

Food Recognition and Nutrition Analysis by Resnet-50

Anshul Diwakar¹, Rashi Agarwal²

¹PG - Computer Science and Engineering, HBTU, Kanpur, Uttar Pradesh, India.

²Associate Professor, Computer Science and Engineering, HBTU, Kanpur, Uttar Pradesh, India.

Emails: anshuldk19@gmail.com¹, rashi@hbtu.ac.in²

Abstract

Diabetes is a serious problem that is of major concern today. Maintaining and monitoring the diet by keeping in check the nutritional values of food intake can help planning a healthy diet. Numerous Computer Vision techniques are encompassed for automatic Food Image Recognition followed by nutritional values estimation. Advancements in convolutional neural networks grounded on Deep Learning methods have resulted in greater accuracy in the image recognition. This study proposes a fine-tuned model trained on Resnet-50 achieving an accuracy of 81.59% for identifying food images on the Food-101 dataset entailing 101 food categories with 1000 images in each category. With an additional MLP for nutritional values estimation through a CSV file maintained beforehand. The fusion of Convolutional Neural Networks (CNN) and Multi-Layer Perceptron (MLP) approaches for analyzing nutritional values in food images involves combining the distinct strengths of these architectures to enhance the overall system's performance, particularly on complex image datasets like Food 101.

Keywords: CNN, Diabetes, MLP, Nutrition

1. Introduction

Food image analysis for recognition and nutritional values estimation is now vital in food science and technology due to advances in computer vision, machine learning, and deep learning. Food quality, safety, and nutritional assessments for diet monitoring fall within this realm. A nutritional diet lowers risk of severe problems like diabetes, malnutrition, obesity and more. Dietary evaluation systems need food image analysis also. Food products may be accurately classified from photos using modern deep learning methods, such as CNNs. This makes food picture categorisation more effective even with complicated and nonlinear datasets, benefiting health-conscious consumers and nutrition evaluations (Islam et al., 2018). Additionally, end-to-end food picture analysis systems include food localisation, categorisation, and portion size calculation. Localisation findings and food energy distribution maps are used to estimate meal quantities and nutritional content, which is essential for controlling and avoiding chronic illnesses (He et al., 2021). Due to the intricacy and visual variability of food photos, issues persist. After

tremendous advances in machine learning and deep learning, managing various datasets and increasing different multiple regional cuisine representation needs continual development (Gilal et al., 2023). Conventional imaging-based machine vision systems help evaluate fresh and packed food quality. These devices can detect exterior faults, color changes, and interior chemical-physical properties of food goods, helping firms manage product quality throughout the distribution chain (Palumbo et al., 2023). Food picture analysis improves categorisation and nutritional evaluation using computer vision and deep learning. Further integration of these technologies with real-time applications will be helpful for dietary management. Deep learning outperforms classical machine learning in food identification. This technique uses food imagery to enhance food category, ingredient, quality, and amount recognition (Zhang et al., 2023). The use of deep learning in food identification algorithms has made nutritional assessments more accurate and dependable than human memory (Salim et al., 2021). CNN is a key approach in deep learning for food identification.

CNNs thrive with grid-like data like photos, making them popular in image recognition. Automatically and adaptively learning spatial hierarchies of characteristics makes them ideal for complicated visual input tasks like food identification (Li, 2022). Transfer learning extends pre-trained models to new, related tasks, minimising the requirement for huge labelled datasets, which are generally a constraint in food identification tasks (Zhang et al., 2023). The considerable intra-class variance and low inter-class variation in food photos create major hurdles for deep learning models in food identification. This is because comparable foods are presented differently and categories change somewhat. Modern systems use data augmentation to increase training dataset variety by artificially manipulating data. Additionally, combining several classifiers trained on various deep learning models improves food identification system accuracy and resilience (Salim et al., 2021). Deep learning has enhanced food identification systems and expanded their commercial and research applications. They are being tested for creative purposes including food safety checks and nutritional evaluations, which might greatly benefit health areas (Zhang et al., 2023; Salim et al., 2021). Food image analysis, a distinct field of picture identification, uses Convolutional Neural Networks (CNNs) effectively. This job suits CNNs since they can handle extremely non-linear datasets like food categorisation datasets (Islam et al., 2018). These networks automatically learn hierarchical characteristics from photos, revealing subtle patterns that typical feature extraction methods ignore. Iteratively convolving input with learnt filters creates a hierarchy of feature maps that increase classification accuracy (Taye, 2023). CNNs are vital for food image analysis because they meet the rising requirement for dietary assessment systems. Food photos are diverse and complicated, making this application difficult. CNNs can handle complicated classification problems, as shown by a 92.86% accuracy in categorising a food dataset (Islam et al., 2018). CNNs are versatile enough to be used in various computer vision areas including object recognition, which helps improve food picture

analysis by detecting particular food items. Due to their resistance to distortions and translations, networks are efficient in many applications (Taye, 2023). CNNs need a lot of computing power to complete billions of multiply-accumulate (MAC) operations for feature extraction. Value prediction algorithms that utilise the spatial correlation of zero-valued activations in CNNs may decrease these operations, optimising efficiency without losing accuracy (Shomron & Weiser, 2019). [1-3]

2. Literature Review

Modern diets and health depend on nutritional analysis to better understand how eating affects health. Digital technology in dietary evaluation indicates a move towards more accurate and personalised nutrition advising. Digital quantitative volume estimate innovations may simplify food intake measurement, improving nutritional evaluations and public health concerns like obesity (Tay et al., 2020). Nutritional analysis also helps discover and treat nutritional deficiencies that might harm health. Deficits in vitamins and minerals may cause oral health issues, emphasising the necessity for full dietary evaluations to promote tooth health and avoid illnesses (Strączek et al., 2023). The development of contemporary merchants and changing food contexts emphasise the necessity of nutritional analysis in understanding dietary trends. These merchants may boost protein and micronutrient availability, but they may also increase ultra-processed food consumption, requiring regulatory rules to encourage healthy food choices (Khonje et al., 2020). Metabolomics and mass spectrometry have improved nutritional evaluation accuracy and specificity. Metabolomics, for instance, creates novel biomarkers to track food intake and its biological impacts to study complicated nutrition-health connections (Llorach et al., 2012). Mass spectrometry allows molecular study of nutrients and metabolites, which is vital for understanding dietary component bioavailability and health effects (Kusmann et al., 2007). Food nutritional value estimation is essential to understanding and supporting healthy eating. To make educated dietary decisions, nutritional value assessment estimates

food nutrients such vitamins, minerals, fats, proteins, and carbs. Nutritional value estimates help customers choose healthier foods. Tools like the Overall Nutritional Quality Index (ONQI) assess items by nutritional content to help customers follow dietary requirements (Katz et al., 2009). By detecting nutrients deficient or abundant, food composition databases help public health nutrition. This informs nutrition-related health policies and initiatives to promote public health (Ocké et al., 2021). Micronutrient ratios may help explain diet-health outcomes better than individual nutrients. For instance, contentious ratios like omega-6/omega-3 and sodium/potassium might affect metabolic health and help develop dietary treatments (Kelly et al., 2018). Estimating nutritional values shows food insecurity-related dietary deficits. Such understanding supports measures to increase resource availability and nutritional intake for disadvantaged groups, especially women who are more food insecure and at risk of health issues (Ma et al., 2021). For sustainable diets, nutritional value estimation is crucial. It promotes plant-based diets while recognising the nutritional value of animal-source foods by assessing diets' nutritional quality and environmental effect (Dave et al., 2021). Understanding food nutrition allows for tailored diets for newborns, toddlers, and older individuals. Estimating nutritional value helps discover dietary items that substantially contribute to these groups' nutrient consumption (Ma et al., 2021). Advanced algorithms can predict individual nutritional needs, making PN a highly individualized approach to health (Gami et al., 2024). However, adapting these technologies in real-world settings requires overcoming significant challenges, such as data privacy, ethical concerns, and ensuring equitable access to services across diverse populations (Donovan et al., 2025). Current frameworks in personalized nutrition also suggest the need for adaptive systems that can provide just-in-time advice tailored to individual needs within real-life food environments. (Bossard et al, 2014) introduced the Food-101 dataset to automate food image classification to detect objects of interest using

random forest showing a conventional method of food object detection with an accuracy rate of 50.76%. . This limited scope suggests the need for more inclusive studies to explore the potential of PN across various demographic and health profiles. Research on UEC-FOOD100 dataset has been done using fine-tuned Deep Convolutional Neural Network CaffeNet that is pre-trained using the ImageNet dataset with 2000 categories achieving an accuracy of 78.77% (Yanai et al., 2015). A mobile application platform was developed by (Pouladzadeh et al., 2017) using CNN for training to analyze food constituents of FoodDD Dataset achieving an accuracy of 94.11%, a precision rate of 93.05%, and average recall rate of 90.98%. In (Aktı et al., 2022) the Middle Eastern Cuisine dataset with 27 classes was used to train the MobileNetV2 model to achieve two goals: prediction of type of food and having low-latency predictions. This is easily flexible for mobile applications though being a shallow architecture and achieved high accuracy compared to other existing models. [4-7]

3. Research Methodology

This section focuses on transfer learning based CNN Resnet50 which is being trained on FOOD-101 dataset followed by Multi-Layer Perceptron for food image classification and nutritional values estimation respectively. [8-10]

3.1. Dataset and Image Preprocessing

Food-101 is a large dataset containing food images used for computer vision and machine learning problems relating to food image detection and classification. 101 food categories and 101,000 photos make up this collection with 1000 images of each food category. The original dataset is split into 75%, 25% for train and test respectively followed by splitting of 75% in a 9:1 ratio. Therefore 68175 images for training, 7575 images for validation and 25250 images for testing. The dataset is preprocessed to avoid noise, distortions and for more accurate detection of food images. The image is given to the preprocessing function i.e. preprocess_input to make it suitable for the intended task. The rotation range, width-shift range, height-shift range, sheer-range and zoom-range have been set to 20, 0.2, 0.2, 0.2, 0.2

respectively. [11-13]

3.2. Model Architecture and Building

Resnet50: Resnet-50 consists of convolutional layers that have several layers followed by batch normalization and ReLU activation to extract features, identity block, convolutional block for processing features and fully connected layers for final output. In this paper Resnet50 is introduced as it uses the residual blocks that improves model performance by extracting desirable features and feeding them. The proposed model Resnet50 is set up by Keras built in function, followed by an MLP head shown in Figure 1 containing the Global Average Pooling layer which takes output from Resnet50 feature selection as an input. Then a dropout layer dropping 50% of neurons is added. A Dense layer with 512 units with activation function 'relu' is added, again followed by a dropout layer dropping 30% of neurons. The output is taken by a Dense layer with 101 classes and activation function being set to 'softmax'. In feature extraction the model is trained with 13 epochs with all layers on freeze. In the next step, for fine-tuning the last 125 layers of feature extracted model are unfreeze, keeping other layers on freeze and trained for 30 epochs with various callbacks EarlyStopping, Model Checkpoint and ReduceLRPlateau. The model is compiled with Adam optimizer with learning rate of 1e-5 on accuracy metric and the results are stored in a CSV file. [14-17]

3.3. Multi-Layer Perceptron

Multilayer Perceptron (MLP), an artificial neural network, has been used with Resnet50 to estimate nutritional values. MLPs are known for their pattern recognition, modelling, and prediction abilities across fields. In this paper in Figure 1 MLP is integrated with Convolutional Neural Networks (CNNs) using visualisation approaches to understand neural network behaviour beyond performance measurements. The MLP is introduced with a Sequential layer with one's output to the next input. The input layer specifies the shape of the data, the number of features in input data. A dense layer with 256 neurons and activation function set to 'relu', takes the shape as input to learn complex patterns.

Then a dropout layer drops 30% of neurons to make modern less dependent on specific neurons followed by a Dense layer with 128 neurons. The output layer is Dense layer with 6 neurons each representing one nutritional value: calories, protein, carbs, fat, sugar, fiber respectively. In data preparation nutritional data is loaded from a CSV file by accessing food class name as dataframe index. Mapping of class indices using train generator is done with food class names. The MLP extracts features from training images using pre-trained feature extractor grounded on CNN for training to predict nutritional targets. Mapping of class index of each training image is done with food class names and then with its nutritional value. Features and nutrition targets are standardized using StandardScaler (). The model is compiled with Adam optimizer on metric mean absolute error and mean squared error loss. The predictions are made on sample test images. Figure 2 shows ResNet-50 with MLP [18-20]

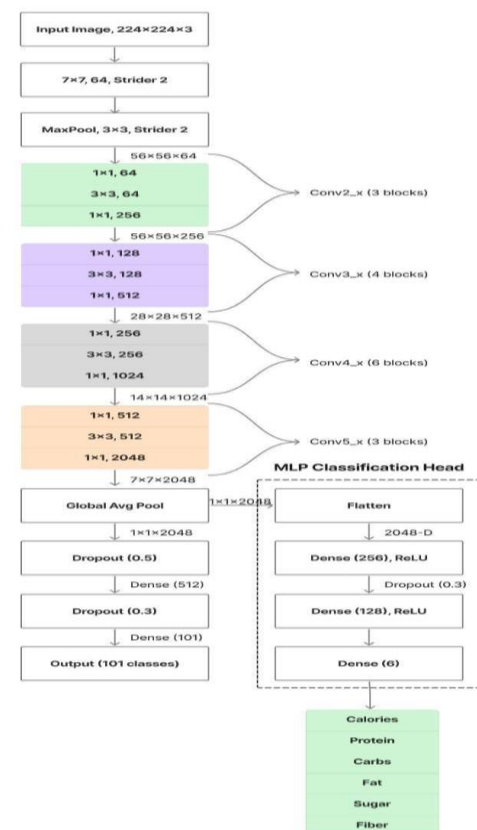


Figure 2 ResNet-50 with MLP

4. Results and Discussions

The model has been trained for feature extraction and respective training and validation accuracies has been plotted. The top-1 % validation accuracy has been obtained by testing on the trained accuracy of the model and the respective graph is plotted in Figure 2. During feature extraction it was observed the model achieves the 60.2% validation accuracy. Figure 2 shows Train Vs Val Accuracy of Model

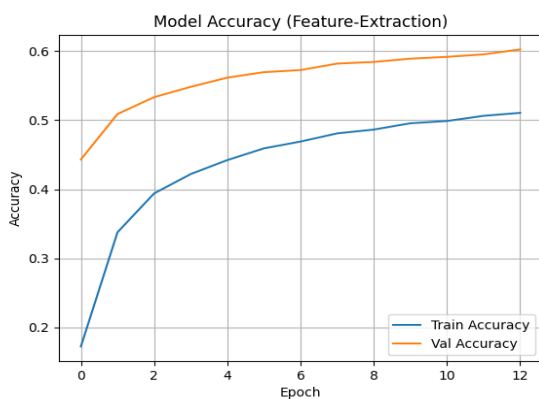


Figure 2 Train Vs Val Accuracy of Model

Correspondingly, the loss occurred during training and on validation has also been calculated successfully. Figure 3 represents the gradual decrease in the training and validation loss with the increase in the number of epochs. The loss initially was 2.35 and it moved to 1.60 for validation. This represents that model extracting salient features. Figure 3 shows Train Vs Val Loss of Model [21-23]

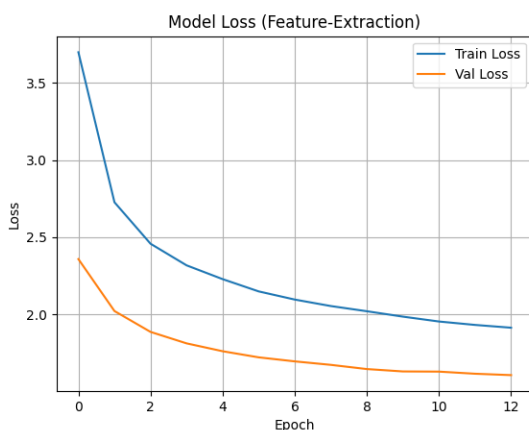


Figure 3 Train Vs Val Loss of Model

Now to enhance performance and accuracy the best model extracted from feature extraction has been fine tuned to utilize existing pre-trained feature extracted model's knowledge. In Fine-Tuning the model has obtained a gradual increase in validation accuracy from 60.2% to 73.90% which shows the performance of the model has increased significantly with the epochs in Figure 4. The notable difference in training as well as validation accuracy can be attributed to the relatively optimal degree of training of the model. Figure 4 shows Train Vs Val Accuracy of Fine-Tuned Model The loss has also been represented in Figure 5.

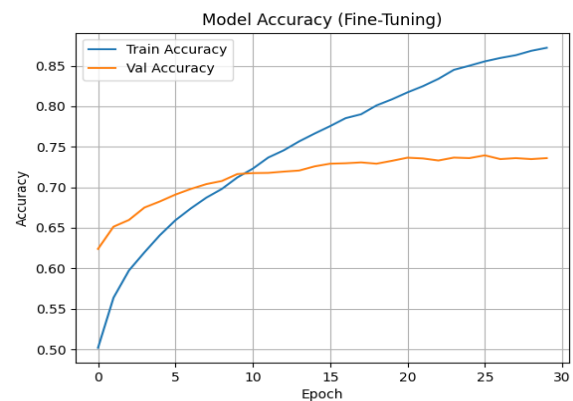


Figure 4 Train Vs Val Accuracy of Fine-Tuned Model

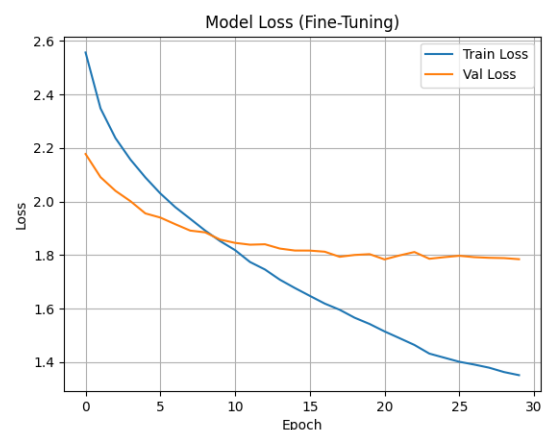


Figure 5 Train Vs Val Loss of Fine Tuned Model

Figure 6 represents the confusion matrix of the classification model indicating the correct and

incorrect predictions of the true labels ranging from 0-100 representing the 101 classes of the Food-101 dataset. Various metrics have been evaluated by the Confusion matrix and respective weighted average has been calculated i.e Accuracy - 0.82, Precision - 0.82, Recall - 0.82 and F1-score - 0.81 for 101 classes of Food-101 dataset. Figure 6 Confusion Matrix for Food-101 Dataset [24]

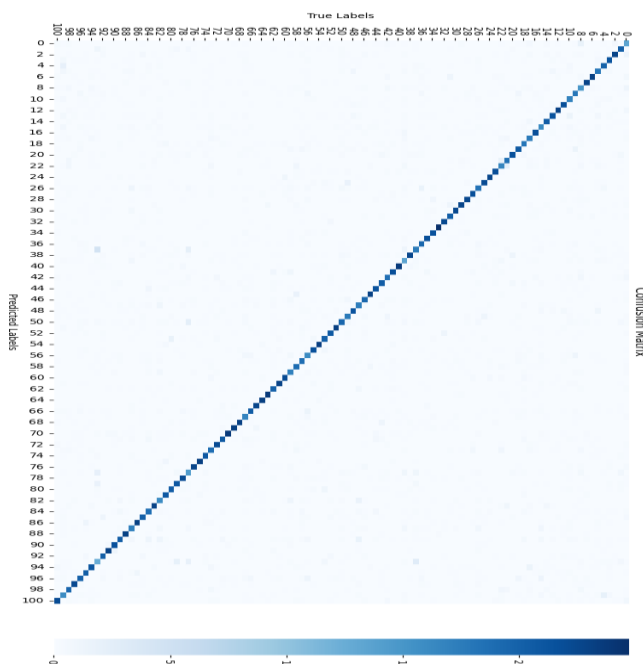


Figure 6 Confusion Matrix for Food-101 Dataset

In final evaluation, on classification on 25250 testing images the model finally achieves accuracy of 79.29% followed by TTA model obtained 81.59% accuracy. Table 1 shows Training and Testing-Accuracy and Loss

Table 1 Training and Testing- Accuracy and Loss

Steps	Train Accuracy (%)	Loss	Val Accuracy (%)	Loss
Feature Extraction	51.06	1.91	60.23	1.60
Fine-Tuning	87.21	1.35	73.90	1.80

After image classification to achieve nutritional targets, mlp training is done. The MLP training shows the mean squared error and mean absolute error observed during MLP training over 50 epochs in Figure 7 and Figure 8.

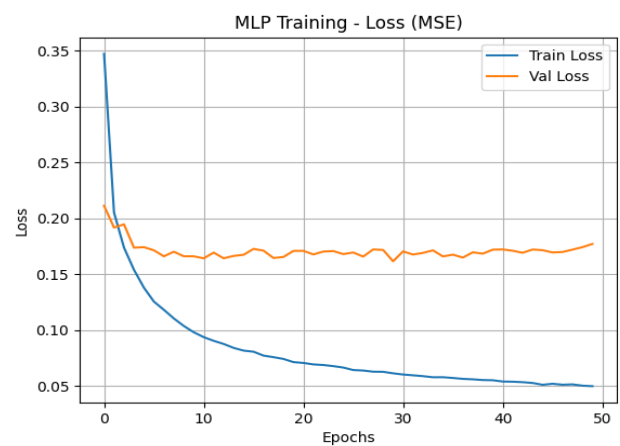


Figure 7 Mean Squared Error in MLP Training

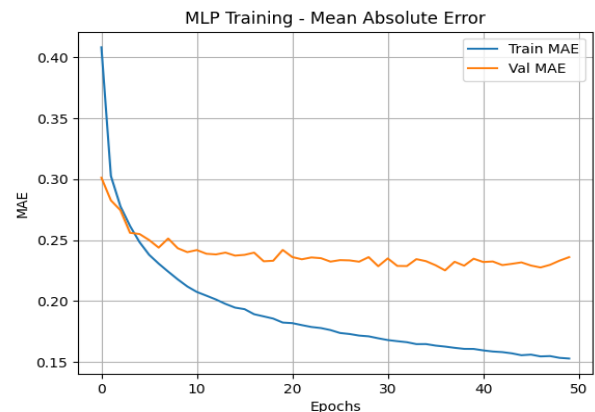


Figure 8 Mean Absolute Error in MLP Training

Conclusion

In this paper, ResNet 50 through transfer learning has achieved an accuracy of 81.59% for image classification and MLP model analysis of the Food 101 dataset, a popular food picture recognition dataset shows a greater accuracy of 97% to identify food image and achieving its nutritional targets i.e, calories, protein, carbs, fat, sugar and fiber. CNN models, ResNet 50 can properly categorise and analyse food photos to assess nutritional value,

according to the research. Diversity in the dataset helps models generalise, improving real-world performance. Transfer learning improved CNN performance when pre-trained models were used. When tested against online photos outside of Food 101, the models showed high generalisation. These results demonstrate the usefulness of sophisticated deep learning models for nutritional analysis and the feasibility of incorporating them into mobile food-tracking apps to meet the growing public health and nutrition emphasis. Transfer learning lets it perform well on smaller in-domain datasets, demonstrating its adaptability and efficiency in many settings. Generating diet-health information via nutrition science and public health improves population health. Public health nutrition has moved from nutritional insufficiency to excess, imbalances, and sustainability. Sustainable food systems need nutrition science to include environmental factors. Nutrition research confronts problems from "nutritional scientism," which simplifies and misuses scientific facts. Misinformation about food and health and ethical concerns might result. Reinterpreting and using science more accurately is needed to avoid these errors. Because nutrition research is complex and limited, public scepticism plagues it. Promoting openness and multidisciplinary teamwork will repair trust and boost nutritional insights' social worth. Identifying dietary patterns that impact health outcomes, nutrition science influences diet and health outcomes. World nutrition issues need focused public health actions to reduce dietary hazards and inequities. Healthy eating policies, community activities, and educational campaigns may combat undernutrition and overnutrition globally. We conclude that nutrition science and public health work together to create population-level health policy.

References

- [1]. Islam, M. T., Siddique, B. N. K., Rahman, S., & Jabid, T. (2018, October). Image recognition with deep learning. In 2018 International conference on intelligent informatics and biomedical sciences (ICIIBMS) (Vol. 3, pp. 106-110). IEEE.
- [2]. He, J., Mao, R., Shao, Z., Wright, J. L., Kerr, D. A., Boushey, C. J., & Zhu, F. (2021). An end-to-end food image analysis system. arXiv preprint arXiv:2102.00645. <https://doi.org/10.2352/issn.2470-1173.2021.8.imawm-285>
- [3]. Gilal, N. U., Al-Thelaya, K., Al-Saeed, J. K., Abdallah, M., Schneider, J., She, J., ... & Agus, M. (2024). Evaluating machine learning technologies for food computing from a data set perspective. *Multimedia Tools and Applications*, 83(11), 32041-32068. <https://doi.org/10.1007/s11042-023-16513-4>
- [4]. Palumbo, M., Cefola, M., Pace, B., Attolico, G., & Colelli, G. (2023). Computer vision system based on conventional imaging for non-destructively evaluating quality attributes in fresh and packaged fruit and vegetables. *Postharvest Biology and Technology*, 200, 112332. <https://doi.org/10.1016/j.postharvbio.2023.112332>
- [5]. Zhang, Y., Deng, L., Zhu, H., Wang, W., Ren, Z., Zhou, Q., ... & Wang, S. (2023). Deep learning in food category recognition. *Information Fusion*, 98, 101859. <https://doi.org/10.1016/j.inffus.2023.101859>
- [6]. Salim, N. O., Zeebaree, S. R., Sadeeq, M. A., Radie, A. H., Shukur, H. M., & Rashid, Z. N. (2021, July). Study for food recognition system using deep learning. In *Journal of Physics: Conference Series* (Vol. 1963, No. 1, p. 012014). IOP Publishing. <https://doi.org/10.1088/1742-6596/1963/1/012014>
- [7]. Li, Y. (2022, January). Research and application of deep learning in image recognition. In 2022 IEEE 2nd international conference on power, electronics and computer applications (ICPECA) (pp. 994-999). IEEE. <https://doi.org/10.1109/icpeca53709.2022.9718847>
- [8]. Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures,

<https://doi.org/10.1109/iciibms.2018.8550021>

- applications, future directions. *Computation*, 11(3), 52. <https://doi.org/10.3390/computation11030052>
- [9]. Shomron, G., & Weiser, U. (2018). Spatial correlation and value prediction in convolutional neural networks. *IEEE Computer Architecture Letters*, 18(1), 10-13. <https://doi.org/10.1109/lca.2018.2890236>
- [10]. [10]. Tay, W., Kaur, B., Quek, R., Lim, J., & Henry, C. J. (2020). Current developments in digital quantitative volume estimation for the optimisation of dietary assessment. *Nutrients*, 12(4), 1167. <https://doi.org/10.3390/nu12041167>
- [11]. Strączek, A., Szalkowska, J., Sutkowska, P., Srebrna, A., Puzio, N., Piasecka, A., ... & Thum-Tyzo, K. (2023). Impact of nutrition on the condition of the oral mucosa and periodontium: A narrative review. *Dental and Medical Problems*, 60(4), 697-707. <https://doi.org/10.17219/dmp/156466>
- [12]. Khonje, M. G., Ecker, O., & Qaim, M. (2020). Effects of modern food retailers on adult and child diets and nutrition. *Nutrients*, 12(6), 1714. <https://doi.org/10.3390/nu12061714>
- [13]. Llorach, R., Garcia-Aloy, M., Tulipani, S., Vazquez-Fresno, R., & Andres-Lacueva, C. (2012). Nutrimetabolomic strategies to develop new biomarkers of intake and health effects. *Journal of Agricultural and Food Chemistry*, 60(36), 8797-8808. <https://doi.org/10.1021/jf301142b>
- [14]. Kussmann, M., Affolter, M., Nagy, K., Holst, B., & Fay, L. B. (2007). Mass spectrometry in nutrition: understanding dietary health effects at the molecular level. *Mass spectrometry reviews*, 26(6), 727-750. <https://doi.org/10.1002/mas.20147>
- [15]. Katz, D. L., Njike, V. Y., Faridi, Z., Rhee, L. Q., Reeves, R. S., Jenkins, D. J., & Ayoob, K. T. (2009). The stratification of foods on the basis of overall nutritional quality: the overall nutritional quality index. *American Journal of Health Promotion*, 24(2), 133-143. <https://doi.org/10.4278/ajhp.080930-quant>
- 224
- [16]. Ocké, M. C., Westenbrink, S., van Rossum, C. T., Temme, E. H., van der Vossen-Wijmenga, W., & Verkaik-Kloosterman, J. (2021). The essential role of food composition databases for public health nutrition—Experiences from the Netherlands. *Journal of Food Composition and Analysis*, 101, 103967. <https://doi.org/10.1016/j.jfca.2021.103967>
- [17]. Kelly, O. J., Gilman, J. C., & Ilich, J. Z. (2018). Utilizing dietary micronutrient ratios in nutritional research may be more informative than focusing on single nutrients. *Nutrients*, 10(1), 107. <https://doi.org/10.3390/nu10010107>
- [18]. Ma, C., Ho, S. K., Singh, S., & Choi, M. Y. (2021). Gender disparities in food security, dietary intake, and nutritional health in the United States. *Official journal of the American College of Gastroenterology| ACG*, 116(3), 584-592. <https://doi.org/10.14309/ajg.0000000000001118>
- [19]. Dave, L. A., Hodgkinson, S. M., Roy, N. C., Smith, N. W., & McNabb, W. C. (2023). The role of holistic nutritional properties of diets in the assessment of food system and dietary sustainability. *Critical reviews in food science and nutrition*, 63(21), 5117-5137. <https://doi.org/10.1080/10408398.2021.2012753>
- [20]. Gami, S. J., Dhamodharan, B., Dutta, P. K., Gupta, V., & Whig, P. (2024). Data Science for Personalized Nutrition Harnessing Big Data for Tailored Dietary Recommendations. In *Nutrition Controversies and Advances in Autoimmune Disease* (pp. 606-630). IGI Global. <https://doi.org/10.4018/979-8-3693-5528-2.ch023>
- [21]. Donovan, S. M., Abrahams, M., Anthony, J. C., Bao, Y., Barragan, M., Bermingham, K. M., ... & Winters, B. L. (2025). Personalized nutrition: perspectives on challenges, opportunities, and guiding principles for data use and fusion. *Critical Reviews in Food*

Science and Nutrition, 1-18.
<https://doi.org/10.1080/10408398.2025.2461237>

- [22]. Yanai, K., & Kawano, Y. (2015, June). Food image recognition using deep convolutional network with pre-training and fine-tuning. In 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1-6). IEEE. <https://doi.org/10.1109/ICMEW.2015.7169816>
- [23]. Pouladzadeh, P., & Shirmohammadi, S. (2017). Mobile multi-food recognition using deep learning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 13(3s), 1-21. <https://doi.org/10.1145/3063592>
- [24]. Aktı, Ş., Qaraqe, M., & Ekenel, H. K. (2022, May). A mobile food recognition system for dietary assessment. In *International Conference on Image Analysis and Processing* (pp. 71-81). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-13321-3_7