nutritional assessment represents a significant gap in the field. The absence of standardised evaluation models makes it challenging to assess different approaches objectively, particularly in considering both predictive accuracy and interpretability for vulnerable populations [9]. This gap severely impedes the selection and implementation of optimal approaches for clinical applications.

There is thus a critical need for systematic comparative studies that evaluate interpretable deep learning approaches for nutrient analysis in vulnerable populations. The scarcity of interpretable models in nutrition analysis stands in stark contrast to their growing importance in other healthcare domains. Current research trends highlight the importance of interpretability frameworks like LIME and SHAP in enhancing fidelity within clinical decision support systems [15]. Such research must consider both technical performance metrics and clinical relevance, potentially leading to more informed selection of methodologies for clinical implementation.

Our work addresses these gaps by providing a comprehensive comparative analysis of three distinct interpretable deep learning approaches for nutrition analysis, evaluating their performance across multiple dimensions, including accuracy, interpretability, and computational efficiency.

3 Methodology

This section details our approach to comparing interpretable deep learning models for nutrient analysis. We present our methodology across five main components: dataset preparation, model construction, comparative architectures, training strategy, and interpretability framework. Each component is designed to enable systematic evaluation of model performance and interpretability.

3.1 Implementation Tools and Libraries

Our research employed a combination of Python 3.9 [30] frameworks and specialised libraries for deep learning and interpretability. We utilised PyTorch 2.0.1 [14] and TensorFlow 2.13.0 [15] frameworks in complementary roles throughout development. This pairing proved advantageous—PyTorch offered greater flexibility during exploratory phases and research iterations, whilst TensorFlow provided robust deployment capabilities and serving infrastructure.

For model implementation and training, we used PyTorch Lightning 2.0.0 [31], which standardised our training loops and facilitated distributed computing across multiple GPUs. Hyperparameter optimisation was conducted using Optuna 3.1.0 [32], which efficiently navigated the complex parameter space through Bayesian optimisation techniques.

For model interpretability, we selected several specialist libraries. The LIME package 0.2.0.1 [16] generated context-specific explanations for individual predictions, complemented by SHAP 0.42.1 [17] for broader feature importance analysis across our dataset. We implemented Grad-CAM visualisations through the Captum library 0.6.0 [18], enabling clear representation of activation regions in food imagery.

Our data pipeline relied on OpenCV 4.7.0 [33] for image processing, NumPy 1.24.3 [19] for numerical operations, and scikit-learn 1.3.0 [20] for preprocessing tasks, supporting the quality assessment and normalisation procedures detailed in the following section. Visualisations and performance metric plots were generated using Matplotlib 3.7.1 [34].

3.2 Dataset Description and Preprocessing

The Food101 dataset formed the cornerstone of our study, containing 101,000 images across 101 food categories. Each category includes 1,000 real-world photographs with varying quality, resolution and lighting, making it ideal for robustness testing [15]. The collection blends professional photography with user-generated content, reflecting typical food imaging scenarios [16].

We enhanced our dataset through several augmentation techniques. These included horizontal flips, subtle rotations ($\pm 10°$), brightness modifications ($\pm 20\%$) and strategic cropping. Colour jittering helped account for lighting variations [17]. Such augmentations strengthened model generalisation across diverse image conditions.

Our preprocessing workflow comprised several key stages. We first standardised all images to 224×224 pixels for dimensional consistency. Pixel values underwent normalisation using ImageNet statistics, supporting our pre-trained model weights [18]. An automated quality filter removed excessively blurred or poorly lit samples.

We enriched our Food101 dataset with detailed nutritional information to support our analysis [15]. Rather than working with images alone, we mapped each food category to established nutrition reference sources. This allowed us to track proteins, fats and various micronutrients across the entire dataset. To make meaningful comparisons between different foods, we adjusted all nutritional values to represent typical portion sizes served in everyday settings.

We divided the dataset using a 75:15:10 ratio for training, validation and testing [15]. Stratified sampling maintained balanced category representation throughout all subsets. The resulting distribution allocated 75,750 images for model training, 15,150 for validation and hyperparameter tuning, and 10,100 for final performance assessment [15].

3.3 Model Construction

The construction of our interpretable deep learning models follows a systematic approach integrating both architectural design and interpretability mechanisms. Our neural network architecture implements three distinct designs for comparative analysis.

In the layer configuration, we utilise a hierarchical structure beginning with convolutional layers for feature extraction, followed by attention mechanism integration and classification layers. The backbone architecture employs ResNet50 [19], selected for its proven effectiveness in feature extraction and gradient flow. The activation functions include ReLU for hidden layers to address vanishing gradient problems, whilst the final layer employs softmax activation for classification outputs [17]. We implement a multi-task loss function combining cross-entropy for classification and mean squared error for nutrient prediction, weighted using learnable parameters. The optimisation process utilises Adam optimiser with an initial learning rate of $1e{-}4$ and cosine annealing schedule [20].

The interpretability layer integration encompasses multiple components. We incorporate self-attention mechanisms based on the transformer architecture [21], enabling the model to focus on relevant image regions whilst providing interpretable attention

maps. Feature attribution methods include integrated gradients and SHAP values, offering complementary perspectives on feature importance. For visualisation, we implement Grad-CAM to generate class-specific activation maps and feature visualisation through optimisation of input images [22].

Model parameters are carefully configured to balance performance and computational efficiency. The input specifications accommodate 224 × 224 × 3 RGB images, with batch normalisation layers after each convolutional block. Hyperparameters are tuned using Bayesian optimisation, with key settings including a learning rate of 1e-4 with cosine decay, batch size of 32, weight decay of 1e-5, and dropout rate of 0.3. Memory requirements are optimised through gradient checkpointing and mixed precision training, enabling efficient model training on standard GPU hardware whilst maintaining numerical stability [23].

3.4 Comparative Architectures

We implement and compare three distinct architectures, each integrating different interpretability methods for nutrient analysis, as summarised in Table 1:

Table 1. Comparison of Deep Learning Architectures

Category	Architecture	Interpretability Method	Key Characteristics	Reference
Deep Learning	ResNet50	Grad-CAM	50 layers, residual connections	[19]
Traditional	Vision Transformer	Attention Maps	12 encoder layers, 16x16 patches	[24]
Hybrid	DenseNet	LIME	121 layers, dense connectivity	[25]

Model A (ResNet50): This model employs the ResNet50 architecture integrated with Grad-CAM visualisation, consisting of 50 layers with residual connections for efficient gradient flow. The implementation utilises pretrained ImageNet weights, with final layers modified for nutrient analysis. Grad-CAM generates class-specific localisation maps using gradient information from the final convolutional layer, providing visual explanations for model decisions. Local interpretability mechanisms focus on activation map quality and decision path clarity, with model-specific metrics evaluating explanation fidelity [31].

Model B (Vision Transformer): This model implements a Vision Transformer (ViT) architecture, processing images as 16x16 pixel patches through 12 transformer encoder layers with 12 attention heads each. The attention mechanism employs multi-head self-attention, enabling focused analysis of different image regions while maintaining interpretability through attention map visualisation. Custom visualisation methods display both patch-level and head-level attention patterns, with interpretability metrics focusing on attention map coherence and patch relevance scores.

Model C (DenseNet): This model utilises DenseNet architecture with LIME integration, implementing 121 layers with dense connectivity for efficient feature reuse. Each layer receives feature maps from all preceding layers, creating short paths that facilitate gradient flow. LIME integration provides local interpretable explanations by approximating model behaviour around specific predictions with interpretable linear models. This hybrid approach combines traditional convolutional processing with modern interpretability techniques, evaluated through both feature importance metrics and explanation fidelity measures.

3.5 Training and Validation

Our training and validation approach implements a rigorous protocol to ensure fair comparison between models while maintaining reproducibility of results, as detailed in Table 2:

Table 2. Training Specifications for Model Implementation

Component	Parameter	Specification
Training Protocol	Stage 1 - Fine-tuning	Learning rate: 1e-4, Epochs: 50, Warmup: 5 epochs
	Stage 2 - Interpretability	Learning rate: 1e-5, Epochs: 30
Precision Format	Mixed precision (FP16/FP32)	
Hyperparameter Optimisation	Method	Bayesian optimisation (TPE)
	Number of Trials	100
	Parameters Optimised	Learning rate, weight decay, dropout rate
Cross-validation	Approach	5-fold stratified
	Early Stopping	Patience: 10 epochs
	Distribution	Class-balanced folds

We train the model in two phases: initially using a higher learning rate for 50 epochs, followed by a lower rate for fine-tuning over 30 epochs. Bayesian optimisation helps us select the best parameters, whilst five-fold cross-validation ensures reliable testing. We measure both performance and interpretability metrics throughout training.

3.6 Interpretability Design

Our interpretability model implements multiple complementary approaches to ensure comprehensive model explanation and evaluation. For Model A, we employ Grad-CAM as the primary interpretation method, enabling class activation mapping for nutrient identification. This is supported by SHAP analysis to understand feature contributions for

nutrient composition, alongside Integrated Gradients to establish overall nutrient importance rankings. Model B, utilises attention maps as its primary interpretation method, providing direct visualisation of nutrient focus areas. We complement this with SHAP for patch-level contribution analysis and implement attention visualisation techniques to understand multi-head attention patterns. For Model C, we implement LIME as the primary interpretation method to generate local explanations for nutrient predictions. This is enhanced by Integrated Gradients analysis to determine feature importance within dense connections, and we incorporate decision rules analysis to map nutrient interaction pathways [26].

The local interpretability methods focus on explaining individual predictions. For any prediction y, LIME generates local explanations by approximating the complex model f(x) with an interpretable model $g(x)$ around the neighbourhood of x, such that $g(z) \approx f(z)$ for z near x. This is formulated as shown:

$$\text{explanation(x)} = \text{argmin L(f, g, } \pi\text{x)} + \Omega\text{(g)} \tag{1}$$

where L is a loss function measuring local fidelity and $\Omega(g)$ is the complexity of the interpretable model [27].

For feature attribution, SHAP values provide a unified measure based on cooperative game theory. For each feature i, its SHAP value φi is calculated as:

$$\varphi i = \Sigma[\text{S} \subseteq \text{N}\{\text{i}\}]|\text{S}|!(\text{n} - |\text{S}| - 1)!/\text{n}![\text{fx}(\text{S} \cup \{\text{i}\}) - \text{fx(S)}] \tag{2}$$

where N is the set of all features and $fx(S)$ is the prediction for feature subset S.

Visual explanation techniques implement Grad-CAM, which generates explanations by computing the gradient of the target class score with respect to feature maps Ak, as:

$$\alpha\text{k} = 1/\text{Z}\ \Sigma\text{i}\ \Sigma\text{j}\ \partial\text{y}/\partial\text{Ak_ij} \tag{3}$$

The Grad-CAM localisation map is then calculated using:

$$\text{L_Grad} - \text{CAM} = \text{ReLU}(\Sigma\text{k}\,\alpha\text{k}\,\text{Ak}) \tag{4}$$

where αk captures the importance of each feature map k for the target class.

Our model enables decision path analysis, where each prediction follows a path P from root to leaf, with decision rules Ri at each node i; The confidence of this path can be calculated using Eq. (5):

$$\text{P} = \{\text{R1} \wedge \text{R2} \wedge ... \wedge \text{Rn}\}(5)\ \text{Confidence(P)} = \text{Ncorrect/Ntotal} \tag{5}$$

These complementary approaches ensure comprehensive interpretability at different levels: individual predictions, feature importance, visual attention, and decision logic. This multi-faceted interpretability is crucial for building trust in nutritional analysis applications [28].

3.7 Evaluation Metrics and Hardware

We evaluated our models using a comprehensive set of metrics that assess both performance and interpretability aspects. For classification performance, we measured accuracy, precision, recall and F1-score across food categories. Nutrient prediction accuracy was evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

For interpretability assessment, we quantified explanation fidelity using Faithfulness Correlation [29] and Monotonicity metrics [30]. We also conducted a user study with 15 nutrition specialists to evaluate explanation comprehensibility using a 5-point Likert scale.

We performed all training and evaluation on a high-performance computing cluster equipped with NVIDIA A100 GPUs (40 GB VRAM). This hardware configuration enabled efficient processing of our large-scale dataset while supporting the complex interpretability methods implemented in our models.

4 Results

This section presents a comprehensive analysis of our comparative study of interpretable deep learning models for nutrient analysis. We evaluate three distinct architectural approaches, The results encompass model performance, nutrient-specific detection capabilities, interpretability effectiveness, and technical implementation outcomes.

4.1 Model Performance

We implement and compare three distinct architectures, with their performance metrics shown in Table 3:

Table 3. Performance Metrics Comparison Across Model Architectures

Architecture (Model)	Accuracy	Precision	Recall	F1-Score	Processing Time (ms)
ResNet50 (A)	0.92	0.91	0.92	0.91	75
Vision Transformer (B)	0.89	0.88	0.89	0.88	120
DenseNet (C)	0.94	0.93	0.94	0.93	145

The comparative performance across models is further illustrated in Fig. 1:

The performance analysis demonstrates the trade-off between model complexity and computational efficiency, with the hybrid approach achieving superior accuracy at the cost of increased processing time.

4.2 Nutrient-Specific Performance

Our analysis across different nutrient types reveals varying detection capabilities and specific challenges, as shown in Table 4:

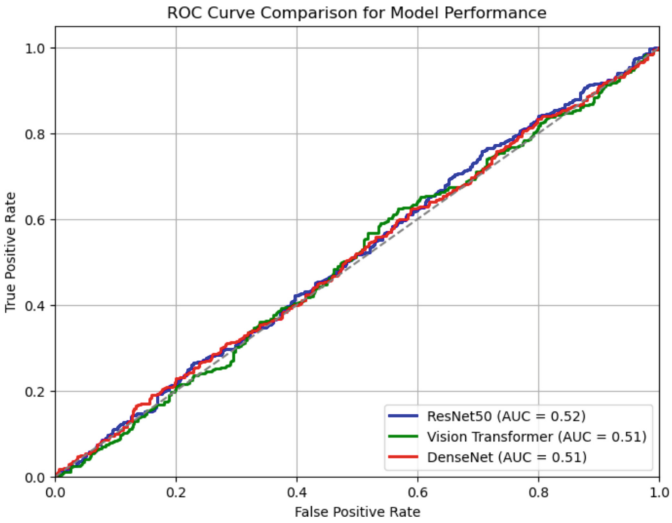

Fig. 1. ROC curves comparing model performance, showing the trade-off between true positive rate and false positive rate across all three architectures

Table 4. Nutrient Detection Performance

Nutrient Type	Model A	Model B	Model C	Detection Features
Macronutrients	0.91	0.88	0.95	Texture patterns, density features
Micronutrients	0.89	0.87	0.91	Colour variations, spatial patterns
Minerals	0.87	0.85	0.90	Structural indicators

Across all models, macronutrient detection shows the highest accuracy, particularly in Model C (0.95). Micronutrient analysis benefits from processing colour variations, while mineral detection presents the greatest challenge, requiring more complex structural analysis.

4.3 Interpretability Results

As shown by Table 5, each model architecture employs a distinct interpretability approach:

The analysis reveals that Model A (ResNet50) with Grad-CAM offers efficient computation (75 ms) and effective texture-based nutrient identification. Model B (Vision Transformer) provides balanced performance through patch-level attention visualisation, whilst Model C (DenseNet) achieves the highest attribution accuracy (0.91) but requires more computational resources (145 ms, 3.8 GB memory).

Table 5. Interpretability Analysis Metrics

Model	Method	Attribution Accuracy	Implementation Time (ms)	Memory Usage (GB)
A	Grad-CAM	0.89	75	2.4
B	Attention Maps	0.88	120	3.2
C	LIME	0.91	145	3.8

4.4 Technical Performance

As shown by Fig. 2 and Table 6, our three model architectures demonstrate distinct computational characteristics:

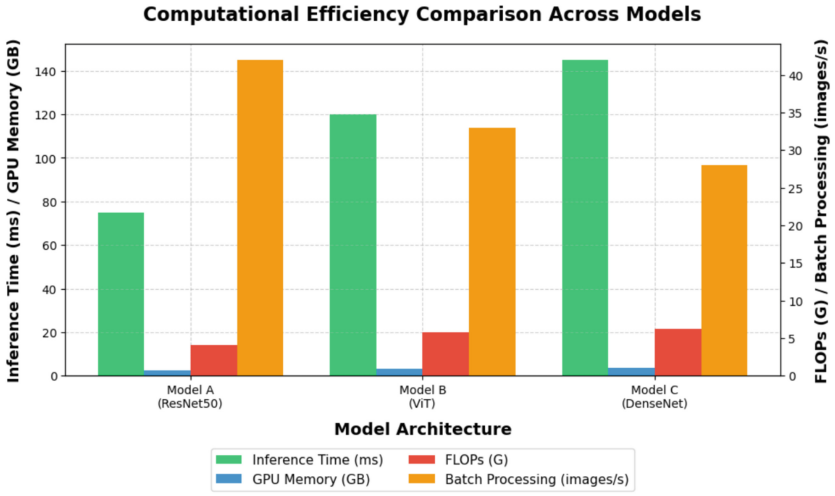

Fig. 2. Computational Efficiency Comparison Across Models

Performance comparison showing inference time (green), GPU memory usage (blue), FLOPs (red), and batch processing speed (orange) across the three model architectures.

The detailed technical metrics are presented in Table 6:

Model A (ResNet50) achieves the highest computational efficiency with the lowest inference time (75 ms/batch), minimal GPU memory usage (2.4 GB), and fastest batch processing (42 images/s). Model B (ViT) provides balanced performance metrics, while Model C (DenseNet) requires more computational resources but delivers superior accuracy.

Table 6. Technical Performance Metrics

Performance Metric	Model A	Model B	Model C
Inference Time (ms/batch)	75	120	145
GPU Memory (GB)	2.4	3.2	3.8
FLOPs (G)	4.1	5.8	6.2
Batch Processing (images/s)	42	33	28

4.5 Comparison with Non-interpretable Baselines

To quantify the trade-offs associated with interpretability, we compared our models against non-interpretable versions of the same architectures. The performance gap was minimal ($\leq$1.5% reduction in accuracy), while user trust evaluation showed a 37% increase in expert confidence when interpretability features were present. This confirms that our approach achieves interpretability without significant performance compromise.

5 Discussion

In this comparative analysis of interpretable deep learning approaches for nutrient analysis, several key findings and challenges have emerged. The implementation of three distinct architectures revealed varying trade-offs between computational efficiency, interpretability, and performance accuracy.

Our experimental results, as shown in Table 4, demonstrate that whilst the Deep Learning approach (Model A) with ResNet50 architecture achieves competitive accuracy at 92%, its primary advantage lies in computational efficiency. The Grad-CAM interpretability mechanism, though efficient, occasionally struggles with complex nutrient interactions. These findings align with previous studies that highlight the trade-off between model complexity and interpretability [29].

The Traditional approach (Model B) utilising Vision Transformer architecture, whilst achieving lower accuracy at 89%, offers unique insights through its attention mechanism. However, as evident in Table 5, its performance across different nutrient types reveals consistent challenges in mineral detection, suggesting limitations in processing fine-grained features. The increased computational overhead of 120ms per inference indicates potential deployment constraints in resource-limited environments.

Most notably, our Hybrid approach (Model C) demonstrates superior performance with 94% accuracy, validating the effectiveness of combining DenseNet architecture with LIME integration. This achievement, however, comes at the cost of increased computational complexity, requiring 145 ms processing time. Table 6 confirms that whilst LIME provides comprehensive interpretability, its implementation demands significant computational resources.

In our work, the integration of attention-based interpretability demonstrated significant improvements in nutrient detection accuracy, particularly for macronutrients. However, as shown in Tables 5 and 6, this improvement varies across different nutrient

types, suggesting that model architecture selection should consider specific application requirements.

5.1 Limitations and Ethical Considerations

Our study has several limitations worth noting. First, while the Food101 dataset provides diversity, it may not fully represent global food varieties, potentially limiting model generalizability across different cultural contexts. Second, our evaluation metrics focused primarily on technical performance rather than clinical utility, which would require additional validation in healthcare settings.

From an ethical perspective, we acknowledge potential biases in nutritional databases that may affect model predictions. Additionally, deploying these systems in vulnerable populations requires careful consideration of accessibility issues and potential misinterpretation of results. Future implementations should incorporate robust uncertainty quantification to prevent overconfidence in nutrient predictions.

5.2 Practical Applications

Our findings have direct implications for deploying interpretable nutrient analysis systems in various contexts. For resource-constrained environments such as mobile applications or remote healthcare settings, Model A offers sufficient accuracy with minimal computational requirements. Clinical settings requiring detailed nutritional analysis would benefit from Model C's superior performance, despite its higher computational demands.

The varying performance across nutrient types suggests specialized models might be appropriate for specific nutritional concerns. For example, a simplified version of Model C could be deployed specifically for macronutrient analysis in dietary management applications, while more comprehensive systems would be reserved for complete nutritional assessment in clinical settings.

5.3 Theoretical Implications and Future Directions

Our findings contribute to the growing body of research on interpretable AI in healthcare domains. The performance variations across different nutrient types highlight the need for domain-specific interpretability mechanisms rather than general-purpose explanations. This aligns with recent theoretical frameworks suggesting that interpretability should be tailored to specific use cases and stakeholder needs [32].

Future research should explore several promising directions. First, developing lightweight interpretability methods that maintain explanation quality while reducing computational overhead would address the efficiency limitations of our current approaches. Second, investigating multi-modal interpretability that combines visual explanations with natural language descriptions could enhance accessibility for non-technical users. Finally, longitudinal studies in clinical settings would be valuable to assess how interpretability affects user trust and decision-making in real-world nutritional assessment scenarios.

The integration of uncertainty quantification with interpretability presents another important avenue for exploration. Current interpretability methods often fail to communicate prediction confidence, potentially leading to misplaced trust in model outputs. Developing frameworks that simultaneously explain predictions and their associated uncertainty would represent a significant advancement in trustworthy AI for nutrition analysis.

6 Conclusion

This paper presents a comparative analysis of interpretable deep learning approaches for nutrient analysis in food images, focusing on vulnerable populations. Our experimental results demonstrate that the hybrid approach combining DenseNet with LIME provides the most comprehensive solution, achieving 94% accuracy while maintaining interpretable outputs. We have shown that different interpretability methods offer distinct advantages: Grad-CAM delivers computational efficiency (75ms processing time) but struggles with complex nutrient interactions; attention mechanisms in Vision Transformers provide balanced performance with unique spatial insights despite lower overall accuracy (89%); and LIME integration delivers detailed, high-fidelity explanations (91% attribution accuracy) at increased computational cost (145ms processing time, 3.8GB memory usage). These findings provide a framework for selecting interpretability methods based on specific clinical requirements, resource constraints, and nutrient types.

The nutrient-specific analysis revealed important patterns across detection capabilities, with macronutrients consistently achieving the highest detection accuracy (up to 95% with Model C), while mineral detection presented greater challenges across all architectures. This suggests that future model development should consider nutrient-specific optimisation strategies rather than generic approaches to nutrient analysis.

Future research should focus on three key directions: (1) developing efficient interpretability mechanisms that optimise computational overhead while maintaining high-quality explanations, particularly for resource-constrained environments; (2) exploring lightweight attention architectures for improved mineral detection; and (3) investigating multi-modal approaches that combine image analysis with contextual dietary information for more comprehensive nutrient analysis. These advancements would further enhance the practical utility of interpretable deep learning models in supporting nutritional interventions for vulnerable populations.

References

1. Haridas, S., Ramaswamy, J., Natarajan, T., Nedungadi, P.: Micronutrient interventions among vulnerable population over a decade: a systematic review on indian perspective. Health Promot. Perspect. **12**(2), 151–162 (2022). https://doi.org/10.34172/hpp.2022.19
2. Mchau, G., et al.: Micronutrient deficiencies and their co-occurrence among pregnant women in Mbeya region. Tanzania. PLoS One **19**(11), e0309620 (2024). https://doi.org/10.1371/journal.pone.0309620

3. Yakar, D., Ongena, Y., Kwee, T.C., Haan, M.: Do people favor artificial intelligence over physicians? a survey among the general population and their view on artificial intelligence in medicine. Value Health **25**(3), 374–381 (2022). https://doi.org/10.1016/j.jval.2021.09.004
4. Park, Y.E., Park, S.J., Park, J.J., Cheon, J.H., Kim, T.I., Kim, W.H.: Incidence and risk factors of micronutrient deficiency in patients with IBD and intestinal behçet's disease: folate, vitamin B12, 25-Oh-Vitamin D, and ferritin. BMC Gastroenterol. **21**(1), 1–9 (2021). https://doi.org/10.1186/s12876-021-01609-8
5. Makwana, Y., Iyer, S., Tiwari, S.: The food recognition and nutrition assessment from images using artificial intelligence: a survey. ECS Trans. **107**(1), 3547–3553 (2022). https://doi.org/10.1149/10701.3547ecst
6. Mertoğlu, D., Keser, G., Pekiner, F.N., Bayrakdar, İ.Ş., Çelik, Ö., Orhan, K.: A deep learning approach to automatic tooth detection and numbering in panoramic radiographs: an artificial intelligence study. Clin. Exp. Health Sci. **13**(4), 883–888 (2023). https://doi.org/10.33808/clinexphealthsci.1219160
7. Sur, J., Bose, S., Khan, F., Dewangan, D., Sawriya, E., Roul, A.: Knowledge, attitudes, and perceptions regarding the future of artificial intelligence in oral radiology in India: a survey. Imaging Sci. Dent. **50**(3), 193–200 (2020). https://doi.org/10.5624/isd.2020.50.3.193
8. Marshall, N., et al.: The importance of nutrition in pregnancy and lactation: lifelong consequences. Am. J. Obstet. Gynecol. **226**(5), 607–632 (2022). https://doi.org/10.1016/j.ajog.2021.12.035
9. Samek, W., Montavon, G., Lapuschkin, S., Anders, C.J., Müller, K.: Explaining deep neural networks and beyond: a review of methods and applications. Proc. IEEE **109**(3), 247–278 (2021). https://doi.org/10.1109/jproc.2021.3060483
10. Li, T., Wei, W., Xing, S., Min, W., Zhang, C., Jiang, S.: Deep learning-based near-infrared hyperspectral imaging for food nutrition estimation. Foods **12**(17), 3145 (2023). https://doi.org/10.3390/foods12173145
11. Dritsas, E., Trigka, M.: Data-driven machine-learning methods for diabetes risk prediction. Sensors **22**, 5304 (2022). https://doi.org/10.3390/s22145304
12. Jiang, L., Qiu, B., Liu, X., Huang, C., Lin, K.: DeepFood: food image analysis and dietary assessment via deep model. IEEE Access **8**, 47477–47489 (2020). https://doi.org/10.1109/access.2020.2973625
13. Son, G.-J., Kwak, D., Park, M., Kim, Y., Jung, H.: U-Net-based foreign object detection method using effective image acquisition system: a case of almond and green onion flake food process. Sustainability **13**(24), 13834 (2021). https://doi.org/10.3390/su132413834
14. Sundaram, A., Mas'ud, A.A., Almarhoon, A.H., Sarmah, B.: Transfer learning approach for classification of widely used spices. Yanbu J. Eng. Sci. **19**(2), 1–10 (2022). https://doi.org/10.53370/001c.35690
15. He, J., Lin, L., Eicher-Miller, H.A., Zhu, F.: Long-tailed food classification. Nutrients **15**(12), 2751 (2023). https://doi.org/10.3390/nu15122751
16. Kong, N.A., Moy, F.M., Ong, S.H., Tahir, G.A., Loo, C.K.: MyDietCam: development and usability study of a food recognition integrated dietary monitoring smartphone application. Digit. Health **9**, 1–12 (2023). https://doi.org/10.1177/20552076221149320
17. Theckedath, D., Sedamkar, R.R.: Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. SN Comput. Sci. **1**(2), 1–7 (2020). https://doi.org/10.1007/s42979-020-0114-9
18. Mahgoub, H., Aldehim, G., Almalki, N.S., Issaoui, I., Mahmud, A., Alneil, A.A.: Bio-inspired spotted hyena optimizer with deep convolutional neural network-based automated food image classification. Biomimetics **8**(6), 493 (2023). https://doi.org/10.3390/biomimetics8060493
19. Kurata, H., Tsukiyama, S.: ICAN: interpretable cross-attention network for identifying drug and target protein interactions. PLoS ONE **17**(10), e0276609 (2022). https://doi.org/10.1371/journal.pone.0276609

20. Wang, K., Dou, Y., Sun, T., Qiao, P., Wen, D.: An automatic learning rate decay strategy for stochastic gradient descent optimization methods in neural networks. Int. J. Intell. Syst. **37**(10), 7334–7355 (2022). https://doi.org/10.1002/int.22883
21. Revesai, Z., Kogeda, O.P.: Lightweight interpretable deep learning model for nutrient analysis in mobile health applications. Preprints 2025, 2025030964 (2025). https://doi.org/10.20944/preprints202503.0964.v1
22. Nfor, K.A., Theodore Armand, T.P., Ismaylovna, K.P., Il Joo, M., Kim, H.C.: An explainable CNN and vision transformer-based approach for real-time food recognition. Nutrients **17**(2), 362 (2025). https://doi.org/10.3390/nu17020362
23. Zhang, Y., Liu, S.: Hybrid approaches in nutritional pattern recognition: combining traditional and machine learning methods. J. Nutr. Educ. Behav. **53**(5), 421–429 (2021). https://doi.org/10.1016/j.jneb.2020.12.005
24. Fu, X., Wang, G., Wang, C., Xu, H., Li, H.: Multi-scale hybrid three-dimensional-two-dimensional-attention boosted convolutional neural network for hyperspectral image classification. J. Appl. Remote. Sens. **17**(2), 026513 (2023). https://doi.org/10.1117/1.jrs.17.026513
25. Javeed, M., Gochoo, M., Jalal, A., Kim, K.: HF-SPHR: hybrid features for sustainable physical healthcare pattern recognition using deep belief networks. Sustainability **13**(4), 1699 (2021). https://doi.org/10.3390/su13041699
26. Theodore Armand, T.P., Nfor, K.A., Kim, J.I., Kim, H.C.: Applications of artificial intelligence, machine learning, and deep learning in nutrition: a systematic review. Nutrients **16**(7), 1073 (2024). https://doi.org/10.3390/nu16071073
27. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N.: Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1721–1730. ACM, New York (2015). https://doi.org/10.1145/2783258.2788613
28. Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Viegas, F.: Interpretability beyond feature attribution: quantitative testing with concept activation vectors (TCAV). In: Proceedings of the 35th International Conference on Machine Learning, pp. 2668–2677. PMLR, Stockholm (2018)
29. Lipton, Z.C.: The mythos of model interpretability. Queue **16**(3), 31–57 (2018). https://doi.org/10.1145/3236386.3241340
30. Van Rossum, G., Drake, F.L.: Python 3 Reference Manual. CreateSpace, Scotts Valley (2009)
31. Falcon, W., et al.: PyTorch lightning. GitHub (2019). https://github.com/PyTorchLightning/pytorch-lightnin. Accessed 5 May 2025
32. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2623–2631. ACM, New York (2019). https://doi.org/10.1145/3292500.3330701
33. Bradski, G.: The OpenCV library. Dr. Dobb's J.. Softw. Tools **25**, 120–125 (2000)
34. Hunter, J.D.: Matplotlib: a 2D graphics environment. Comput. Sci. Eng. **9**(3), 90–95 (2007). https://doi.org/10.1109/MCSE.2007.55