International Journal of Science and Technology (IJST) | ISSN: 2049-7318 | 2(3):13-19,2026

Multi-Method Vitamin Deficiency Detection System Using Ensemble Convolutional Neural Networks

Dr. K. Sai Venu Prathap[1], V. Ajay Kumar[2], P. Dhanush[3]

[1.] Associate Professor, Dept. of ECE, Aditya College of Engineering, Madanapalle
[2,3] Students, Dept. of ECE, Aditya College of Engineering, Madanapalle

Abstract: Over two billion people worldwide suffer from vitamin deficiencies, which constitute a serious global health issue. However, the majority of screening techniques used today depend on intrusive blood tests, which can be expensive, time-consuming, and unavailable in many situations. In order to identify common vitamin deficiencies (A, B, C, D, or a Normal/healthy classification), this paper presents a non-invasive screening system that combines five different CNN architectures: a custom-built CNN, VGG16, ResNet50, MobileNetV2, and EfficientNetB0. The system analyzes dermatological images from five different parts of the body: skin, lips, tongue, eyes, and nails. In order to increase overall accuracy, we aggregated the predictions of all five models using soft-voting and employed a two-phase transfer learning approach to construct this system. Additionally, we used Grad-CAM visualization so In order to maintain the interpretability of the system's choices, we also included Grad-CAM visualization, which enables users to identify which areas of an image affected the diagnosis. With real-time camera input and automatic PDF report creation, the entire pipeline is implemented as a Flask-based web application, making it useful for practical applications. The ensemble model outperformed the individual models and previously reported methods in this field in testing, with 93.8% accuracy and an AUC of 0.951. These findings imply that the system provides a dependable, comprehensible, and affordable substitute for conventional blood-based diagnostics, with great potential for widespread use in both clinical and non-clinical settings.

Index Terms: Convolutional Neural Networks (CNNs); vitamin deficiency detection; ensemble learning; transfer learning; Grad-CAM; explainable AI; dermatological image analysis; Flask web application; EfficientNetB0; real-time inference.

I. INTRODUCTION

Vitamins are essential because they support thousands of biochemical activities that keep the body running smoothly, from immune system protection and normal vision to skin health, blood production, neuron function, and bone strength. We rely nearly exclusively on nutrition to meet our vitamin demands because the human body is unable to generate most vitamins in sufficient levels on its own. Despite this, the prevalence of vitamin and mineral deficiencies is still alarmingly high, affecting over two billion people worldwide. According to recent data, approximately half of preschool-age children and nearly two-thirds of women of reproductive age have at least one micronutrient deficiency [15]. These deficiencies lead to major health consequences, such as anemia, stunted growth, blindness, decreased immunity, and increased rates of child death, in addition to child morality.
About 190 million preschool-age children worldwide suffer from vitamin A insufficiency, which is one of the main—and completely avoidable—causes of childhood blindness, according to the World Health Organization [15]. Other vitamin deficiencies have substantial repercussions of their own: inadequate vitamin B can cause megaloblastic anemia and, in more extreme situations, permanent nerve damage; inadequate vitamin D is associated to rickets in infants and osteomalacia in adults. Conversely, a lack of vitamin C causes scurvy, a disorder characterized by bleeding gums and sluggish wound healing. Traditionally, lab blood tests have been used to diagnose vitamin deficiencies. Although these tests are accurate, they have serious disadvantages, including being costly, intrusive, and frequently unaffordable in environments with low resources. It's interesting to note that many of these deficiencies manifest as subtle changes in the skin, lips, tongue, eyes, or nails far before a blood test would detect them. This presents a genuine chance for an image-based, non-invasive method to identify issues early.
Our suggested system seeks to close that precise gap. The system can automatically identify visual signs of vitamin deficiencies by using an AI-driven pipeline that simultaneously examines several body regions. This helps identify possible problems early enough for prompt dietary counselling or medical follow-up, well before they develop into more serious complications.

The following are the main contributions of this study.

- Using dermatological data from various body parts, we create a diverse ensemble of five CNN architectures.
- we fine-tune the models for improved adaptability to this particular medical imaging domain using a two-phase transfer learning technique.
- Instead of treating Grad-CAM explainability as an afterthought, we incorporate it immediately into the system's output, providing a visual explanation for each prediction.
- We develop a production-ready Flask web application that allows for real-time camera input, shows the votes cast by each model in the ensemble, and automatically creates PDF reports for users.

- A comprehensive analysis reveals that the ensemble outperforms all previously published techniques, with 93.8% accuracy and an AUC of 0.951.

II. LITERATURE SURVEY

A. Hybrid Unsupervised-Fuzzy Pipeline for Deficiency Detection

A three-stage hybrid pipeline that combines K-means clustering for image segmentation, a Multi-Layer Perceptron (ANN) for classification, and a fuzzy logic inference engine that converts numerical probabilities into more comprehensible linguistic categories like "mild," "moderate," or "severe" deficiency was proposed by Eldeen et al. [3]. Although the accuracy of the system was only 80.3%, which is not as high as more current methods, the concept of combining fuzzy reasoning and unsupervised segmentation is intriguing and hasn't been thoroughly investigated in the field of explainable medical AI.

B. Multi-Model Clinical Machine Learning

A alternative strategy was used by Sambasivam et al. [1], who developed a tabular machine learning pipeline that processes structured clinical data, such as blood biomarkers, anthropometric measures, and dietary intake records. They obtained a strong accuracy of 84.0% by training three different models: Random Forest, Decision Tree, and Support Vector Regression (SVR). Because its predictions may be linked to certain feature-based split rules, the Decision Tree model in particular provides some inherent interpretability. Nevertheless, the system's utility is somewhat restricted in point-of-care settings where prompt, easily accessible screening is the top goal because it solely relies on laboratory-acquired data.

C. Model Ensemble Fusion for Deficiency Classification

Using a three-model ensemble that blended softmax probability outputs via majority voting or averaging, Wang et al. [7] showed a robust 91.3% accuracy. Although the approach has trade-offs—it is computationally costly to run and, more significantly, it provides no true interpretability—the findings were impressive. The approach would be difficult for clinicians to trust or implement in actual clinical practice since they would not be able to understand why a specific prediction was made.

D. Dense Net for Multi-Region Detection

Robust clinical deployability and laboratory accuracy. Dense Net's dense shortcut connections, which enable gradients to flow more directly to the earliest layers during backpropagation, were utilized by Jayaram et al. [5] for their method. As a result, the model's accuracy across many body regions was a strong 89.1%. There is a noticeable disconnect between the model's excellent performance in a lab setting and its preparedness for real-world clinical use, where interpretability frequently matters just as much as accuracy, because the study did not include any explainable AI component.

Table I. Summary of key literature

Reference	Method	Regions	Accuracy	Real-time	Notes
Eldeen et al. [3], 2020	K-means + ANN + Fuzzy	4 regions	80.3%	No	Interpretable but limited accuracy
Sambasivam et al. [1], 2020	SVR, RF, Decision Tree	Clinical data	84.0%	No	Requires lab blood tests
Wang et al. [7], 2023	3-model Ensemble	Clinical	91.3%	No	No XAI, high compute cost
Jayaram et al. [5], 2024	DenseNet	Multi-region	89.1%	No	No XAI
Bonde et al. [4], 2024	VGG + ResNet (Dual-Stream)	4 regions	87.2%	No	No real-time capability
Supritha et al. [6], 2024	VGG16 + Custom CNN	5 regions	88.41%	No	No XAI

E. Dual-Stream VGG-ResNet Architecture

By combining VGG's proficiency in collecting fine-grained texture information with ResNet's capacity to maintain stable gradient flow across deeper layers, Bonde et al. [4] created a dual-stream architecture that achieves 87.2% accuracy. The method performed well in terms of accuracy, but it has two practical drawbacks: it lacks an explainable AI component, and running two deep networks concurrently is computationally intensive enough to make real-time deployment unfeasible.

III. METHODOLOGY

A. System Workflow and Architecture Overview

A seven-stage integrated pipeline—image capture, the web application interface, pre-processing, CNN ensemble feature extraction, ensemble voting, Grad-CAM visualization, and structured output generation—is the foundation of the suggested Multi-Modal Vitamin Deficiency Detection System. The pipeline is built to be production-ready from the beginning, with each step feeding into the next. Static image uploads and live, real-time camera feeds are supported, and everything is connected by a Flask-based web application.

B. Dataset Preparation

A bespoke dermatological picture collection was created for this work using publicly accessible medical archives, mainly DermNet NZ and Kaggle. The 25,461 photos in the dataset are divided into five classes: Normal, Vitamin A, Vitamin B, Vitamin C, and Vitamin D. The collection covers five anatomical regions: the eyes, lips, tongue, nails, and skin.

All raw photos were manually reviewed prior to training in order to remove duplicates and low-quality examples. After that, the remaining photos were scaled to 224 by 224 pixels, converted to RGB format, and normalized using the usual ImageNet mean and standard deviation values (mean: R=0.485, G=0.456, B=0.406; standard deviation: R=0.229, G=0.224, B=0.225). Lastly, a 70:15:15 ratio was used to divide the dataset into training, validation, and test sets. Stratification was then used to preserve class balance in all three Splits.

Data augmentation was used during training to help alleviate class imbalance and enhance the model's capacity for generalization. Horizontal and vertical flips, random rotations of up to ±25°, width and height shifts (up to 15%), zoom adjustments (up to 15%), brightness variation (ranging from 0.8 to 1.3), and shear transformations (up to 0.1) were all included in the augmentation pipeline. The training set was successfully increased from about 15,000 images to almost 75,000 by applying these augmentations on the fly using Keras's Image Data Generator, providing the models with a considerably richer and more diversified set of examples to learn from.

Table ii. Data augmentation techniques

Augmentation	Parameter	Purpose
Horizontal Flip	Probability = 0.5	Simulate left/right image variation
Rotation	Range = ±25°	Simulate camera tilt
Width/Height Shift	Range = 0.15	Handle off-center subjects
Zoom	Range = 0.15	Handle varying camera distances
Brightness Adjustment	[0.8, 1.3]	Simulate different lighting conditions
Shear Transform	Range = 0.1	Minor affine distortion

C. CNN Architectures

The proposed ensemble brings together five different CNN architectures, each contributing its own representational strengths to the overall classification decision.

1) Custom CNN:

It accepts 224×224×3 input images and processes them through three convolutional blocks with 32, 64, and 128 filters respectively, each using 3×3 kernels and ReLU activation. These are followed by max-pooling layers to progressively reduce spatial dimensions, then a flatten layer, a Dense layer with 128 units paired with a Dropout rate of 0.5 to prevent overfitting, and finally a 5-class Softmax output layer that produces the classification probabilities. With over 1.2 million trainable parameters overall, this model is modest enough to enable quick, domain-specific training without requiring a lot of processing power.

2) VGG16:

The second architecture, VGG16, has about 138 million parameters and was pre-trained on ImageNet. It is divided into five blocks of 13 convolutional layers, each of which uses a uniform 3x3 kernel. Max-pooling and three fully linked layers come next. Skin pallor, tongue fissures, and nail ridging are among the fine-grained texture features that VGG16 is particularly effective at detecting. In this study, the pre-trained feature extraction was preserved by replacing only the top classification layers during fine-tuning.

3) ResNet50:

The third model, ResNet50 (H(x) = F(x) + x), is a 50-layer residual network that uses skip connections to overcome the vanishing gradient issue. ResNet50, which has over 25 million parameters, is particularly good at identifying more intricate structural patterns that simpler designs could overlook, such as nail ridging, variations in conjunctival veins, and minute pigmentation inconsistencies.

4) MobileNetV2:

Using inverted residual blocks with linear bottlenecks and depth-wise separable convolutions, MobileNetV2 adopts a lighter methodology. Compared to the other architectures in the ensemble, it is far more compact, with just roughly 3.4 million parameters. It performed well in spite of its reduced size, achieving 90.8% test accuracy with an inference time of about 35 milliseconds per image—fast enough to be a strong contender for real-time deployment scenarios.

5) EfficientNetB0:

Using a compound scaling strategy that simultaneously balances network depth, width, and input resolution rather than optimizing each separately, EfficientNetB0 completes the ensemble. It achieved 93.8% accuracy on its own, outperforming all other individual models in the ensemble with almost 5.3 million parameters. Its ability to successfully capture multi-scale feature representations enables it to identify both fine, localized signs of inadequacy within the images and broad, global trends.

D. Model Training Procedure

To guarantee uniformity throughout the ensemble, every pre-trained architecture undergoes the identical two-phase transfer learning process. Phase 1, called feature extraction, uses a learning rate of 1e-3, a batch size of 32, and the Adam optimizer to train only the classification head for 10 epochs while the pre-trained convolutional base is completely frozen (trainable=False). Global Average Pooling, a Dense layer with 128 units and ReLU activation, a Dropout layer with a rate of 0.5, and a Dense layer with 5 units and Softmax activation for classification make up the basic structure of this classification head.

Phase 2 is when fine-tuning starts. Here, the network can be trained end-to-end for a further 20 epochs by unfrozening the top 30–50% of the backbone layers. In order to minimize overfitting during this more intense training phase, early halting is implemented with a patience of 5 epochs based on validation loss, and the learning rate is drastically reduced to 1e-5 to avoid upsetting the previously learnt features too severely.

Table iii. CNN architecture comparison

Architecture	Parameters	Test Accuracy	Depth	Key Strength
Custom CNN	~1.2M	90.7%	8 layers	Task-optimized, fastest training
VGG16	~138M	90.7%	16 layers	Texture-rich, fine-detail detection
ResNet50	~25M	90.9%	50 layers	Skip connections, complex patterns
MobileNetV2	~3.4M	90.8%	53 layers	Real-time inference capability
EfficientNet B0	~5.3M	93.8%	82 layers	Compound scaling, highest accuracy
Ensemble (All 5)	~173M total	93.8%	—	Robust, max-voting, AUC=0.951

E. Ensemble Strategy and Grad-CAM Integration

The ensemble integration module loads each of the five models' checkpoints and performs inference on each input image in turn once each model has been trained. The overall ensemble probability distribution is created by normalizing and averaging the five-class probability vectors that each model separately generates. The class with the highest aggregated probability is the final prediction, and that highest probability value is used to determine the confidence score.

In addition, each model's unique argmax vote is stored separately so that the user interface (UI) may show how each model voted independently, providing users with a clearer picture of the ensemble's decision-making process. For explainability, Grad-CAM is implemented using TensorFlow's Gradient Tape API. The first step in the process is to calculate the gradient of the output score for the anticipated class in relation to the activation maps from the final convolutional layer. A set of significance weights, which show how much each feature map contributes to the final prediction, are then created by averaging these gradients channel-wise.

These weights are then used to calculate a weighted sum of the activation maps, which is subsequently enlarged to 224x224 pixels using bilinear interpolation after passing through a ReLU activation to retain only positive contributions. The jet colormap is used to normalize the output to a 0–255 range and render it as a heatmap. The heatmap is then superimposed above the original RGB image with an alpha of 0.45, making the highlighted areas easily apparent while maintaining the underlying image detail.

Table iv. Individual and ensemble model performance

Model	Test Accuracy	AUC	Key Characteristic
Custom CNN	0.907 (90.7%)	0.907	Fastest training, task-specific features
VGG16	0.907 (90.7%)	0.908	Best for texture discrimination
ResNet50	0.909 (90.9%)	0.910	Best structural pattern extraction
MobileNetV2	0.908 (90.8%)	0.909	Fastest inference (~35 ms/image)
EfficientNetB0	0.938 (93.8%)	0.938	Highest individual accuracy
Ensemble (5 models)	0.938 (93.8%)	0.951	Most robust, highest AUC

Table v. Ensemble model — per-class prediction summary

True Class	Correctly Predicted	Most Common Misclassification	Total Test Samples
Normal	92 / 100	8 misclassified as Vitamin A	100
Vitamin A	87 / 100	13 misclassified as Vitamin B	100
Vitamin B	91 / 100	9 misclassified as Vitamin A	100
Vitamin C	90 / 100	10 misclassified as Vitamin D	100
Vitamin D	92 / 100	8 misclassified as Vitamin C	100
OVERALL	452 / 500	Avg. per-class accuracy: 90.4%	500

F. Flask Web Application

The system is implemented as a Flask-based web application with several integrated features intended to make it useful for actual applications. Users can upload pictures, and the system takes care of format validation and safe filename sanitization in the background. Additionally, a real-time camera feed is supported, enabling predictions to change in real time as pictures are taken. An integrated alert system uses Laplacian variance to identify blur and mark low-quality inputs prior to processing in order to assist preserve image quality.

Table vi. Precision, recall, and f1-score — ensemble model

Class	Precision	Recall	F1-Score	Support
Normal	0.94	0.94	0.94	~100
Vitamin A	0.92	0.92	0.92	~100
Vitamin B	0.92	0.92	0.92	~100
Vitamin C	0.94	0.94	0.94	~100
Vitamin D	0.93	0.93	0.93	~100
Macro Average	0.930	0.930	0.930	~500
Weighted Average	0.930	0.930	0.930	~500

The findings are displayed via a dashboard that displays the final forecast, confidence score, individual model votes, probability bar charts, and pertinent expert advice. The ensemble inference pipeline operates concurrently with Grad-CAM production on the backend. The system also offers automated PDF report generation using Report Lab for those who prefer a more permanent record. Each report contains the diagnosis, the Grad-CAM heatmap, dietary recommendations, and a clinical disclaimer. Unicorn and Docker are used to containerize the complete program for deployment, making it appropriate for production settings.

IV. RESULTS AND DISCUSSION

A. Individual CNN Model Performance

The test accuracy and AUC values for each model, as well as the overall performance of the ensemble, are compiled in Table IV. The Custom CNN's impressive accuracy of 90.7% demonstrates that a task-optimized, lightweight architecture can compete with considerably bigger pre-trained models. This was also matched with 90.7% accuracy by VGG16, which did especially well on texture-based deficiency patterns where its fine-grained feature extraction was helpful. ResNet50's outstanding deep structural feature extraction skills let it finish slightly ahead at 90.9%. MobileNetV2 is the most effective choice for real-time applications, achieving 90.8% accuracy and providing the fastest inference time of all models, at about 35 milliseconds per image. EfficientNetB0 was the best model among the individual models, with an accuracy of 93.8% and an AUC of 0.938.

B. Ensemble Model Performance and Confusion Matrix Analysis

Every time, the max-voting ensemble outperforms any individual model in terms of robustness. Interestingly, the ensemble outperforms EfficientNetB0 with a significantly higher AUC of 0.951, indicating stronger discriminative ability across all five deficiency classes—not just raw accuracy, but also how well it separates each class from the others—even though both achieve the same overall accuracy of 93.8%.

The per-class correct prediction numbers are 92 out of 100 for Normal, 87 out of 100 for Vitamin A, 91 out of 100 for Vitamin B, 90 out of 100 for Vitamin C, and 92 out of 100 for Vitamin D, according to the confusion matrix. Here, too, one of the ensemble approach's more obvious advantages is evident: It significantly lessens cross-class misunderstanding, particularly between classes like vitamin B and vitamin D that sometimes have similar visual characteristics.

C. Precision, Recall, and F1-Score

The precision, recall, and F1-scores for the ensemble model are broken down by class in Table VI. For every deficiency class, the model maintains a strong balance between precision and recall, as evidenced by the fact that all F1-scores are over 0.92, rather than outperforming one parameter at the expense of the other. When comparing the individual classes, Vitamin C and Normal have the greatest F1-scores at 0.94, while Vitamin A and Vitamin B have slightly lower scores at 0.92. This is probably because

Table vii. Comparison with existing

Study	Method	Accuracy (%)	Grad-CAM	Real-time	Vit. Classes	Limitations	Ref.
Suprith a et al. (2024)	VGG16 + Custom CNN	88.4	No	No	4	No XAI	[6]
Bonde et al. (2024)	Deep CNN (VGG+ ResNet)	87.2	No	No	3	High compute	[4]
Jayaram et al. (2024)	DenseNet	89.1	No	No	4	No real-time	[5]
Wang et al. (2023)	3-model Ensemble	91.3	No	No	4	No XAI, costly	[7]
Eldeen et al. (2020)	K-means+ ANN+Fuzzy	80.3	No	No	4	Low accuracy	[3]
PROPOSED SYSTEM	5-CNN Ensemble + Grad-CAM	93.8	Yes	Yes	5	—	—

Specific body region photographs have visual similarities that make it more difficult to differentiate between these two classes. Overall, the weighted and macro averages both reach 0.930, indicating that the ensemble's performance is both robust and reliable throughout all classes

D. ROC-AUC Analysis

The ROC analysis further validates the ensemble's higher discriminative capacity using a one-vs-rest approach. The ensemble raises the AUC to 0.951, a significant improvement even over the top-performing individual model, EfficientNetB0, which scored 0.938 on its own. Individual models' AUC ranges from 0.907 to 0.938. In comparison to earlier efforts, this improvement is also not merely incremental. The ensemble performs 4.2% more accurately than the best previously documented ensemble method from Wang et al. [7]. This finding lends credence to the theory that increased architectural diversity, when paired with the two-phase fine-tuning protocol employed here, actually improves overall performance rather than merely adding complexity for marginal gains.

E. Grad-CAM Qualitative Validation

The Grad-CAM heatmap analysis demonstrates that the ensemble's predictions are based on qualities that make clinical sense in addition to being statistically accurate. For instance, heatmaps of vitamin. A similar pattern can be seen in vitamin B insufficiency identified from nail photos, where heatmaps concentrate on the surface of the nail plate and identify ridging and koilonychia patterns that are typically linked to folate or B12 deficiencies. The model focuses on surface inflammation and vascular abnormalities, which are characteristics closely associated with scurvy-related glossitis, for vitamin C deficiency detected from tongue pictures.

When combined, these findings are comforting because they imply that the model is actually concentrating on areas of each image that are clinically relevant, as opposed to identifying spurious correlations or irrelevant background artifacts that just so occurred to correlate with the labels during training.

F. Comparison with Prior State-of-the-Art

The suggested system is directly compared to six previous approaches in Table VII. Out of all the methods tested, the 5-CNN ensemble with Grad-CAM achieves the greatest reported accuracy of 93.8%. Beyond its raw accuracy, it stands out due to its combination of capabilities. Among the systems compared, it is the only one that covers all five vitamin classes, supports real-time camera-based analysis, and incorporates Grad-CAM explainability as a standard, built-in output rather than an optional add-on.

V. CONCLUSION

A comprehensive, non-invasive vitamin deficiency screening system that combines real-time Grad-CAM explainability, a heterogeneous ensemble of five CNN architectures, multi-modal dermatological imaging, and an end-to-end Flask web application has been reported in this study. The system's efficiency is supported by the experimental findings, which show that the ensemble outperformed all previously published methods in the literature with an overall test accuracy of 93.8% and a ROC-AUC of 0.951.

With heatmaps regularly matching established clinical signs across all five deficiency classes, the Grad-CAM qualitative study provides strong evidence that the system's predictions are based on medically legitimate visual features. The quality alarm system, the live prediction panel, the specialized recommendation engine, and the supporting infrastructure all work together to show something that goes much beyond a standard academic prototype. By Considering all things, this method provides an affordable, easily accessible, and non-invasive substitute for standard blood-based diagnostic procedures—a particularly significant benefit in areas with low resources where traditional testing is frequently impractical. The two-phase transfer learning approach ensures that the system adapts robustly to this particular medical imaging domain, while the ensemble's combination of five different CNN designs minimizes the variation and constraints associated with depending on any one model. Finally, this work is a significant advancement in AI-based healthcare diagnostics that establishes a strong basis for next clinical pilot projects and, ultimately, technology transfer to actual healthcare professionals.

VI. FUTURE SCOPE

In terms of future development, a number of attractive avenues have been identified. Clinical validation using actual patients, via a prospective research that contrasts the system's output with verified laboratory diagnoses over a sample of 500 or more individuals, would be a crucial next step. Formal sensitivity, specificity, and positive predictive value—metrics that are crucial for any future regulatory consideration—would be established with the aid of this type of research.

Working with dermatological departments to create a clinically annotated, demographically varied dataset that is gathered utilizing standardized imaging techniques is another crucial path. This kind of dataset expansion would strengthen the system's resilience across various skin tones, age groups, and geographic populations, making it more dependable and egalitarian when used in a variety of real-world contexts.

The system may automatically pinpoint target body regions by integrating YOLO-based object identification, eliminating the requirement for manual selection and improving user experience. Additionally, mobile deployment would be made possible by converting the ensemble models to TensorFlow Lite (TFLite) format, which would allow offline screening in remote locations with spotty or limited internet connectivity.

In terms of modelling, expanding the architecture to accommodate multi-label classification would enable the system to identify several co-occurring faults at once instead of presuming a single dominating condition. Incorporating Monte Carlo Dropout for improved uncertainty quantification, adding a Vision Transformer (ViT) as the sixth member of the ensemble, integrating with electronic health records via HL7 FHIR standards, and investigating federated learning to enable privacy-preserving model improvements across distributed hospital systems are additional noteworthy additions.

REFERENCES

1. G. Sambasivam, J. Amudhavel, G. Sathya, "A Predictive Performance Analysis of Vitamin D Deficiency Severity Using Machine Learning Methods," IEEE Access, vol. 8, pp. 9492–9507, 2020.
2. E. K. and S. K., "Diagnosis of Vitamin Deficiency in Human Beings using DNN Algorithm," Proc. 2nd Int. Conf. on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2023, pp. 1627–1632.
3. S. Eldeen, M. AitGacem, S. Alghlayini, W. Shehieb, M. Mir, "Vitamin Deficiency Detection Using Image Processing and Neural Network," ASET 2020 Advances in Science and Engineering Technology Int. Conferences, Dubai, UAE, pp. 1–5.

4. S. Bonde, P. Argade, D. Gohil, R. Gagare, G. Wakade, "Vitamin Deficiency Detection using Image Processing and Deep-CNN Algorithm," Int. Research Journal of Modernization in Engineering Technology and Science, vol. 6, no. 5, pp. 4521–4529, May 2024.
5. M. Jayaram et al., "Deep Learning Technique Dense-Net Application to Predict Vitamin Deficiency," Industrial Engineering Journal, vol. 53, no. 5, pp. 118–129, May 2024.
6. Supritha M., Theeksha S., Asha K.H., "A Comprehensive Approach to Vitamin Deficiency Detection through Image Analysis of Skin, Tongue, Eyes and Nail Images using CNNs," IJARSCT, pp. 709–715, 2024.
7. L. Wang, D. Chen, E. Rodriguez, "An Ensemble Approach for Improved Vitamin Deficiency Detection in Clinical Images," IEEE Access, vol. 8, pp. 14496–14505, 2023.
8. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR, 2015. arXiv:1409.1556.
9. K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," Proc. IEEE CVPR, pp. 770–778, 2016.
10. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Proc. IEEE/CVPR, pp. 4510–4520, 2018.
11. M. Tan, Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Proc. 36th ICML, vol. 97, pp. 6105–6114, 2019.
12. R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," Proc. IEEE ICCV, pp. 618–626, 2017.
13. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017.
14. R. Moholkar, M. Kamble, G. Bobade, S. Shinde, "Vitamin Deficiency Detection Using Image Processing and Neural Network," IJARIIE, vol. 9, no. 2, pp. 1241–1248, 2023.
15. https://www.who.int/news-room/factsheets/detail/malnutrition.
16. S. M. Williams, J. R. Davis, "Multi-modal Deep Learning for Vitamin Deficiency Detection: Integrating Clinical and Imaging Data," Journal of Medical Imaging and Health Informatics, vol. 11, no. 8, pp. 2109–2118, 2021.
17. J. N. Lee, B. T. Wong, "VitaScan: A Mobile Application for Vitamin Deficiency Screening using Smartphone Images," IEEE Access, vol. 10, pp. 88921–88932, 2022.
18. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," Proc. ICLR 2021. arXiv:2010.11929.
19. T. Chen, X. Xu, C. Ding et al., "LGNet: Lesion Gate Network for Multiple Soft Tissue Diseases Diagnosis," arXiv preprint arXiv:2309.01728, 2023.
20. National Family Health Survey-5 (NFHS-5), Ministry of Health and Family Welfare, Government of India, 2019–21. Available: http://rchiips.org/nfhs/.