Exploring the Potential of AI in Nutrition: Food Recognition and Classification for Improved Dietary Assessment

Jan Chodora*, Md Rakibul Islam†, Johanna Melander‡
School of Innovation, Design and, Engineering
Mälardalens University, Västerås, Sweden
Email: *jca22003@student.mdu.se, †md.rakibul.islam@mdu.se, ‡jmr19002@student.mdu.se

***Abstract*—This study aims to develop an AI-based system focused on recognizing food items to promote a manageable healthy lifestyle and encourage healthier eating habits. The study explores various methods for food recognition and classification, with a focus on Convolutional Neural Networks (CNNs) for feature extraction and classification. The VGG-16 model is used for feature extraction, and different classification models, including Support Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), Naive Bayes (NB), Gradient Boosting (GB), Deep Neural Network (DNN), and Extreme Learning Machine (ELM), are evaluated. The accuracy and training time of each model are compared. Additionally, explainability of the AI system is addressed using the LIME (Local Interpretable Model-Agnostic Explanations) technique to explain the decision-making process. The hardware and software used for the experiments are described, and the Food-101 dataset is used for training and testing. The results show that combining CNN for feature extraction with SVM for classification achieves the highest accuracy of 80%. The limitations and future work of the project are discussed, including implementing the AI model on mobile devices and estimating the amount of food presented in the images for nutrition calculation. The study contributes to the field of food recognition and classification and provides insights into the development of AI systems for dietary assessment and healthy lifestyle promotion.**

***Index Terms*—Convolutional Neural Networks (CNNs), Deep Neural Network (DNN), Extreme Learning Machine (ELM), Gradient Boosting (GB), K-Nearest Neighbor (KNN), Naive Bayes (NB), Principal Component Analysis (PCA), Random Forest (RF), Support Vector Machine (SVM), eXplainable AI (XAI)**

I. Introduction

A healthy lifestyle and diet are crucial for our wellness and can help prevent life-threatening and chronic diseases [1], [2]. To evaluate one's diet, dietary assessment is conducted by collecting data on the person's food and beverage consumption during a fixed time period [3]. The information is then recorded and analysed using food composition tables to calculate the intake of energy and nutrients. Mobile applications are popular tools for monitoring diets, but the applications require a lot of input from the user [1], [2], [4].

The purpose of this project is to promote a manageable healthy lifestyle by encouraging healthier eating habits through the development of an AI-based system focused on recognizing food items.

In the related work section, several research papers and studies were reviewed. The focus was to study various methods for food recognition and classification, as well as nutrition calculation. In the area of classification, Convolutional Neural Networks (CNNs) were a commonly used solution, mainly due to their high accuracy [5], [6], [7], [8], [2]. They excel at tasks like food recognition and segmentation because they can learn hierarchical representations of visual features directly from raw input data without the requirement for manual feature engineering [7]. On the other hand, it requires relatively big dataset for training. Extracted features may contain information which is not relevant for classification, techniques such as PCA, can reduce the dimensionality of the data on the output layer of the CNN. The reason for using PCA was to reduce dimensionality and thus make training and classification faster. As mentioned, CNNs require large datasets to be trained, it also requires massive computational resources, to work around this a pretrained model was introduced[2]. The transfer VGG-16 was a common model used for image recognition problems and provided good accuracy, for that reason, the VGG-16 model was used[2]. For classification, several models were examined, some commonly used in this area [2], such as Support Vector Machines (SVM), Random Forrest (RF), K-Nearest Neighbor (KNN), Naive Bayes (NB) and some that are not commonly used for these kinds of problems, such as Gradient Boosting (GB), Deep Neural Network (DNN) and Extreme Learning Machine (ELM). The ELM, GB and DNN were selected for the reason of not being commonly used models, the purpose for that was to investigate if these three models could help improve the classification accuracy and explore the time interval for classification. For better understanding how does the black-box model make prediction, XAI will be used. XAI strengthen trust and transparency in AI systems by explaining how complex deep learning models, generate their result [9], [10], [11]. Explainability of AI is an important research area, it is especially important in the healthcare sector because of the importance to build trust in the system [12]. To provide explanation, XAI technique LIME was chosen since it can provide model agnostic explanation for the importance of extracted features for classification [13].

II. Problem Formulation

The main research questions that need to be addressed in this project are:

RQ1: How can CNN-PCA be used for feature extraction?

RQ1 focuses on feature extraction, which is a crucial step in building an accurate food recognition model. Using CNN-PCA for feature extraction can help to reduce the dimensionality of the features and improve the model's performance.

RQ2: Which model fits for classification?

RQ2 aims to address the challenge of accurately classifying the food images. Finding the best model for classification is important as it improves the accuracy and efficiency in the model.

RQ3: How can the process behind decision-making be explained?

RQ3 focuses on explainability, which is becoming increasingly important in AI applications. Using techniques such as LIME can help to explain the decision-making process and provide insights into the important features that contribute to the classification outcome.

III. Background

A. Convolutional Neural Network (CNN)

Convolutional neural networks (CNNs) are composed of layers, with each layer responsible for decomposing the input image into a set of feature maps. Then, using these feature maps, inferences can be drawn about the image's contents. Most importantly, a CNN performs a convolution operation, which involves passing the input image through a collection of learnable filters to produce a series of feature maps [14].

Each filter in the convolutional layer can be thought of as a small window that is swept across the input image to produce a dot product between the filter and the corresponding region of the image. Through the use of backpropagation [7], the CNN is able to learn a set of filters that allow it to recognise local features such as edges, corners, and blobs. In subsequent network layers, these local features can be combined to form more sophisticated features.

In addition to convolutional layers, a regular CNN also has pooling layers. These layers determine the highest or average value across relatively tiny sections of the feature maps, hence downsampling the feature maps. This lessens the dimensionality of the feature maps and improves the network's performance as a whole.

The representations that CNNs can learn are translation invariant, meaning that they are unaffected by even small shifts in the position of the object being recognised. This is arguably one of the greatest benefits of using a CNN. This allows the network to identify items regardless of their position or orientation inside the image [14].

Using a large dataset of food images labelled with the segmentation mask or food category, a convolutional neural network (CNN) can be trained to recognise and segment food. During training, the network is trained to identify and extract visual elements that are most useful for classifying foods and determining their boundaries. With these features, the network can more reliably evaluate food photos for quality. Automatic food detection and segmentation is made possible by using the trained network to generate predictions on new photos.

Convolutional neural networks (CNNs) are a powerful class of neural networks that are particularly well-suited to image processing and computer vision applications. They have been effectively applied to many different challenges in food detection and segmentation, and they remain a hotspot for AI researchers.

B. Principal Component Analysis (PCA)

Dimensionality reduction and feature extraction are two common applications of Principal Component Analysis (PCA) in the field of machine learning. It relies on finding a dataset's Principal Components (PC). Linear combinations of the original features known as principal components capture the most variation in a dataset [15]. PCA is particularly useful when dealing with high-dimensional data, as this type of data often contains many features that are highly correlated or redundant.

Finding the covariance matrix's eigenvectors and eigenvalues is essential to PCA's mathematical formulation. Principal components are symbolised by eigenvectors, while the amount of variance explained by each component is shown by its corresponding eigenvalue. Each successive principal component is computed as the eigenvector with the next largest eigenvalue, with the restriction that it must be orthogonal to all of the components that came before it. The first principal component is the eigenvector with the largest eigenvalue.

PCA is useful for reducing the dimensionality of the feature space and extracting relevant features for further use in classification and regression when applied to image data [16]. In the context of food recognition and the estimation of nutritional values, for example, PCA can be used to extract useful features from images of food. Colour, texture, and shape are all examples of features that can be fed into a machine learning algorithm.

C. Classifiers

1) Support Vector Machine: The main idea behind Support Vector Machine (SVM) is to project datapoints into high dimensional feature space where they are linearly separable [17].

2) Random Forest: The Random Forest forms a forest-like structure by creating decision trees through the process of random sampling with replacement[18].

3) K-Nearest Neighbor: K-Nearest Neighbor (KNN) is a classification method which builds its decision upon classes of samples which are in future space close to the classified sample. This method is especially usable when there is little prior knowledge about the data distribution [19].

4) Naive Bayes: The Naive Bayes is a probabilistic classifier that uses Bayes' theorem with the "naive" assumption of independence between features. It calculates the probability of a class given a set of features by multiplying the individual probabilities of each feature occurring in that class[20].

5) Gradient Boosting: Gradient Boosting is an algorithm for classification or regression tasks that frequently fits new models to make the estimate of the response variable more accurate. The GB iteratively trains new models to minimize the gradient of the loss function with respect to the ensemble's predictions[18], [21].

6) Deep Neural Network: Deep Neural Network (DNN) is an Artificial Neural Network (ANN) model which contains multiple hidden layers [22], [23].

7) Extreme Learning Machine (ELM): In an ELM, the weights between the input and hidden layers are initially generated at random, and the hidden layer's output is calculated using a nonlinear activation function, such as a sigmoid or radial basis function. After that, solving a linear system of equations to obtain the weights of the output layer, which involves locating the Moore-Penrose generalised inverse of the output matrix of the hidden layer [24]. The weights of the output layer can then be derived from this process. To prevent overfitting, ELM includes a regularisation term within the objective function, which is minimised.

D. Explainable AI

"Explainable AI," or "XAI," refers to an AI system's ability to offer a computationally tractable explanation for its inferences and conclusions. To better understand how the AI algorithm recognises and categorises foods and how it determines nutritional information, XAI can be applied to this project.

One approach that could be taken with XAI is to use saliency maps, which highlight the parts of an image that were most important in the AI's decision-making process. This is merely one of several viable options. Users can gain insight into how a model made its classification decisions by examining saliency maps, which can shed light on why a given food item was assigned a particular category. Users can also gain insight into the reasoning behind a dish's classification with the aid of saliency maps [25].

Decision trees and decision rules are another method of XAI that can be used to explain the AI's thought process and how it arrived at a conclusion. When presented with decision trees, which can simplify the otherwise convoluted decision-making process and break it down into a series of straightforward steps, users will have an easier time understanding and trusting the AI's output [26].

By spotting potential biases or errors that were missed during the model's development phase, XAI is able to help improve the model's accuracy and robustness. XAI can do this in addition to explaining the AI's decision-making process in a way that users can comprehend.

The incorporation of XAI techniques into this project can, all things considered, help to ensure that the AI algorithm is transparent and trustworthy, which, in the end, can lead to greater user acceptance and adoption of the technology.

IV. Related Work

There is numerous research and projects done in area of diet tracking as well in area of image recognition. Yet, connecting these two areas may rise many problems which need to be solved in order to develop system which can help with diet tracking through estimation of food features like number of calories. This section will provide an overview of some of the work related to this project.

In a paper from 2016 authors addressed zero-shot retrieval problem [5], in other words food classification based on the appearance of the whole dish in contrast with first understanding its composition. Based on this problem, authors proposed a system for ingredient recognition and receipt retrieval using CNN.

Another CNN-based architecture was NutriNet [6], which was trained on a dataset of 225,953 images from 520 food and drink classes. The model was tested on 130,517 images, on which it achieved 94.47% accuracy. Yet on real world images it achieved Top-5 accuracy of 55%. The model is used as part of a mobile app designed to assist patients struggling with Parkinson's disease.

In a paper from 2018 authors introduced a new dataset which consists of 247,636 images classifiable into 457 food classes [7]. In this paper authors aimed to investigate CNN architectures from which they chose a Residual Network with 50 layers as the best. Authors are introducing new benchmark food database Food-475 and in conclusion, authors are proposing that in order to be able to increase any model's performance, a larger database needs to be created.

In another work which introduced new dataset, authors collected 29,515 real word images using wearable camera eButton [8], these images were later tagged by already existing CNN model and classified by handwritten function achieving 91.5% food detection accuracy.

A novel system to automatic estimate food attributes such as ingredients and nutritional value, with up to 80% classification accuracy [27]. The system is implemented as a mobile app.

Authors are using their own dataset, on which they achieved 83.8% classification accuracy on single corp and 85% on multi corp data. In comparison with Food-101 dataset on which accuracy achieved was 78.3% on single corp and 79.22% on multi corp data. The work considers three approaches for calculating food attributes. First approach might be food volume estimation with combination of nutrition tables to calculate food attributes. A second possible approach would be direct estimation of the food attributes. Instead, the authors chose the approach of vector space embedding with vector space representation of words from a large dataset.

In this work [28] authors are using a collection of 50 categories which consists of 1000 images. In the next step, the authors performed food recognition using pre-trained learning model MobileNet using TensorFlow Lite, which performed food recognition with 80% accuracy. Authors performed data collection into Google Images database Food-256, training of the MobileNet model and transfer learning model using cloud database setup. Online process includes mobile application development.

In a review from 2022 [2] about image-based food-recognition systems, they reviewed 78 research papers. The main goal of the review was to present potential solutions that integrate camera-equipped mobile devices with integrate cameras on a mobile device with computer vision methods for supporting dietary assessment. To facilitate cost assessments, the process was divided into different steps, these steps integrated the camera on a mobile device with computer vision techniques. The different steps were presented in a table, where they compare the different methods. The different columns represented: user input preprocessing, segmentation, feature extraction, dimensionality reduction, classification, volume estimation, and dataset and performance. They also presented a table with a summary of classification methods where Artificial Neural Network (ANN), Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (KNN), Random Forest (RF), Convolutional Neural Network (CNN) were presented. The result showed that CNN appeared to be the most effective solution with just one disadvantage, that a large dataset is needed for the training of the CNN. The highest accuracy for using the Food-101 dataset was 90.27% and the lowest accuracy was 55.3%. The highest rate was achieved by using CNN for feature extraction, no dimensionality reduction and CNN(WISeR) for classification. The lowest rate was by using the VGG-16 model for feature extraction, no dimensionality reduction and adaptive CNN for classification. Another model using the CNN(VGG) and SVM for classification achieved an accuracy of 72.88% on the dataset Instagram800K.

The research paper from 2019 [29] introduces the Ville Cafe dataset, which collected over 35,000 images across 16 simple categories, such as salad, fruit, toast, eggs, and hamburger. The authors based their model on R-CNN with union postprocessing. On combination of their dataset with the Food-256 dataset, the authors achieved an accuracy of 99.86%. Additionally, the authors experimented with weight estimation on 8 food classes, achieving an average relative error of 0.13.

V. METHOD

A. Hardware

Experiments were run on a device which had a processor AMD Ryzen 7 5800H and 16 GB RAM. This device was running operating system Microsoft Windows 11 Home.

B. Software

The model was implemented in Python using libraries Scikit-Learn, TensorFlow, Keras.

C. Dataset

The dataset Food-101 consists of 101,000 images of 101 food categories [30]. On purpose, 75% of the images in this dataset contain noise, which might represent more real-world cases and might be used as training set. The images have been rescaled and have a length of 512 pixels maximum. For the model, 5 classes were used for the training of the model, each class consisting of 1000 images. Out of this subset, 4000 images were used for training and 1000 images for testing. The dataset was provided via TensorFlow [30].

D. Pre-processing

Preprocessing is a crucial step in building an accurate classification system. The dataset Food-101 contains images of various sizes, to make the data compatible with the model, the data needs to be resized to a specific size. The images in the dataset are adjusted to 128x128 pixels. The dataset was split with probability of 80% for train set and 20% for test set.

E. Feature Extraction

Visual Geometry Group 16 (VGG-16) was used for feature extraction. The model have 13 convolutional layers, 5 max pooling layers and 3 dense layers but has only 16 weight layers, which is the learnable parameters layers [31], [32]. To use the VGG-16 model, the TensorFlow library was used.

F. Dimensionality Reduction

To preform the dimensionality reduction, the PCA was used. First data is standardized and then PCA is used to get reduced feature representation, number of PCs was set to 300, this number was chosen experimentally.

G. Classification

For the classification, several models were tested through the sklearn library.

1) Support Vector Machine was used via Scikit-Learn library as probabilistic C-Support Vector Classification.
2) Random Forest was used via Scikit-Learn library.
3) K-Nearest Neighbor was used via Scikit-Learn library.
4) Naive Bayes was used via Scikit-Learn library as Gaussian Naive Bayes.
5) Gradient Boosting was used via Scikit-Learn library.
6) Deep Neural Network was used via Scikit-Learn library as Multilayer Perceptron.
7) Extreme Learning Machine was implemented from scratch.

H. Explainable AI

LIME (Local Interpretable Model-Agnostic Explanations) was implemented through the LIME library in Python.

I. Evaluation Metrics

To evaluate the results, accuracy metric is used. To tell how well a classification algorithms performs, the accuracy rate will measure how good it is[1], [2]. In a binary system, accuracy is determined by the following calculation:
$A = \frac{TP+TN}{TP+TN+FP+FN}$ where A stands for the accuracy, TP stands for the true positives, TN stands for the true negatives, FP stands for the false positives and FN stands for the false negatives. TP describes the number of cases that has a positive label and are classified as members of the positive class, the TN describes the number of cases that have a negative

label and are classified as members of the negative class. FP describes the number of cases that have a negative label and are classified as a member of the positive class, and FN describes the number of cases that have a positive label and are classified as a member of the negative class.

J. Experiment

Two experiments was tested, figure 1 represents the experiment where the tranfer model VGG-16 was used for feature extraction, PCA for dimensionality reduction followed by the classifiers. The other figure (figure 2) represents the second experiment where the dimensionality reduction is removed.

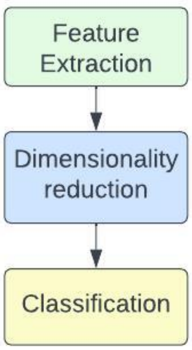

Figure 1. CNN-PCA-Classifier

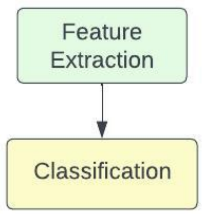

Figure 2. CNN-Classifier

VI. Results

The results form the experiment are represented in table. The first one represents the accuracy in [%] of each model and the second one is showing the training time in [s].

Accuracy [%]		
Classification Model	CNN-PCA	CNN
SVM	38.5	80.0
RF	40.9	75.3
KNN	34.7	50.5
NB	32.4	60.7
GB	41.6	75.4
DNN	33.8	75.6
ELM	21.6	21.6

Training Time [s]		
SVM	6.213	195.413
RF	5.630	9.481
KNN	0.003	0.028
NB	0.005	0.156
GB	186.844	1632.920
DNN	1.892	22.046
ELM	23.518	237.214

Figure 3 depicts LIME explanation for one representative point. Each line represents one class and the decision to classify the sample as this class.

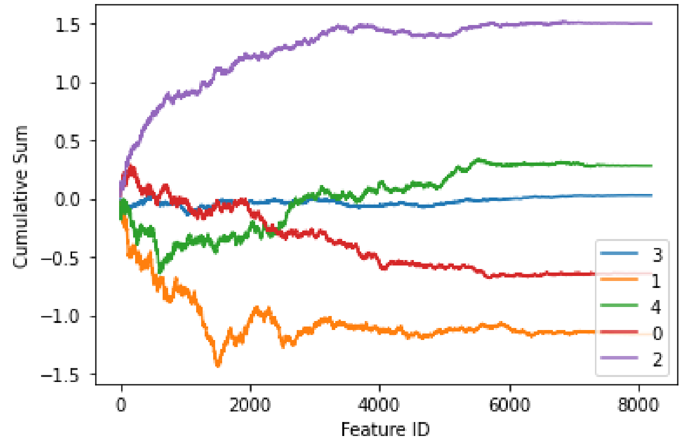

Figure 3. LIME explanation for feature importance for decision-making

VII. Discussion

To evaluate our result the accuracy metric was used, different results were achieved depending on the model's structure. According to the review [2], a result over 70% is a good result and a result of over 80% is an excellent model. The best result was through combining the transfer model VGG-16 for feature extraction with SVM classifier, which gave a result of 80%.

Other works using similar models (VGG for feature extraction and SVM classification) achieved an accuracy of 72.88% on the Instagram800K dataset. The CNN seems to be the most commonly used method for feature extraction, but the accuracy of the model varies depending on the selected method for classification.

In the review [2], the best accuracy of the dataset Food-101 was by combining the CNN for feature extraction and CNN(WIeR) for classification (90.27%), so compared to the best result achieved in this project (CNN for feature extraction and SVM for classification), the CNN(WIeR) seems to be a better method for the classification problem. The lowest result in the review on the same dataset was by combining the VGG-16 model for feature extraction and adaptive CNN for classification, the accuracy was 55.3%, solution proposed in this paper demonstrated an improvement in the accuracy.

Interesting results were achieved by using not so commonly used methods for food recognition. By combining the transfer model with GB, the accuracy was 75.4% and the accuracy for the DNN was 75.6%. The training time was much longer for the GB model compared to other models, one reason for that might be the structure of the GBs algorithm and also the amount of data used. The iterative process builds on a sequence of models to correct the mistakes by the previous models, this means that it learns from the past mistakes. These steps may slow the GB model down.

DNN posses the ability to learn non-linear dependencies as well as it can handle high dimensional data, this might be the reason for the classification based on extracted features to result in high accuracy.

For the ELM model the algorithms before the classification did not affect the result of the accuracy, it stayed firmly at 21.6%. Unsatisfactory performance of ELM could have been caused by possible factors such as insufficient dataset size or data preprocessing, as well as internal causes like hyperparameter tuning. When using the combination of VGG-16 for the feature extraction and PCA for dimensionality reduction, the training time decreased significantly, but the accuracy achieved was not high. The reason for that might be in high amount of noise and high variance contained in the input images.

Direct classification after feature extraction without dimensionality reduction points into too much information being excluded by the reduction. This might have been caused by possibly non-linear relationship in input data. Exploring different methods of dimensionality reduction might bring improvement. As it is indicated by LIME technique that some information contained in the input image could be reduced.

A. Limitations of this project

Several limitations were encountered in this project that have impacted the progress. The foremost constraint has been the time and the time structure through this project. The problem was far more complex than expected. At the beginning of the project research questions were selected which were too broad and complex for this amount of time, narrowing the problem down immediately would have been more beneficial. During the course of the project, more details about the area and selected approach were discovered, and thus research questions changed. Another challenge was finding the right dataset for this project, it was soon discovered that the perfect dataset that would enable us to accurately calculate nutrition and energy, did not exist. In order to utilize the large dataset that was selected, substantial computational power was required, which was not easily available. In order to save training time, 5 classes of the dataset were used.

VIII. Conclusion

The PCA did not improve the accuracy of our model but using it for dimensionality reduction it reduced the training time. XAI technique LIME which was used to provide insight into classifier's prediction revealed that a prediction is mostly based on summary of all features rather than a small subset. The highest accuracy was achieved by combining the CNN and SVM. For explanation of which CNN extracted features had the biggest impact on the decision made, LIME can be used.

A. Future Work

Improving the AI model that is used for food recognition is the primary focus here.

The main goal of this research is to make dietary assessment available and usable for anyone. As a proposal for future work is investigation of the feasibility of implementing the previously developed AI model on mobile devices such as smartphones. The objective is to create an easy-to-use smartphone application that makes use of the artificial intelligence model that has been developed.

This research would involve the exploration and implementation of a variety of techniques for adapting the AI model to work on smartphones. These techniques might include compressing the size of the model, optimising its performance for lower computing power, and developing a user-friendly interface for uploading images and receiving nutrition reports. In addition to this requirement, the developed smartphone application would need to protect users' privacy and keep their data secure.

Another challenge is to find and implement a technique which will help to estimate the amount of the food presented in the image data.

For right dietary assessment, it is important to estimate nutrition values of the meal. The objective is to devise a technique that is both precise and all-encompassing in order to compute the nutrient density of the food depicted in the pictures. This research might involve exploring various methods for calculating the nutrition values, such as using external nutrition databases, developing custom algorithms based on recognised food items, and exploring techniques such as feature extraction and machine learning to improve the accuracy of nutrient calculation. Ultimately, the goal of this research is to develop a more comprehensive understanding of how to improve the accuracy of nutrient calculation.

Acknowledgment

The authors would like to thank Carl Ahlberg for the support and guiding through this project and Mälardalens Univeristy for making this project possible.

References

[1] R. Kaur, R. Kumar, and M. Gupta, "Deep neural network for food image classification and nutrient identification: A systematic review," *Reviews in Endocrine and Metabolic Disorders*, Mar. 2023. [Online]. Available: https://doi.org/10.1007/s11154-023-09795-4

[2] K. V. Dalakleidi, M. Papadelli, I. Kapolos, and K. Papadimitriou, "Applying Image-Based Food-Recognition Systems on Dietary Assessment: A Systematic Review," *Advances in Nutrition (Bethesda, Md.)*, vol. 13, no. 6, pp. 2590–2619, Dec. 2022.

[3] M. C. Dao, A. F. Subar, M. Warthon-Medina, J. E. Cade, T. Burrows, R. K. Golley, N. G. Forouhi, M. Pearce, and B. A. Holmes, "Dietary assessment toolkits: an overview," *Public Health Nutrition*, vol. 22, no. 3, pp. 404–418, Mar. 2019. [Online]. Available: https://www.cambridge.org/core/product/identifier/S1368980018002951/type/journal_article

[4] M. Tosi, D. Radice, G. Carioni, T. Vecchiati, F. Fiori, M. Parpinel, and P. Gnagnarella, "Accuracy of applications to monitor food intake: Evaluation by comparison with 3-d food diary," *Nutrition*, vol. 84, p. 111018, Apr. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0899900720303014

[5] J. Chen and C.-w. Ngo, "Deep-based Ingredient Recognition for Cooking Recipe Retrieval," in *Proceedings of the 24th ACM international conference on Multimedia*, ser. MM '16. New York, NY, USA: Association for Computing Machinery, Oct. 2016, pp. 32–41. [Online]. Available: https://dl.acm.org/doi/10.1145/2964284.2964315

[6] S. Mezgec and B. Koroušić Seljak, "NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment," *Nutrients*, vol. 9, no. 7, p. 657, Jul. 2017, number: 7 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-6643/9/7/657

[7] G. Ciocca, P. Napoletano, and R. Schettini, "CNN-based features for retrieval and classification of food images," *Computer Vision and Image Understanding*, vol. 176-177, pp. 70–77, Nov. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314218302467

[8] W. Jia, Y. Li, R. Qu, T. Baranowski, L. Burke, H. Zhang, Y. Bai, J. Mancino, G. Xu, Z.-H. Mao, and M. Sun, “Automatic food detection in egocentric images using artificial intelligence technology,” *Public Health Nutrition*, vol. 22, no. 7, pp. 1168–1179, 2019.

[9] Y.-C. Wang, T.-C. T. Chen, and M.-C. Chiu, “An improved explainable artificial intelligence tool in healthcare for hospital recommendation,” *Healthcare Analytics*, vol. 3, p. 100147, Nov. 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S277244252300014X

[10] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, “Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022),” *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107161, Nov. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169260722005429

[11] A. Kale, T. Nguyen, F. C. Harris, C. Li, J. Zhang, and X. Ma, “Provenance documentation to enable explainable and trustworthy AI: A literature review,” *Data Intelligence*, vol. 5, no. 1, pp. 139–162, Mar. 2023. [Online]. Available: https://direct.mit.edu/dint/article/5/1/139/109494/Provenance-documentation-to-enable-explainable-and

[12] S. F. Nimmy, O. K. Hussain, R. K. Chakrabortty, F. K. Hussain, and M. Saberi, “An optimized Belief-Rule-Based (BRB) approach to ensure the trustworthiness of interpreted time-series decisions,” *Knowledge-Based Systems*, vol. 271, p. 110552, Jul. 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0950705123003027

[13] C. Lim, K. Goh, and L. Lim, “Explainable Artificial Intelligence in Oriental Food Recognition using Convolutional Neural Network,” 2021, pp. 218–223.

[14] R. Chauhan, K. K. Ghanshala, and R. Joshi, “Convolutional Neural Network (CNN) for Image Detection and Recognition,” in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, Dec. 2018, pp. 278–282.

[15] H. Abdi and L. J. Williams, “Principal component analysis,” *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010, publisher: Wiley Online Library.

[16] T. Kurita, “Principal component analysis (PCA),” *Computer Vision: A Reference Guide*, pp. 1–4, 2019, publisher: Springer.

[17] S. Suthaharan, “Support Vector Machine,” in *Machine Learning Models and Algorithms for Big Data Classification*. Boston, MA: Springer US, 2016, vol. 36, pp. 207–235, series Title: Integrated Series in Information Systems. [Online]. Available: https://link.springer.com/10.1007/978-1-4899-7641-3_9

[18] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, “A comparative analysis of gradient boosting algorithms,” *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021. [Online]. Available: https://link.springer.com/10.1007/s10462-020-09896-5

[19] L. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009. [Online]. Available: http://www.scholarpedia.org/article/K-nearest_neighbor

[20] R. R. Yager, “An extension of the naive Bayesian classifier,” *Information Sciences*, vol. 176, no. 5, pp. 577–588, Mar. 2006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0020025504003457

[21] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Frontiers in Neurorobotics*, vol. 7, 2013. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnbot.2013.00021/abstract

[22] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0893608014002135

[23] ——, “Deep Learning in Neural Networks: An Overview,” 2014, publisher: arXiv Version Number: 4. [Online]. Available: https://arxiv.org/abs/1404.7828

[24] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006, publisher: Elsevier.

[25] M.-H. Song, “A Study on Explainable Artificial Intelligence-based Sentimental Analysis System Model,” *International Journal of Internet, Broadcasting and Communication*, vol. 14, no. 1, pp. 142–151, 2022, publisher: The Institute of Internet, Broadcasting and Communication.

[26] A. Sathyan, A. I. Weinberg, and K. Cohen, “Interpretable AI for bio-medical applications,” *Complex engineering systems (Alhambra, Calif.)*, vol. 2, no. 4, 2022, publisher: NIH Public Access.

[27] R. Yunus, O. Arif, H. Afzal, M. Amjad, H. Abbas, H. Bokhari, S. Haider, N. Zafar, and R. Nawaz, “A Framework to Estimate the Nutritional Value of Food in Real Time Using Deep Learning Techniques,” *IEEE Access*, vol. 7, pp. 2643–2652, 2019.

[28] N. Jamil, S. Roslan, R. Hamzah, and I. Ramli, “Food recognition of Malaysian meals for the management of Calorie intake,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 1.6 Special Issue, pp. 355–360, 2019.

[29] M.-L. Chiang, C.-A. Wu, J.-K. Feng, C.-Y. Fang, and S.-W. Chen, “Food Calorie and Nutrition Analysis System based on Mask R-CNN,” in *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, Dec. 2019, pp. 1721–1728.

[30] “Food-101 – Mining Discriminative Components with Random Forests.” [Online]. Available: https://data.vision.ee.ethz.ch/cvl/datasets_extra/food-101/

[31] Q. Guan, Y. Wang, B. Ping, D. Li, J. Du, Y. Qin, H. Lu, X. Wan, and J. Xiang, “Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study,” *Journal of Cancer*, vol. 10, no. 20, pp. 4876–4882, 2019. [Online]. Available: http://www.jcancer.org/v10p4876.htm

[32] Y. Meir, O. Tevet, Y. Tzach, S. Hodassman, R. D. Gross, and I. Kanter, “Efficient shallow learning as an alternative to deep learning,” *Scientific Reports*, vol. 13, no. 1, p. 5423, Apr. 2023. [Online]. Available: https://www.nature.com/articles/s41598-023-32559-8