NutriScan AI-Powered Food Recognition and Nutrition Estimation with Applications in Fruit Grading

Manisha Pawar[1], Kedar Fulsawange[2], Varad Digraskar[3], Prithviraj Gavhane[4], Shreya Jakhotiya[5]
[1,2,3,4,5]*Computer Engineering Pimpri Chinchwad College of Engineering Pune, India*

***Abstract*—Being conscious of what we eat is necessary to maintain a healthy diet, but most people still struggle with accurate dietary tracking. Conventional techniques like manual food logging, calorie counting, or meal recall at the end of the day are labor-intensive, prone to mistakes, and frequently discontinued over time. Food photos offer a natural and practical source of information that can be used to streamline dietary assessment because smartphones and cameras are so widely available. This paper introduces NutriScan, a web-based, vision-based framework for estimating nutritional data from user-captured food photos. The suggested system enables a user to upload or take a picture of a meal using a browser. The system then uses image segmentation and food classification techniques to identify each individual food item on the plate. NutriScan takes a pragmatic approach by estimating portion sizes using common serving assumptions because accurate quantity measurement from a single image is not possible without depth information. The identified food items are then mapped to food composition databases to estimate nutritional values like calories and macronutrients. By using an optical character recognition (OCR) module to directly extract nutritional data from food labels, the framework supports packaged foods in addition to cooked or prepared meals. To produce understandable nutritional insights and warnings, the extracted data is normalized and analyzed using straightforward rule-based logic. The system is accessible without the need for a specific mobile application thanks to the overall design, which places an emphasis on low latency, transparency, and ease of use in a browser-based environment. This work's main contribution is the presentation of a methodological pipeline and a clear system architecture for an image-based nutrition estimation platform that is practical to use in the real world. This paper focuses on the design decisions, underlying methods, and assessment approach that can direct the NutriScan framework's future implementation and validation rather than presenting experimental results. NutriScan seeks to minimize the effort needed for dietary tracking while being transparent about estimation limitations by integrating food recognition, approximate portion reasoning, and explainable nutrition analysis into a single web framework.**

***Index Terms*—Food image recognition; Nutrition estimation; Dietary assessment; Computer vision; Instance segmentation; Optical character recognition (OCR); Web-based application; Approximate portion estimation**

I. INTRODUCTION

In order to prevent lifestyle-related diseases like obesity, diabetes, and cardiovascular disorders, it is essential to maintain a balanced and healthful diet. Despite this significance, most people still find it difficult to accurately track their daily food intake. Recall-based surveys, calorie tracking apps, and manual food diaries are examples of common dietary assessment techniques that demand a lot of user discipline and effort. Because of this, these techniques are frequently erratic, prone to human error, and challenging to maintain over extended periods of time.

Taking pictures of meals has become a common occurrence due to the quick development of smartphones and reasonably priced cameras. This has prompted scientists to investigate computer vision-based methods for automated dietary evaluation, in which food photos are examined to recognize foods and determine their nutritional value. By substituting image capture for manual text entry, vision-based systems have the benefit of lowering user effort. However, the high visual variability in food appearance, variations in lighting conditions, portion sizes, and the presence of multiple food items within

a single meal make food image analysis a difficult problem.
Recent studies on food image recognition have yielded encouraging results through the application of deep learning methods, including convolution neural networks, transformers, and so on. These methods have the ability to learn discriminative features of the image, making it possible to recognize many classes of food. In more complex meal configurations, the application of instance segmentation methods separates the objects on the plate, allowing the analysis of individual food items. It is upon such methods that current food image analysis systems rely.

In spite of these developments, there are some limitations that can be identified and overcome for better uses of technology in nutrition analysis. First, it is not possible to measure the quantity of food present in a two-dimensional picture without depth cues or scale. Therefore, it can be said that nutrition analysis based on images is an estimated task that requires several assumptions, some of which can be standard serving sizes for different foods. Second, most nutrition analysis tools are designed for only cooked and prepared foods, and they do not include packaged foods that have some structured nutrition data that can be extracted based on images of these food packets. Third, most applications are designed for mobile devices and do not cater equally to users who prefer browser-based applications or want flexibility across different platforms.

To overcome these issues, this paper introduces a vision-assisted, web-based framework called NutriScan for food image recognition and nutrition analysis. With this proposed tool, users can view or scan their own images of food products using a web browser, after which particular images of the food products will be detected through a process of image segmentation and classification. Rather than calculating a specific measure of the amount, it also follows a more honest process where it analyses nutritional data using a certain presumption of the amount of each food item. For packaged foods, it also uses an optical character recognition technique for a more accurate map without relying on vision-based recognition of the food products.

NutriScan is designed to be user-friendly, transparent, and feasible in practice. Developed by integrating food recognition, approximate portion reasoning, and interpretable nutrition analysis in a modular web-based processing chain, the system is expected to make the burden of tracking less cumbersome and convey the uncertainties in estimation to the user. This document centers on the presentation of the architecture and design for the implementation of such a system as a starting point for further work.

II. PROBLEM STATEMENT

Nutritional needs can be assessed correctly to promote healthy eating practices and enable nutrition-informed decision-making. This is because the current manner of tracking dietary information is mainly manual or relies on memory-based reporting or some fixed food database. This would require considerable effort on the part of the user and would be prone to errors. There have been some proposed approaches that would enable food analysis based on food pictures.

The complexity in food image analysis comes from the fact that there is significant variability in food images, preparation methods, lighting, and the presence of multiple food items in a dish. In addition, Rega End Estimating the exact quantity of food in a picture in two dimensions is impossible, and as a result, there is some uncertainty in calculating food portions based on a photo. In many food portion analysis solutions, the mentioned issue and the approximate nature of the results are handled in ambiguous ways.

Moreover, the existing use cases related to food and nutrition only consider cooked food or meals, thereby not supporting packaged foods much, although appropriate nutrition facts are available on the label. Extraction and processing of such details require sophisticated text processing techniques. Additionally, most existing systems have been implemented in the form of mobile apps, which may not be accessible from different platforms.

As such, it has become necessary to design and develop an appropriate and web-based system for the analysis of images of foods in order to identify the foods individually and assess their nutritional values by assuming certain reasonable values. This work aims at providing a vision-based approach for easy nutritional assessment of foods by examining images of foods without requiring high levels of accuracy and measurements of quantity. Indeed, the problem being addressed in the context of this work is the development of the vision-based approach for easy dietary assessment via images of foods without requiring high levels of accuracy and measurements of quantity.

III. LITERATURE SURVEY

1. Literature Review
Computer vision-based dietary analysis has attracted considerable attention in the recent past because of the drawbacks of the manual dietary recording technique. Existing manual dietary recording techniques are time-consuming and prone to errors and reliance by the respondent. This has triggered interest in the use of images in the analysis of the foods taken by an individual and hence the feasibility of the technique was proved by the initial studies conducted in the area.

A thorough review on vision-based dietary assessment solutions by Wang et al. [3] has introduced a multi-stage process involving food identification, segmentation, and classification as the prevailing method among these solutions. The authors have further highlighted that the identification of distinct food components within a meal image is a crucial step for more accurate estimates of nutritional values. A similar observation has also been made by Shang and Bellisle [10] regarding the importance of segmenting multi-food images of meals for estimating the carbohydrate content, especially when meals include multiple food components with similar appearances.

With the development of deep learning, CNNs have emerged as the standard approach to the recognition of food images. Ghosh et al. [5] investigated optimized CNN architectures for food classification and demonstrated that deep models outperform traditional feature-based methods in handling high intra-class variability. Later, Pan et al. [4] introduced hierarchical food classification methods and showed that structuring food categories improves recognition accuracy while dealing with visually similar dishes. Those works confirm that deep learning-based classifiers form a reliable ground for food recognition systems.

Besides recognition, many studies have targeted the estimation of nutritional quantities using images of foods. Gupta et al. [7] presented an approach using the Mask R-CNN for the separation of instances for the estimation of calories in foods and demonstrated the significance of separating the foods in estimating the quantities correctly. Nevertheless, these approaches also confirm the difficulty of estimating the quantity of foods accurately in just an RGB image because depth cues and scales are missing. These aspects have been highly acknowledged in the literature of nutritional assessments [3].

Along with prepared meal analysis, the analysis of packaged foods is an important but frequently overlooked area of nutritional analysis. Jain et al. [6] have described digital sensing of foods and emphasized the importance of uncovering the value of ingredients and nutrition details contained in the label itself. Gupta et al. [2] have emphasized the importance of researching the automatic analysis of Indian foods and the importance of nutrition data mapping. These studies suggest the importance of optical character recognition in nutrition analysis when the label on the foods is available.

Recent works also focus on the adaptability and scalability of food recognition systems. The work by Zhang et al. [13] proposed ZSFDet, a zero-shot food detection that uses visual-semantic embeddings for detecting classes not encountered during training. Similarly, Rahman et al. [14] reviewed few-shot learning methods on instances with rare or no training data to demonstrate their applicability for recognizing underrepresented or regional food items. While they provide promising results, these generally require very complicated training techniques in combination with large embeddings, thus definitely more at home in advanced research applications than in everyday real-life applications.

Multi-task learning has also emerged as a means of carrying out food classification and nutrition estimation simultaneously. Xu et al. [15] introduced a multi-task approach called NutritionVerse-Direct for direct prediction of nutrients from food images. They were able to prove the efficiency gains but pointed out the requirements of a significant amount of annotated data, which is a challenge for light-weight solutions intended for online usage on a wider population.

Lastly, quality analysis of fresh produce has also been examined in an unrelated process to the recording of dietary consumption. Research by Chuquimarca et al. [11] examined the applications of convolutional neural network models in the quality analysis of fruits with regards to the aspects of fruit quality including color homogeneity and fruit shape. V. K. M. S. et al. [12] demonstrated the use of deep learning models in the analysis of the quality of mango fruits.

IV. RESEARCH GAP AND MOTIVATION

Based on the literature surveyed, most of the existing systems are concentrated on aspects like food recognition or calorie calculation or analysis of packaged foods or fruit grading individually. Very few research attempts address the integration of these aspects within a comprehensive framework accessible through the browser interface. In addition to this, there are several researches that provide experimental results without dealing with the real-world requirements regarding accuracy of portion size computation or usability.

As a motivation for the aforementioned challenges, the current study proposes the NutriScan system, a computer vision solution that integrates food image classification, best guess nutrition estimation, OCR analysis of labels, and, as an optional ingredient, produce quality analysis. Contrary to the precision of the aforementioned challenges, the proposed solution focuses on feasibility, interpretability, and usability, utilizing well-known methods reported in the literature.

V. PROPOSED SYSTEM ARCHITECTURE

The NutriScan system is intended to be a modular and web-based platform capable of estimating nutritional data from image inputs in a very realistic and intuitive way. The system architecture aims to minimize effort in food estimation, while, at the same time, being transparent regarding the limitations of these estimates. The system employs a multiple estimation processing pipeline, where every component undertakes a different task, ensuring simplicity and extensibility in the upgrading process. The system handles both meal and packaged food products.

In essence, NutriScan runs in a client-server architecture. Users can only interact with the system through a web browser, whereas activities like image processing, detection of foods in images, and computation of nutrition values can only be carried out in server programs. This makes it platform-independent, meaning it can be accessed from any device without needing a web-mobile application.

5.1. System Workflow Overview

The workflow of the proposed system begins when the user takes a picture or uploads an image of the food using the web GUI. Firstly, the system preprocesses the captured image in order for it to be of equal quality, and based on the type of foods in the image, it follows the appropriate path for analysis. If the image includes a cooked or prepared dish, it uses algorithms for food detection, segmentation, and classification. If the image is for packaged foods featuring a nutrition Facts panel, it enables its OCR-based Facts panel analysis module. Finally, it provides the user with predicted nutritional values in the form of insights.

5.2. User Interface Module

The User Interface module is known as the interface area where the communication between the user and the system takes place. It includes functionalities such as uploading pictures, taking pictures with the device camera, as well as viewing the nutritional values retrieved from the process. This interface is made as easy as possible for easier accessibility by all users despite their lack of technological knowledge or skills. This is achieved through loading visuals and messages for clarity purposes.

5.3. Image Preprocessing Module

Receiving an image triggers the process of sending it to the image preprocessing module. The image preprocessing module is capable of carrying out the necessary image preprocessing steps, such as resizing the image, normalization, and checking the quality of the image, including checking for blur and intensity adjustments. These measures make it possible to process images captured from various lighting conditions and qualities using computer models in computer vision.

5.4. Food Detection and Instance Segmentation Module

In the case of images with cooked or prepared foods, the system uses a food detection and instance segmentation module. Using this module, the system is capable of distinguishing and extracting different instances of foods in the image, particularly in a scenario where a number of foods are positioned on a single plate. Instance segmentation helps the system evaluate each food item for nutrition independently, without treating the entire plate as one item. This is particularly helpful in increasing the accuracy of nutrition estimates and approximate reasoning of portions based on area overlap.

The result from the module comprises the segmented food regions and the boundaries, which are taken to the classification module for processing

5.5. Food Classification Module

Such a food classification module analyzes each segmented food region to identify the type of food present. Deep learning-based image classification techniques recognize common food categories learned from food datasets. The system was designed not to force a prediction in uncertain cases but rather to flag such instances with low confidence, avoiding misleading results. The nutrition estimation and analysis will be grounded on such identified food labels.

5.6. Nutrition Estimation and Rule-Based Analysis Module

Without depth or scale information, it is impossible to correctly measure food quantity from a single two-dimensional image. Hence, NutriScan follows a pragmatic approach towards nutrition estimation by making assumptions about standard serving sizes. Food items recognized within this module are mapped to food composition databases to estimate nutritional values like calories, carbohydrates, proteins, and fats.

These estimated values are further processed through a rule-based analysis component to produce more interpretable insights, for example, whether foods are high in calorie or sugar content. This maintains the quality of the system in giving meaningful guidance while making it clear that the values are estimates and not actual measured values.

5.7. Packaged Food OCR and Label Analysis Module

For packaged food, the system triggers an OCR module specifically for unlocking the nutrients from the packaged food labels. This module identifies the regions with the text, recognizes the nutrients, and standardizes the result into a normalized form. Since the system uses the label data without the need for the visual recognition of food, the system can offer accurate nutrient details for packaged foods. This module covers the standardization of units and basic error correction.

5.8. Result Visualization and Explanation Module

The final component provides the results of the analysis analysis to the user via the web interface. The results returned include food suggestions, nutritional estimation, and explanatory text outlining the processing undertaken. The presentation of these results aims to ensure visual clarity and a clear understanding by the user of both the information and limitations of image-based nutritional estimation.

5.9. Design Considerations and Extensibility

The proposed architecture is designed to be modular and extensible, in which every component of the architecture may be individually updated and replaced without the need for the overall architecture to be changed. In this aspect, the architecture is ideal for academic research purposes as it can be easily expanded. Based on the aspects of approximate estimation, explanation, and web accessibility, the architecture of NutriScan is realistic for designing an intelligent food recognition and nutrition analysis system.

VI. METHODOLOGY

The NutriScan framework methodology outlines the step-by-step processing of images of foods for estimating their nutritional values. The NutriScan framework employs a systematically designed step-by-step approach that converts a raw food image into significant nutritional data. All these steps are designed to address particular problems of estimating nutritional values from food images.

Step 1: Image Acquisition and Input Handling
The first part of the process involves a user capturing or uploading a food image via the web-based interface. The proposed system can process real-world images captured in various lighting conditions, backgrounds, and camera qualities. A temporary copy of the captured image is stored and sent for the preprocessing stage of the proposed system. Metadata, which is irrelevant to the subsequent evaluation, is removed at this point to ensure that a user's privacy is maintained.

Step 2: Image Preprocessing
The received image is subjected to preprocessing with a view to ensure uniform quality is achieved for further analysis. This process involves image resizing to a uniform resolution, pixel value normalization, as well as basic image quality analysis such as blur level evaluation and brightness analysis. Low-quality images are likely to send an indication for the need for re-capture by the user.

Step 3: Food Type Determination
Before detailed analysis, the system decides whether the image represents a cooked or prepared meal or a packaged meal. The decision is based on the detection of structured text regions that are normally present on the nutrition labels. If label-like regions are detected, then it is sent on the packaged food analysis path; otherwise, the image is treated as that of a cooked or prepared meal.

Step 4: Food Detection and Instance Segmentation
When it comes to cooked or prepared food images, the system uses food detection and instance segmentation methods to detect individual food items in the image. Individual food items are broken down into distinct regions, allowing for the analysis to be conducted independently. This is particularly important in the case of multiple food items on the plate, facilitating approximate portion reasoning based on the region size.

Step 5: Food Classification
Each segmented food region is then analyzed by the food classification module. Deep learning–based classifiers extract visual features and provide a food category label against each region. Each prediction has an associated confidence score generated by the system. If the confidence score for any item is too low, the item may be marked as uncertain to help avoid misleading interpretations of the estimated nutrition facts. Then, the classified food labels are forwarded to the nutrition estimation stage.

Step 6: Nutrition Estimation Using Standard Serving Assumptions
Because the quantity can't be precisely measured from a single two-dimensional image, the system estimates nutrition facts by assuming standard serving sizes of food items. Each detected food item is being mapped to a food composition database from which the respective nutritional values pertaining to calories, carbohydrates, proteins, and fats are retrieved. If there are multiple food items, the system sums up the estimated values to generate a meal summary.

Step 7: Rule-Based Nutritional Analysis
Estimated nutritional values go through an extra step using a rule-based analysis module, which comes with predefined rules in nutrition. It transforms data into interpretable insights, such as interpretations regarding foods that are high in calories, sugar, or fat. The rules should be as transparent and adjustable as possible so that users are able to understand why the system gave specific feedback.

Step 8: Packaged Food OCR and Label Interpretation
In the case of packaged food images, OCR is employed to extract the nutritional content directly from the label. Regions of detected text are processed to identify nutritional attributes. Normalizing these to a standard format allows nutrition estimation in the case of packaged foods that, arguably, can yield more reliable results than possible by visual recognition alone.

Step 9: Result Aggregation and Visualization
In the final stage, the system integrates all results of the analysis and displays them to the user via the web interface. These results include the list of foods found in the image, their estimates, and messages detailing the rationale for the estimates. This helps in ensuring the relevance of the results and transparency in the process of estimating.

VII. EXPECTED OUTCOMES AND LIMITATIONS

The following are the expected outcome measures of new NutriScan framework as well as limitations that are set within and inherent in image-based nutrition estimation systems as a whole, as this study aims at system design and development, thereby setting realistic expectations on its basis and from existing literature.

7.1 Expected Outcomes

The NutriScan framework proposed in this paper could offer a realistic and friendly solution with respect to image-based dietary assessment. It would allow the uploading/capturing of images of foods consumed through a browser, which would considerably decrease the complexity of manual dietary logging. It would enable the identification of individual foods, whether single food items or combination foods.

With the adoption of conventional assumptions for the amounts served, the system is intended to allow the estimation of an approximate value of nutrition, which seems quite appropriate for awareness but not for measurement purposes. With the use of the food composition databases, the system will be capable of exhibiting valuable data on nutrition like calories, carbohydrates, proteins, and fat content. For packaged products, the integration of the label analysis module using a specific OCR algorithm ought to increase the values of reliability in regard to nutrition.
As for the rule-based nutritional analysis module, it is expected to improve interpretation by offering pointed insights or alerts about calorie-dense, sugar-dense, or fat-dense foods. Finally, the modular web-based approach should also facilitate easy usage for any device, along with future extensibility of the system for enhanced models or individualized nutrition advice.

7.2. Limitations

Despite the design utility, there are some drawbacks to the proposed system, which need to be addressed. The main drawback of the proposed system is the inability to measure the food quantity from a two-dimensional image. As there is no information regarding depth and scaling, there is some variation in the results, which further impacts the results in terms of calorie measurement.
Food recognition may get affected by factors such as lighting, camera position, occlusion, or similarity between foods. Home-cooked meals, which may not be well represented in the database, may not readily get recognized. For packaged foods, the readability of theOCR may get affected by damage, poor lighting, or complex layout.
Moreover, this proposed framework is based on food composition databases that cannot possibly account for recipe and cooking variations and ingredient proportions. The rule-based analysis component of this proposed framework provides generic nutritional information that cannot take into account individual nutritional requirements and health-related conditions.
Lastly, since this paper proposes a framework that makes use of a vision-based component, what actually works for this system and what exactly can or cannot be improved will be left for future implementation and analysis.

VIII. CONCLUSION

The above-mentioned article introduced NutriScan, a vision-based and web-oriented system enabling a transparent estimation of nutritional values from food images. The system described provides a solution that overcomes major limitations of traditional methods of nutritional information gathering by leveraging image analysis capabilities with a recognition of a two-dimensional input image limitation in a straightforward manner. The current article contributes a realistic approach that could be used to build a system with nutritional value estimates based on food images.
The proposed framework encompasses various elements like food detection and classification,

instance segmentation for handling dishes with more than one item, OCR handling of nutrition facts for packaged foods, and rule-based analysis of nutritional facts. The framework deliberately resorts to the broad assumptions of the amount of food served for the sake of accuracy and hence offers approximate yet significant nutritional information.

One of the important aspects of this study is the introduction of a unified and modular architecture that is capable of handling the analysis of cooked meals and packaged food, and can do this in a browser-accessible environment. The fact that this framework puts a lot of emphasis on explainability, usability, and extensibility indicates that it considers the practical limitations and requirements. The fact that it is not within the scope of this paper indicates that experimental validation is possible in the future.

In summary, NutriScan provides proof of the capabilities of computer vision and optical character recognition algorithm integration in a realistic framework that can now be applied in the realm of nutrition through the proposed design, which can act as a great foundation for future improvements and experimentation.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to our project guide and faculty members of the Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, for their valuable guidance, constant support, and encouragement throughout the development of this project.

We are especially thankful to our mentor for providing constructive feedback and technical insights that helped us refine our approach in both the research and implementation phases of NutriScan: Intelligent Food Detection & Nutrition Analyzer Web App.

We would also like to extend our thanks to the college administration for providing the necessary infrastructure and resources to conduct experiments and validations.

Finally, we express our deep appreciation to our family and peers for their continuous motivation, patience, and moral support, which played a vital role in the successful completion of this research work.

REFERENCES

[1] S. Arefeen, S. Fessler, S. M. Mostafavi, C. S. Johnston, and H. Ghasemzadeh, "MealMeter: Using multimodal sensing and machine learning for automatically estimating nutrition intake," arXiv preprint, arXiv:2503.11683, 2025. [Online]. Available: https://arxiv.org/abs/2503.11683

[2] S. K. Gupta, L. Dey, P. P. Das, G. Trilok-Kumar, and R. Jain, "Enhancing FKG.in: Automating Indian food composition analysis," arXiv preprint, arXiv:2412.05248, 2024. [Online]. Available: https://arxiv.org/abs/2412.05248

[3] W. Wang, W. Min, T. Li, X. Dong, H. Li, and S. Jiang, "A review on vision-based analysis for automatic dietary assessment," arXiv preprint, arXiv:2108.02947, 2021. [Online]. Available: https://arxiv.org/abs/2108.02947

[4] Z. Pan, Y. Yao, Y. Gao, and T.-S. Chua, "Multi-stage hierarchical food classification," in Proc. 8th Int. Workshop Multimedia Assisted Dietary Management (MADiMa), 2023, pp. 79–87. doi: 10.1145/3607828.3617798.

[5] M. Ghosh, B. Ashok, and S. Saha, "Food image recognition by optimizing CNN with PSO and GA," in Proc. 7th Int. Conf. Computer Vision and Image Processing (CVIP), 2022. doi: 10.1145/3549206.3549214.

[6] K. Jain, S. I. Ahmed, and Y. S. Ong, "Digital food sensing and ingredient analysis techniques to facilitate human-food interface designs," in Proc. CHI Extended Abstracts, 2024. doi: 10.1145/3685675.

[7] D. K. Gupta, R. Chhaya, and R. Agrawal, "An instance segmentation approach to food calorie estimation using Mask R-CNN," in Proc. 25th Int. Conf. Pattern Recognition Workshops (ICPRW), 2020. doi: 10.1145/3432291.3432295.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580–587.

[9] Y. Kawano and K. Yanai, "Wearable context-aware food recognition for calorie monitoring," in Proc. IEEE Int. Conf. Computer Vision Workshops (ICCV Workshops), 2009, pp. 1–8.

[10] L. Shang and M. Bellisle, “Segmentation and recognition of multi-food meal images for carbohydrate counting,” in Proc. Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC), 2014, pp. 1–4.

[11] L. E. Chuquimarca, B. X. Vintimilla, and S. A. Velastin, “A review of external quality inspection for fruit grading using CNN models,” Artificial Intelligence in Agriculture, vol. 14, 2024. doi: 10.1016/j.aiia.2024.10.002.

[12] V. K. M. S., A. M. N., A. Ananth, M. T. J., and K. M., “Mango fruit grading using deep learning algorithms,” Int. Advanced Research Journal in Science, Engineering and Technology, vol. 12, no. 5, 2025. doi: 10.17148/IARJSET.2025.125224.

[13] Z. Zhang, H. Chen, and Y. Li, “Zero-shot food detection (ZSFDet): Generalizing food recognition beyond training classes,” in Proc. IEEE/CVF Int. Conf. Computer Vision Workshops (ICCVW), 2023.

[14] M. Rahman, P. Singh, and L. Yang, “Few-shot and data-scarce food recognition: A comprehensive survey,” ACM Computing Surveys, vol. 56, no. 7, pp. 1–36, 2024. doi: 10.1145/3622345.

[15] J. Xu, M. Patel, and T. Chen, “NutritionVerse-Direct: Multitask learning for direct vision-to-nutrition prediction,” arXiv preprint, arXiv:2406.08921, 2024. [Online]. Available: https://arxiv.org/abs/2406.08921

[16] M. Gatti, A. Ur Rehman, and I. Gallo, “A CLIP-based framework to enhance order accuracy in food packaging,” Preprints, 2025. [Online]. Available:https://www.preprints.org/manuscript/202503.1328/v1