Journal of Computer Science

Review

Computer Vision for Reducing Food Waste in an Institutional Canteen: A Literature Review and Performance Analysis

[1]Ana Correia, [1]Clara Aidos, [1,2]João M. L. P. Caldeira and [1,2,3]Vasco N. G. J. Soares

[1]*Polytechnic University of Castelo Branco, Av. Pedro Álvares Cabral n 12, 6000-084 Castelo Branco, Portugal*
[2]*Instituto de Telecomunicações, Rua Marquês d'Ávila e Bolama, Covilhã, Portugal*
[3]*AMA-Agência para a Modernização Administrativa, Rua de Santa Marta, Lisboa, Portugal*

Article history
Received: 12-08-2024
Revised: 15-08-2024
Accepted: 23-08-2024

Corresponding Author:
Vasco N. G. J. Soares
Polytechnic University of Castelo Branco, Av. Pedro Álvares Cabral n 12, 6000-084 Castelo Branco, Portugal
Email: vasco.g.soares@ipcb.pt

Abstract: Food waste in today's society has been the subject of growing interest and discussion, given its economic, environmental, social, and nutritional implications. Although food waste is present throughout the food supply chain, in developed countries it tends to be higher in the final stages of consumption (e.g., households and food services). This study focuses on institutional canteens, where food waste includes prepared meals that have not been sold (i.e., leftovers), as well as food served that is left on plates after the meal has been consumed (i.e., scraps). It presents a first step towards developing a prototype/solution based on computer vision techniques to identify and quantify food waste in an institutional canteen. It begins by introducing the related concepts. It then surveys the state-of-the-art and categorizes existing solutions, presenting their main characteristics, strengths, and limitations. Inception-V3 and ResNet-50 are identified as the most promising computer vision techniques, and their performance has been evaluated. Information is also provided on open questions and research directions in this area.

Keywords: Food Waste, Food Classification, State-of-the-Art, Computer Vision, Convolutional Neural Networks, Object Detection, Performance Evaluation, ResNet-50, Inception-V3

Introduction

Food waste is emerging as an issue of growing interest and discussion, given the economic, environmental, social, and nutritional implications. It is a global problem, manifested by the considerable loss of food throughout the supply chain (Gustavsson *et al*., 2011). This phenomenon, which is particularly evident at the retail and final consumption stages (Correia *et al*., 2022), is the result of the behavior adopted by retailers and consumers, leading to a substantial reduction in edible food mass (Storup, 2016). The magnitude of this challenge is alarming, considering that approximately one-third of food intended for human consumption, equivalent to around 1.3 billion tons per year, is lost or wasted on a global scale (Gustavsson *et al*., 2011). The significance of addressing this matter is unquestionable, underscoring the importance of target 12.3 of the Sustainable Development Goals (SDGs). This target seeks to halve per capita food waste by 2030, encompassing retailers and consumers, as well as production and supply chains (Kateřina and Adriana, 2023).

The work presented in this study focuses on institutional canteens, where food waste covers both prepared meals that have not been sold (i.e., leftovers) and food that remains on plates after the meal has been consumed (i.e., scraps). Nonetheless, even though the main focus is on institutional canteens, all sectors of food waste are explored, including "Households," "Retail and Distribution," "Restaurants and Food Services," "Food Production," and "Primary Production." This approach allows us to understand ongoing efforts and practices in these sectors and to gain insights into the current landscape of food waste management. Additionally, it is of particular interest to further explore what is being done in computer vision across these sectors, as it will provide valuable knowledge for application in institutional canteens.

This study serves as the basis for the future development of a viable prototype/solution that can contribute to reducing food waste in an institutional canteen. It will incorporate computer vision techniques, low-cost IoT components, and cloud computing.

Thus, this study presents a state-of-the-art analysis of current knowledge about food waste in the final stages of consumption through the analysis of similar or related published work. It aims to provide a comprehensive overview of what has been done in related fields. It

DOI: 10.3844/jcssp.2025.851.868

studies, identifies, and evaluates computer vision techniques and datasets that will make it possible to create a system capable of recognizing and analyzing the most wasted food in an institutional canteen. This system will have the potential to support process optimization and promote conscious practices. Furthermore, it may assist in the decision-making process, contributing to more efficient management and a consequent reduction in food waste. This study also raises open questions and research directions for the field.

Food Waste

Food waste manifests itself as a decrease in edible food mass at the end of the food supply chain, predominantly at the retail and final consumption stages. It is largely influenced by the behavior of retailers and consumers. Approximately one-third of the food produced for human consumption, which equates to around 1.3 billion tons per year, is lost or wasted worldwide (Gustavsson et al., 2011). The importance of reducing food waste is clear and defined as a target in the SDGs in goal 12, target 12.3: "By 2030, halve per capita food waste globally, at both retailer and consumer levels and reduce food waste along production and supply chains, including post-harvest" (Kateřina and Adriana, 2023).

Minimizing food loss and waste is crucial for economic and environmental sustainability. It has a direct impact on food security, nutrition, and several SDGs. For example, reducing food waste affects the SDGs on hunger, the environment, poverty, economic growth, and inequality (Food and Agriculture Organization of the United Nations, 2019). Progress on other SDGs, such as gender equality and clean energy, can, in turn, help reduce food waste (Food and Agriculture Organization of the United Nations, 2019). Figure (1) illustrates the interconnection between reducing food waste and various SDGs, highlighting the wider implications for sustainability and human well-being.

Food Waste in Numbers

According to Eurostat (2023) statistics and as can be seen in Fig. (2), in 2021, around 131 kilograms (kg) of food waste per inhabitant were generated in the European Union (EU). Restaurants and catering services, which include institutional canteens, were responsible for 12 kg of food waste per person (9%) (Union, 2023b).

According to the Instituto Nacional de Estatística (2022), food waste in Portugal reached 183.5 kg per inhabitant in 2020, corresponding to 1.9 million tons. Restaurants and similar services were responsible for 13% of this waste, as can be seen in Fig. (3).

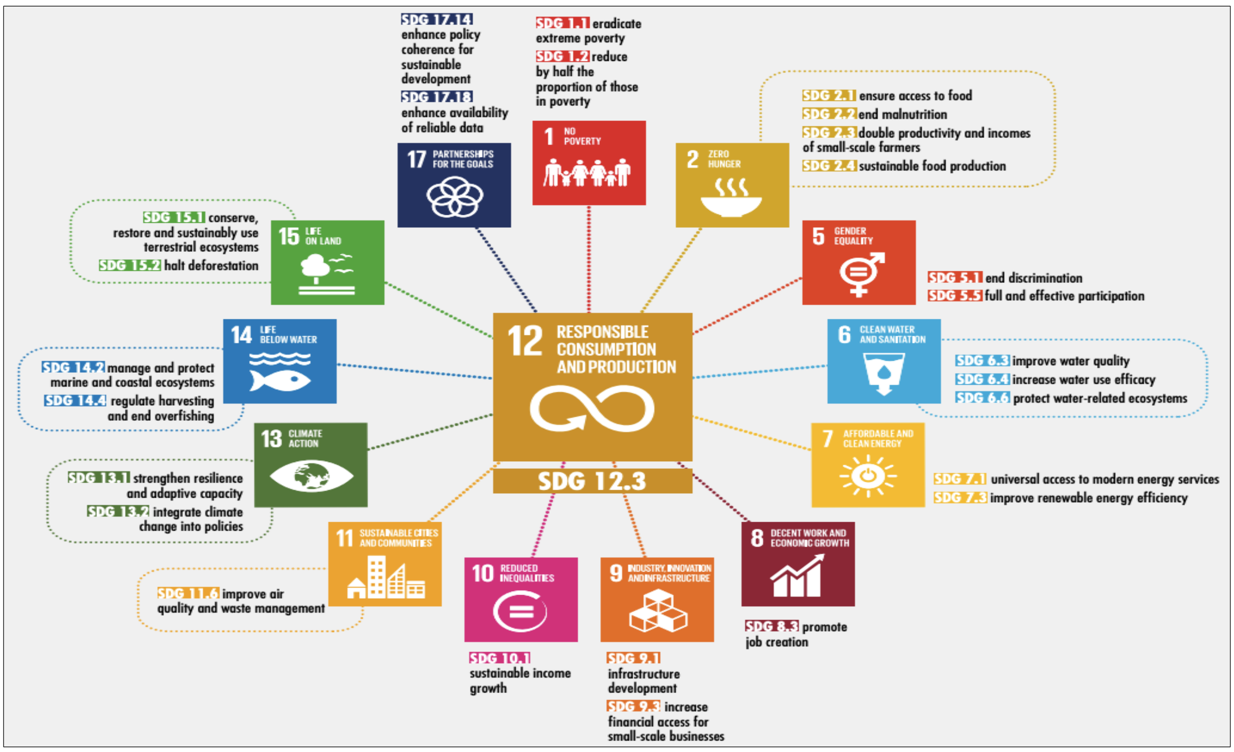

Fig. 1: Food loss and waste and the sustainable development goals. Source: Food and Agriculture Organization of the United Nations (2019)

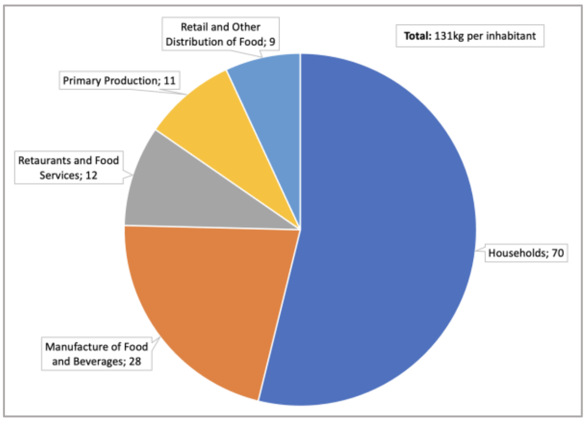

Fig. 2: Food waste in kg per inhabitant in the European Union by main economic sectors. Adapted from: Union (2023a)

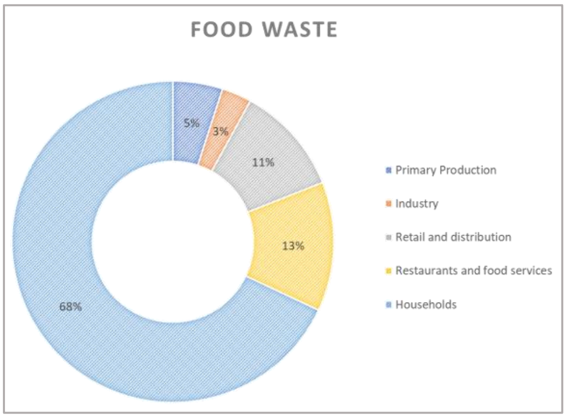

Fig. 3: Food waste generated in 2020 in Portugal. Adapted from Correia *et al*. (2022)

Cost of Food Waste

The Food and Agricultural Organization of the United Nations (FAO) and the World Bank are warning about the economic impact of food waste (Gustavsson *et al*., 2011). The FAO states (Gustavsson *et al.*, 2011) that the food intended for human consumption that is lost or wasted every year translates into a cost of 680 billion dollars in developed countries and 310 billion dollars in developing countries. At the same time, the World Bank estimates the global economic impact of food loss and waste to be around 940 billion dollars worldwide (Gustavsson *et al.*, 2011). In the EU, according to an estimate in a 2016 report (Stenmarck *et al*., 2016), the cost of food waste in 2012 was 143 billion euros.

The Fig. (4) shows the costs of food waste by sector per year. These numbers underline the profound financial impact of inefficiencies along the food supply chain, affecting economies on a global scale. Reducing food waste is, therefore, not only an environmental imperative but also a crucial strategy for minimizing economic losses, promoting financial sustainability, and strengthening global economic resilience (Gustavsson *et al*., 2011).

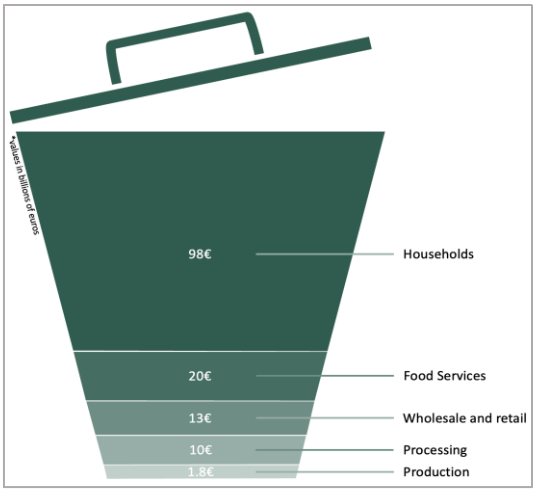

Fig. 4: Cost of food waste in billions of euros by sector per year. Adapted from: Urry (2023)

How to Reduce/Prevent Food Waste

Throughout the food supply chain, there are various forms of food waste. Figure (5) details these different forms.

In order to reduce and prevent food waste, the EU has established important initiatives, including the EU platform on food losses and food waste (Commission, 2023a) and the EU food loss and waste prevention hub (Commission, 2023b), which serve as facilitating channels for the effective sharing of good practices, resources, and knowledge. There was also funding available to boost concrete food waste prevention actions, such as €2,250,000 in support for the hotel and restaurant sector in 2022. This funding was intended to improve the measurement of food waste and help implement food waste prevention measures in the operations of the organizations, as can be seen in the initiative entitled "grants for stakeholders to improve measurement of food waste and help implement food waste prevention in their operations and organizations" (Health and Agency, 2024). Member States are encouraged to run consumer campaigns, integrate food waste prevention into school curricula, and facilitate food donations through legislative measures. The citizens' panel (Commission, 2023c) recommendations serve as a guide, and the document provides quick tips for individuals to reduce food waste. By implementing these measures, the EU aims to empower citizens to actively participate in reducing food waste and promote significant changes in consumption habits (Commission, 2023c).

Mitigating food waste requires a comprehensive approach, integrating various measures and strategies. Examples of crucial initiatives include information, awareness, and communication campaigns aimed at educating the population about efficient food storage, preparation, and use practices. The implementation of educational programs proves to be effective in promoting more conscious household practices, highlighting the importance of portion control and the use of smaller plates (Wansink and van Ittersum, 2013).

DOI: 10.3844/jcssp.2025.851.868

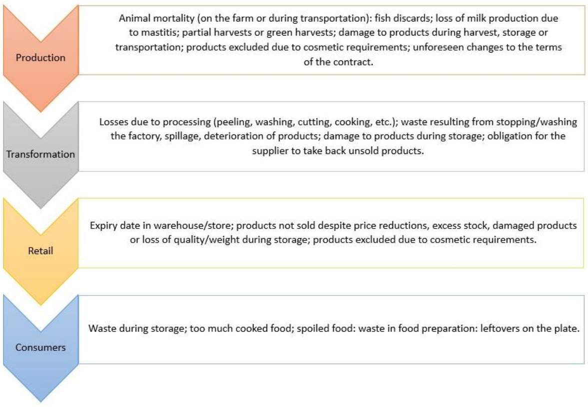

Fig. 5: Factors that result in food waste and losses along the food chain. Adapted from: Storup (2016)

The theory of social influence has emerged as a valuable tool, as evidenced by the positive impact, for example, of messages in university canteens (Stöckli *et al.*, 2018). Psychosocial approaches are essential for changing behaviors related to food waste. Promoting training and qualifications, combined with boosting innovation and technological development, appears to be a promising strategy (Gabinete de Planeamento Politicas e Administração Geral, "National Strategy and Action Plan to Combat Food Waste, 2017).

Therefore, the prevention and effective reduction of food waste requires cooperation between sectors, the education of society, and the integration of innovative technologies, all of which play interdependent roles in building a more sustainable future (Stöckli *et al*., 2018; Gabinete de Planeamento Politicas e Administração Geral, "National Strategy and Action Plan to Combat Food Waste, 2017).

Technologies to Fight Food Waste

Information and communication technologies have an important contribution to make in the fight against food waste. Technology plays a central role in addressing this complex challenge, as evidenced by the essential collaboration between food service companies and technology providers. Technological advances, including applications and data provided by these partners, provide valuable information to food service companies, allowing for more efficient waste management (Martin-Rios *et al*., 2020).

The potential of Artificial Intelligence (AI) to fight food waste and promote a more efficient circular economy is well recognized (Onyeaka *et al.*, 2023). The application of AI can trigger more efficient processes, provide for better informed decision-making, and promote innovative solutions to the challenges facing the global food system. Monitoring and optimizing food production (Sebastian *et al.*, 2023), redistributing surpluses to those in need (Onyeaka *et al.*, 2023), and supporting waste reduction efforts (Fang *et al.*, 2023) are specific areas where AI can play a crucial role.

In addition, the importance of technology in reducing food waste also manifests itself in improving food production and increasing efficiency, controlling food quality, and automating activities such as inventory management, order fulfillment, and delivery. Technologies such as Machine Learning (ML) make it possible to identify trends (Merdas and Mousa, 2023), personalize menus (Naik, 2020), optimize packaging and storage (Wang *et al*., 2023) and detect food safety risks (Wang *et al*., 2022).

The implementation of automated sorting and grading systems that use image recognition to evaluate fruits and vegetables is indeed a promising strategy for reducing food waste. In the context of image recognition, Convolutional Neural Networks (CNNs) can be used to automate the process of inspecting the quality of products such as cereals, fruits, and vegetables, enabling the rapid and accurate identification of defects and quality problems, thus reducing waste caused by human errors and recalls. In addition, the use of CNNs in image recognition systems helps to optimize the efficiency of the supply chain, directing products to appropriate destinations based on their condition, thus preventing unsuitable products from reaching consumers. Therefore, the integration of CNNs into automated sorting and grading systems using image recognition technology aligns with the strategy of reducing food waste by detecting and classifying defects, ultimately also contributing to extending the shelf life and improving the quality of agricultural products (Onyeaka *et al*., 2023).

The application of ML is not just limited to detecting defects and contamination but also to analyzing customer feedback. This analysis makes it possible to assess the impact of various packaging and preservation techniques, providing companies with valuable information to optimize their processes and, consequently, reduce food waste (Onyeaka *et al*., 2023).

Another significant contribution of ML lies in the optimization of supply chain management. ML, including computer vision applications, enables more efficient inventory management, helping companies to reduce operating costs while improving the efficiency of logistics processes (Praveen *et al*., 2020). Computer vision can be harnessed to develop automated food recognition systems capable of identifying food-related objects and ingredients (Pandey *et al.*, 2023).

In addition to these advances, specific technologies are highlighted in the fight against food waste, such as the automatic classification of food waste. For example, Kitro (2024) uses ML to accurately quantify and categorize food waste. Winnow (2024) has developed a smart meter technology connected to food waste containers, which enables efficient measurement and tracking of waste in commercial kitchens. Orbisk (2024) uses computer vision and an AI data logging terminal to monitor and analyze food waste in commercial kitchens, especially in food-to-order establishments. In addition, some solutions integrate data network connectivity with waste disposal machines (Martin-Rios *et al*., 2020), such as smart containers that can communicate with other devices or systems, like the devices used by waste management workers, to provide information on the status of the containers (Czekała *et al*., 2023). This integration allows for real-time monitoring and data sharing, contributing to more efficient waste management practices.

Thus, the combination of these technologies allows for a comprehensive and innovative approach to minimizing food waste, from production to waste disposal, with growing relevance in promoting sustainable practices in food management (Martin-Rios *et al*., 2020; Onyeaka *et al.*, 2023; Pandey *et al*., 2023).

State of the Art

This section presents state of the art from previous research, analyzing scientific articles that contain information related to the technologies and techniques used for detecting and accounting for food waste and for recognizing and classifying food in images.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology was used to review the scientific articles. According to (Prisma, 2023), this methodology focuses mainly on reporting reviews that evaluate the effects of interventions, but it can also be used as a basis for reporting systematic reviews with objectives other than evaluating interventions.

Approaches to Food Waste in Different Sectors

By analyzing Fig. (3), it was decided that the food waste sectors to be analyzed in this document would be "Households", "Retail and distribution", "Restaurants and food services", and "Food production". Figure (3) also shows another sector, "Primary production". However, after carrying out the research adapted to each sector, it was decided not to include it. After applying all the filters, only one article remained that was not aligned with the parameters we had been looking for.

It should be noted that the search strategy was conducted using a query based on research questions, considering the keywords related to the topic of the work and the technologies that are expected to be used. The research questions that should be answered are:

1. How does artificial intelligence contribute to reducing food waste in various sectors such as retail, food distribution, and restaurants?
2. What roles do convolutional neural networks and deep learning play in food waste reduction?
3. How effective is computer vision in food recognition and classification for minimizing waste in food production and services?

Query (main):

"Food waste", and ("Food recognition" or "Food segmentation" or "Food classification") and ("Artificial intelligence", or "Convolutional neural networks" or "Deep learning",) and "Computer vision" and "Machine learning"

Next, given that the search aimed to include various sectors of food waste, keywords related to each sector were added to the main query. In total, four queries were used. The search terms added were "retail"," "food distribution"," "food production"," "restaurants"," "food services"," "canteen"," "households", and "house"." In order to filter out the relevant studies, inclusion and exclusion criteria were defined for the state-of-the-art articles, which are shown in Table (1).

The articles selected for a detailed analysis were chosen based on their title, abstracts, and conclusions. As expected, only articles that address pertinent topics related to the subject of this study were selected. The database used to carry out the research was B-On (B-On, 2023), as its use for queries offers several advantages, including access to reliable resources, constant updates, ease of navigation, and advanced search tools. This research was carried out in November 2023. Figure (6) shows the flowchart describing the various stages of searching for the articles to be studied as part of this study.

Sectors of Food Waste

This subsection describes the various studies that were identified in the previous stage and organized by sector.

Households

In Konstantakopoulos *et al*. (2024), a literature review study is presented, with an exhaustive evaluation of the methods and techniques applied to segment food images, classify their food content, and calculate volume. The study mentions datasets of food images that were used to evaluate automatic food recognition methods, including Food 101 (Bossard *et al*., 2014), UEC-Food 100 (Matsuda and Yanai, 2023), VIREO Food-172 (Chen and NGO, 2023) and UEC-Food 256 (Kawano and Yanai, 2023). The methods studied were categorized into three groups:

(i) Semi-automatic and automatic food image segmentation methods
(ii) Methods based on ML and traditional Learning and based on ML and Deep Learning (DL) for food image classification
(iii) Food volume estimation methods use 3D reconstruction, pre-built shape models, perspective transformation, depth camera, and ML and DL methods. These methods were evaluated in terms of performance, and their strengths and limitations were analyzed. The study concluded that both CNNs and

Deep Convolutional Neural Networks (DCNNs) have been widely used in food image recognition studies. And that the most widely used CNN models have been built specifically for each problem based on the Inception-V3 and deep food models

In Kumar *et al.* (2022) an innovative model was presented for detecting and classifying fresh and damaged fruit using ML and DL techniques. The aim was to develop an intelligent fridge with as few sensors as possible to help reduce food waste, in this case, fruit. To this end, the use of CNNs to recognize images of fresh and damaged fruit was considered. YOLOv3, Faster R-CNN, and SSD were studied. In addition, the pre-trained models Inception-V3 and VGG16 were used to improve classification accuracy. The models were trained on a large dataset of fruit images. The accuracies for distinguishing between fresh and damaged fruit were evaluated. The authors concluded that Faster R-CNN outperformed YOLOv3 and SSD. The results showed that the Faster R-CNN model can correctly differentiate between fresh and damaged fruit. The Accuracy of this model was evaluated on a test dataset, and a mean Average Precision (mAP) score of 78.9% was achieved.

Table 1: Criteria for inclusion and exclusion of articles

Inclusion criteria	Exclusion criteria
Extracted from the B-On database	Published in years prior to 2020. From websites and opinion articles
Scientific in nature and validated by other researchers	
Uses computer vision techniques to classify and detect food in images	

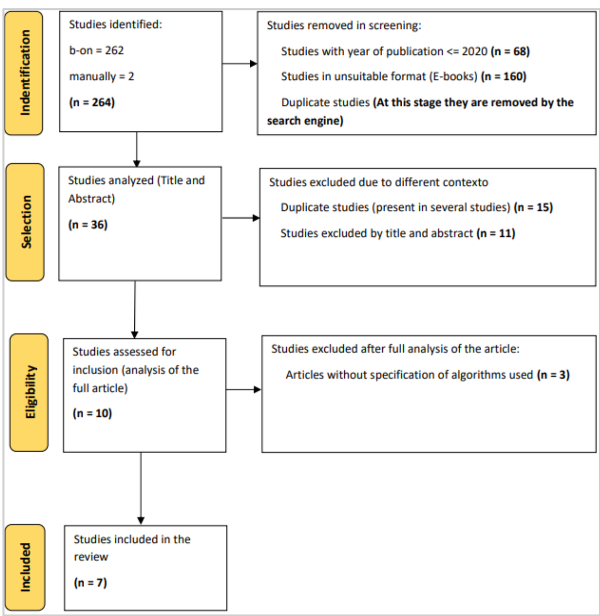

Fig. 6: Flowchart of the research phases

Retail and Distribution

In Hosseinnia Shavaki and Ebrahimi Ghahnavieh (2023), a systematic literature review was carried out with the aim of investigating the application of DL models in operations and Supply Chain Management (SCM). The study covers 43 articles and presents the problems, the DL models adopted, and discusses the open points. It presents a list of DL models applied in the field of SCM and concludes that CNNs are the most widely used. CNN architectures have been used to solve SCM problems such as work forecasting, stock optimization, price forecasting, and fraud detection, among others.

Restaurants and Food Services

In Lubura *et al.* (2022), a CNN model was developed for recognizing and estimating food waste. The model was trained with 157 different food categories and achieved high Accuracy (over 98%) in classifying food images. For food recognition, two datasets were used: The UEC Food 100 (Matsuda and Yanai, 2023), with around 15,000 images divided into 100 classes, and a proprietary dataset with images of the most common foods on the Serbian market, with a total of 23,552 images and 157 classes. The proposed model was built using Keras (2023), an Application Programming Interface (API) for DL methods that use the Python programming language (Python, 2023). This model contains two convolutional layers, a fully connected layer and an output layer with 157 neurons. The CNN model showed good predictive capabilities, obtaining an accuracy of 0.988 and a loss of 0.102 after the network's training cycle. It was estimated that the average food waste per meal for Serbian students was 21.3%.

In Zahisham *et al.* (2020), a learning model was proposed using a DCNN to correctly distinguish foods and recognize them in different orientations. To this end, the ResNet50 model was trained with three datasets: ETHZ Food 101 (Bossard *et al.*, 2014), UEC-Food 100 (Matsuda and Yanai, 2023), and UEC-Food 256 (Kawano and Yanai, 2023). The results of the three trained models showed that high Accuracy was achieved in food recognition: 91.5% in ETHZ-FOOD101, 87.5% in UEC-Food100, and 84.4% in UEC-Food 256.

The study (Huang *et al.*, 2021) focused on accurately and efficiently estimating the Carbon/Nitrogen (C/N) ratio of the organic fraction of municipal solid waste, a crucial factor in the context of automated composting control. To this end, the Mask R-CNN model used was trained with the organic waste-3 dataset (Sekar, 2023). This model is an extension of Faster R-CNN, also based on CNN architectures. To obtain the initial weights, the model was pre-trained with the COCO dataset (COCO, 2024). This is a large-scale dataset made up of 91 common object classes. The results of the study demonstrate the effectiveness of the Mask R-CNN model, which was tested on three different types of organic waste: Lettuce, steamed rice, and bananas. Regression analyses revealed strong linear correlations between the ground truth and the measured volumes of banana (R_2 = 0.985), lettuce (R_2 = 0.955), and rice (R_2 = 0.970).

Food Production

The study presented by (Sood and Singh, 2021) focused on the challenges of limited food production, declining quality, waste, and loss of food products in the field of food production and agriculture. The authors presented an analysis of statistical and computer vision approaches used in food production and agriculture. They found that DL-based approaches produce better results, specifically for image processing applications. They provided a list of publicly available datasets and models used in related studies. They concluded that the datasets Food-101 (Bossard *et al*., 2014), UEC-Food256 (Kawano and Yanai, 2023), and UEC Food 100 (Matsuda and Yanai, 2023) are the most used. The models mentioned in the study were not tested for performance.

Table (2) presents a summary of the studies described above, in what is considered the most relevant aspects: Year of publication, architectures, models, and datasets used.

Critical Analysis of Results

The analysis of the researched papers allows for a conclusion on the suitability and potential of the different types of AI models, as well as on the datasets considered. As can be seen in Fig. (7), the neural network architecture most used in the studies was CNN, with a percentage of use of 62%. DCNN had a percentage of 38%. Although there are distinctions between CNNs and DCNNs, both can be within the same domain of architecture. CNNs represent a specific DL model widely used in image classification (Carvalho, 2023). DCNNs, while belonging to the above, are composed of multiple, fully connected layers and are commonly used to learn complex representations of input data, as well as speech recognition, natural language processing, and classification of data organized in tables (Chaudhari *et al*., 2023).

An analysis of the datasets used in the articles studied shows that seven different datasets are referenced, as illustrated in Fig. (8). The datasets referred to in the articles are Food-101 (Bossard *et al*., 2014), UEC-Food 256 (Kawano and Yanai, 2023), UEC Food 100 (Matsuda and Yanai, 2023), OrganicWaste-3 (Sekar, 2023), COCO. (2024) and VIREO Food-172 (Chen and NGO, 2023). In the analysis carried out, it was noted that one of the studies did not provide information on the dataset used and, therefore, appears in the graph as Not Applicable (N/A). Some of the studies have built and used their own dataset. Figure (8) shows that the most used dataset was UEC-Food 100. On consulting the documentation (Matsuda and Yanai, 2023), it was noted that this is a dataset composed of images of mostly Japanese food.

Table 2: Summary of the scientific articles studied

Reference	Article/Study	Year of publication	Architecture (s)	Model (s)	Dataset (s)
Konstantakopoul *et al.* (2023)	A review of image-based food recognition and volume estimation artificial intelligence systems	2023	CNN, DCNN	The model proposed by the authors based on Inception V3, deep food	Food-101, UEC Food 100, VIREO Food 172, UEC Food 256
Kumar *et al*. (2022)	A novel model to detect	2022	CNN	Faster RCNN	Own
	Classify fresh and damaged fruits to reduce food waste using a deep-learning technique			Incept ionV3, VGG16	
Hosseinnia Shavaki and Ebrahimi Ghahnavieh (2023)	Applications of deep Learning into supply chain management: A systematic literature review and a framework for future research	2023	CNN	N/A	N/A
Lubura *et al.* (2022)	Food recognition and food waste estimation using convolutional neural network	2022	CNN	The model proposed by the authors	Own, UEC Food100
Zahisham *et al.* (2020)	Food recognition with ResNet50	2020	DCNN	ResNet-50	Food-101, UEC Food 100, UEC Food 256
Huang *et al.* (2021)	Method for C/N ratio estimation using Mask RCNN and a depth camera for the organic fraction of municipal solid wastes	2021	CNN	Mask RCNN	Organic Waste-3, COCO
Sood and Singh (2021)	Computer Vision and ML-based Approaches for Food Security: A review	2021	DCNN	N/A	Food 101, UEC Food 100, UEC Food 256

DOI: 10.3844/jcssp.2025.851.868

The 2nd and 3rd most studied datasets were Food-101 and UEC-Food 256. UEC-Food 256 is also composed of images mainly of Japanese food, as it is a similar dataset to the previous one, differing only in the number of classes it contains. Food-101 stands out, with 101 different classes of food from a wide variety of cultures, Japanese food is not predominant.

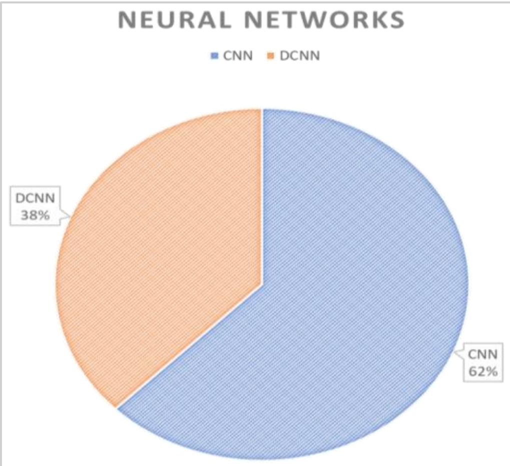

Fig. 7: Neural network architectures used in the articles

Fig. 8: Datasets used in state-of-the-art articles

Fig. 9: Computer vision models used in state-of-the-art articles

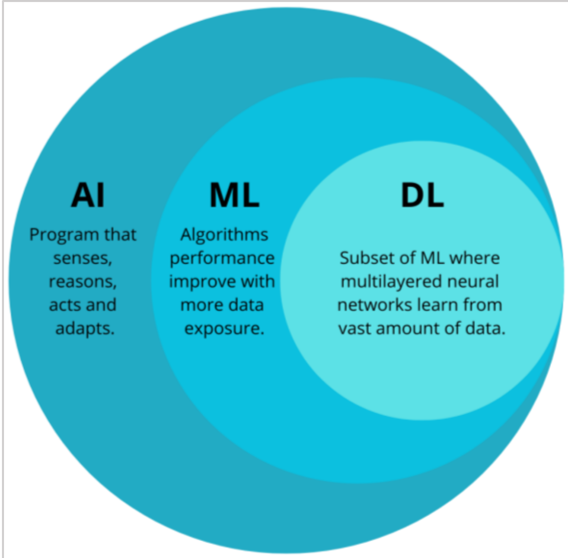

Fig. 10: AI, ML, and DL hierarchy. Adapted from: Alzubaidi *et al.* (2021)

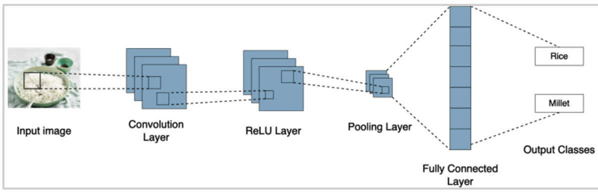

Fig. 11: Typical architecture of a CNN. Adapted from: Géron (2022)

Likewise, an analysis was carried out to see which CNN and DCNN models were considered in the studies analyzed in the state-of-the-art. Figure (9) shows the seven models mentioned in the different articles, namely: Faster R-CNN, Inception V3, VGG16, Deep Food, ResNet-50 and Mask R-CNN. In addition to these, one of the articles did not clarify which model was used. There are also other studies that have developed their own model based on existing ones. They are represented in the graph as "Model Proposed". Figure (9) allows us to conclude that the most widely used model was Inception-V3, which was developed for computer vision tasks, particularly for classifying objects in images.

Ultimately, there are articles on the subject that are the focus of this study. However, several of these articles lack substantial results or omit details about the models adopted. Regarding the articles that have been analyzed, there are various types of models and architectures that can be applied in the context of classifying and detecting food in images. Thus, it can be concluded that InceptionV3 is the most widely used model. So, its performance should be studied and evaluated. The same applies to the ResNet-50 model described in (Zahisham *et al.*, 2020), which is

DOI: 10.3844/jcssp.2025.851.868

relevant to the "Restaurants and food services" sector. Thus, the performance of these models will be evaluated in the context of this study to conclude on their suitability for the task of classifying food in images.

Computer Vision Techniques

Computer vision techniques refer to the methods and algorithms used to enable machines to interpret and understand visual data from the world around them (Shorten and Khoshgoftaar, 2019). These techniques include image processing, pattern recognition, ML, DL, and AI. They are used to analyze and extract information from images and videos, such as object detection, segmentation, tracking, and recognition (Shorten and Khoshgoftaar, 2019).

AI is a vast domain that encompasses many subdomains, including ML and DL, as illustrated in Fig. (10). ML is a subset of AI that focuses on algorithms that can learn and make predictions based on data. DL is a subset of ML that uses neural networks with many layers to learn complex representations of data. One of the most popular types of neural networks used in DL is CNNs (Alzubaidi *et al.*, 2021). This section aims to provide an understanding of some of the main concepts behind CNNs and of two specific CNN models, Inception-V3 and ResNet-50, which have been identified as the most prominent for the task of classifying food in images.

Convolutional Neural Networks

CNNs are a type of artificial neural network designed for processing grid-structured data such as images. They have been proven to be highly effective in computer vision tasks such as image classification, object detection, and image segmentation (Géron, 2022).

Models based on CNNs are usually made up of convolution layers, pooling layers, and fully connected layers (Géron, 2022), which represent the presence of various features in the input image. Figure (11) illustrates an example of a CNN architecture for image classification. It starts with an input image that passes to the convolutional layer, which plays a crucial role in extracting features from images (Alzubaidi *et al.*, 2021). This layer has a set of filters or kernels that can be trained. These are combined with the input image to generate feature maps. Each filter detects specific patterns in the image, such as edges, textures, or shapes. The output of the convolutional layer consists of a collection of feature maps (Maurício *et al*., 2023).

The Rectified Linear Unit (ReLU) layer introduces non-linearity into the CNN architecture by applying the rectification function to the output of the previous layer. This non-linearity is crucial for the network to learn and perform more complex tasks. ReLU is an activation function commonly used in CNN architectures due to its simplicity and effectiveness in training deep neural networks (Alzubaidi *et al.*, 2021).

The pooling layer is responsible for reducing the sample size of the feature maps generated by the convolutional layers (Alzubaidi *et al.*, 2021). It reduces the spatial dimensions of the feature maps while maintaining the essential information (Maurício *et al.*, 2023). Various pooling methods can be used, such as max pooling and average pooling (Alzubaidi *et al.*, 2021). Pooling helps to control overfitting, reduce computational complexity, and maintain the dominant information in the feature maps (Maurício *et al*., 2023).

The fully connected layer is responsible for creating high-level abstractions and the final classification in a CNN. It receives mid- and low-level features from the previous layers and connects each neuron to each neuron in the previous layer (Maurício *et al*., 2023). The flattened output of the previous layers is fed into these fully connected layers, allowing the network to learn complex relationships between features and make predictions (Alzubaidi *et al.*, 2021). The fully connected final layer produces an output that represents the predicted class probabilities for the input image, allowing the classification of the learned features into different classes (Maurício *et al*., 2023).

Examples of CNN models are listed in Shorten and Khoshgoftaar (2019); Alzubaidi *et al.* (2021); Géron (2022) and include Alex Net (Krizhevsky *et al*., 2017), Network in Network (NIN) (Lin *et al*., 2013), Zf Net (Zeiler and Fergus, 2014), Visual Geometry Group (VGG) (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy *et al.*, 2015) with its versions Inception V1, V2, V3 (Szegedy *et al*., 2016), V4 (Szegedy *et al*., 2017), Residual Network (Res Net) (He *et al*., 2016), Densely Connected Convolutional Network (Dense Net) (Huang *et al*., 2017), Xception (Chollet, 2017), Squeeze-and-Excitation Network (SE Net) (Hu *et al*., 2018), ResNeXt (Xie *et al*., 2017), MobileNet (Howard, 2017), Cross Stage Partial Network (CSPNet) (Wang *et al*., 2020) and EfficientNet (Tan and Le, 2019).

Choosing the most suitable CNN model for a task involves careful consideration of several key factors (Alzubaidi *et al.*, 2021). Firstly, it is necessary to assess the specific requirements of the task, such as image classification or object detection (Alzubaidi *et al.*, 2021). Evaluating the performance of different models on reference datasets should consider factors such as model size, Accuracy, and speed on both the Central Process Unit (CPU) and the Graphics Processing Unit (GPU) (Maurício *et al*., 2023). It is important to consider the available computing resources, as some models can be computationally demanding (Géron, 2022). In addition, depending on the complexity of the task and the data set available, it is necessary to consider the depth (number of layers) and width (number of neurons in each layer) of the CNN architecture (Alzubaidi *et al.*, 2021). Regularization techniques and optimization methods should be chosen

DOI: 10.3844/jcssp.2025.851.868

based on the dataset and architecture. The potential benefits of transfer learning, where pre-trained CNN models are adjusted for the specific task at hand, should be evaluated (Géron, 2022). Finally, consider the interpretability of CNN models, especially if the task requires understanding the model's decision-making process (Alzubaidi *et al.*, 2021; Géron, 2022).

Based on the articles analyzed in section 3, it was concluded that the Inception-V3 (GoogLeNet) and ResNet-50 models would be the most promising to apply in the context of this study. The paper (Hosseinnia Shavaki and Ebrahimi Ghahnavieh, 2023) supports this assertion, pointing out that CNNs have shown significant success in identifying various types of food and estimating their nutritional values. CNN models such as VGG, GoogLeNet, and ResNet have been effectively applied to food image recognition.

Inception-V3 and ResNet-50

Inception-V3, which evolved from Inception-V1 and was introduced by GoogLeNet in 2014 (Géron, 2022), is a model that is part of the Inception family of CNNs. This model has a depth of 48 layers, exhibiting an error rate of 3.57% in image classification tasks (Alzubaidi *et al.*, 2021). The input size for the images is 229×229×3. This means that the images have a resolution of 229 pixels wide by 229 pixels high and three-color channels (RGB) (Alzubaidi *et al.*, 2021). This model was designed to address the challenges of efficiency, scalability, performance, and resource constraints in the context of CNNs for computer vision, focusing on parameter reduction through factorized convolutions, regularization, and batch normalization (Szegedy *et al*., 2016).

A distinctive feature of Inception networks is the repetition of blocks throughout the artificial neural network. This consists of stacking Inception modules, with each module containing several repeated blocks. These blocks have the function of extracting characteristics from the input images, contributing to the network's effectiveness in classifying images (Andrew and Santoso, 2022).

Figure (12) shows an example of an Inception-V3 model. It starts with factored convolutions, which are used to reduce the number of parameters in the network. This involves splitting traditional convolutions into smaller convolutions. For example, the traditional 7'7 convolution is factorized into three 3'3 convolutions (Szegedy *et al.*, 2016). Pooling layers are applied to reduce the sampling of the feature maps and their dimensionality (Andrew and Santoso, 2022). Inception modules are designed to efficiently capture features at various scales and complexities. These modules are composed of parallel branches that incorporate convolutions with various filter sizes, pooling operations, and the factorization of convolutions to process feature maps efficiently. This design highlights the reduction in the size of the grids between the Inception modules while maintaining the dimensions of the feature maps. The concatenation of filters in the modules makes it possible to combine features extracted from different convolutional branches, enriching the representation of the features (Szegedy *et al*., 2016). The fully connected layers play a crucial role in the final classification phase of the model. After passing through the last Inception module, the output is subjected to global average pooling to calculate the spatial average of the feature maps (Andrew and Santoso, 2022).

In Fig. (12), the fully connected layers are composed of a flattened layer, which is responsible for converting the output of the previous layer into a one-dimensional tensor, preparing it to be processed by the dense layer. The dense layer applies a linear transformation to the input data. Each neuron in this layer is connected to all the neurons in the previous layer, ensuring a total connection. Finally, the softmax activation function, often used in the output layer for multiclass classification, normalizes the original outputs into a distribution of probabilities in different classes (Shazia *et al*., 2021).

The ResNet models (He *et al*., 2016), developed in 2015, are known for their ability to train very deep neural networks using skip connections, also known as shortcut connections (Géron, 2022). ResNet-50, a variant of the ResNet model, has a depth of 50 layers (Andrew and Santoso, 2022). It has an error rate of 6.71% (Top-5 Error Rate) and accepts images with an input size of 224'224 (He *et al*., 2016).

The components of this model include residual blocks. Each residual block consists of convolutional layers, batch normalization, activation functions (e.g., ReLU), and jump links (He *et al*., 2016). The convolutional layers, which include 1'1, 3'3, and 1'1 convolutions, are responsible for extracting features from the input data at various spatial scales and depths within the network (He *et al*., 2016). Batch normalization is used to normalize the activations of each layer, which helps stabilize and accelerate the formation of deep neural networks (He *et al*., 2016). ReLU activation functions are used to allow the network to learn complex representations from input data (He *et al*., 2016). The jump connections allow the gradient (variation of the adjustment to be applied to each weight, with a view to minimizing the network's error (Ferreira *et al*., 2022) to fluctuate more efficiently during training, providing alternative paths for the propagation of the gradient (He *et al*., 2016). ResNet-50 also uses pooling and fully connected layers to achieve accurate classification in image recognition tasks (Andrew and Santoso, 2022). Figure (13) shows an example of a ResNet-50 model.

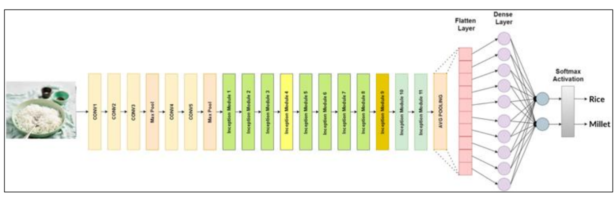

Fig. 12: Diagram of the Inception V3 model. Adapted from Shazia *et al*. (2021)

DOI: 10.3844/jcssp.2025.851.868

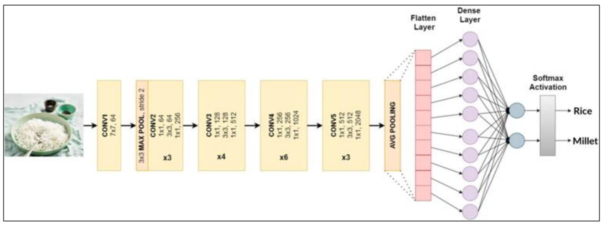

Fig. 13: Diagram of the ResNet-50 model. Adapted from Shazia *et al*. (2021)

Performance Evaluation

The research carried out in this project addresses the challenge of contributing to reducing food waste in an institutional canteen. So, it will be relevant to detect food served that is left on plates after the meal has been consumed (i.e., scraps). The findings from the related work presented above highlight Inception-V3 and ResNet-50 as the most promising models for classifying and detecting food in images. Therefore, it is interesting to compare their performance. In this section, the dataset chosen for use in the tests is described first. Then, the scenarios for implementing the models are presented, as well as the performance metrics. Finally, the results of the tests carried out are discussed.

Dataset

As the dataset has an extensive list of classes, just a few random images of the classes are exemplified. The Fig. (14) shows some of the images in the dataset duly captioned with the class to which they belong.

Based on the conclusions from the state-of-the-art review, the most suitable dataset for testing the models is Food-101 (Bossard *et al*., 2014). This public dataset is made up of food images organized into 101 classes of food types. Each class contains 1,000 images, making a total of 101,000 images (Bossard *et al*., 2014). The images in this dataset have already been properly classified. In addition to the images, the dataset folder provides information in text files about the labels and the classes to which they belong. Table (3) lists the 101 classes that make up the dataset.

The images for training and testing are divided immediately after downloading the dataset in both trained models. 75% of the images were used for training, and the remaining 25% for testing. Dividing the dataset, as previously mentioned, leaves a total of 75750 images for training and the remaining 25250 for testing.

Performance Metrics

The performance metrics applied to evaluate the models in the tasks of classifying food in images were Accuracy and loss. Each model was trained for a specific number of epochs. The decision to stop training the models is made taking several factors into account. When it is detected that the model's performance on the set of test images is no longer improving or is getting worse, training should be stopped. This approach avoids overfitting the model on the dataset and ensures good performance on new images (Prechelt, 1998). Both models had their training interrupted when they showed signs of stagnation in the performance metrics. ResNet-50 was trained for 42 epochs, while Inception-V3 was trained for 32 epochs.

Accuracy represents the proportion of correct predictions in relation to the total number of examples (Developers, 2024). The formula for calculating Accuracy is shown in Eq. (1). Accuracy can be calculated in both the training and testing phases. When it is calculated during training, it means that the model has correctly classified a percentage of the predictions in the training set. While test accuracy, i.e., the Accuracy calculated on the test set, means that the model is correctly classifying a percentage of the images in the dataset (Lehn, 2024):

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ of\ predictions} \tag{1}$$

Loss is an ML model performance metric that can be used on both training and test data. When it is calculated during the training process, it reflects how the model is adapting to the dataset. It should decrease as the epochs progress. However, a very low loss does not necessarily mean that the model will perform well on new data, as it may have adjusted too much to the training data. In the classification models whose performance was evaluated, the loss function used to calculate the loss value was the cross-entropy loss (also known as logarithmic loss or logistic loss) (Vijay, 2024). Equation (2) shows the formula for calculating the cross-entropy loss:

$$Loss = -\frac{1}{Output\ Size}\sum_{i=1}^{Output\ size} y_i * \log \hat{y_i} + (1 - y_i) * \log (1 - \hat{y_i}) \tag{2}$$

Fig. 14: Examples of images from 15 classes of the food-101 dataset

The loss calculated during training means that the model, on average, has an error relative to the training data. Whereas the loss calculated during testing, i.e., when the model was tested with images from the test set, on average made an error relative to that data.

Implementation Scenario

To evaluate the performance of the ResNet-50 model, a notebook available on GitHub (Herick-Asmani, 2023) with Python code was used to download the Food-101 dataset and to train that model. Google Colab platform (Google, 2023) was used to train the model. This platform offers free hardware resources, such as GPUs and Sensory Processing Units (TPU), support for various languages, and integration with Google Drive and GitHub. The machine provided by the platform in the free plan has the following characteristics: NVIDIA T4 graphics card with 16 GB of VRAM and 13 GB of RAM for the system.

The trial environment was run on a machine with an Intel Core i7-1165G7 processor, 16.0 GB of RAM, and integrated Intel Iris Xe Graphics. To carry out the trials, code was implemented in Python, using PyTorch libraries (PyTorch, 2023), just like the notebook used to train the ResNet-50 model.

This trial code makes it possible to use the trained model file extracted from the notebook. To do this, the corresponding checkpoint file must be provided. In addition, an image must be included as an input for classification, from which the code generates an output. The output consists of a string representing the name of the class that the model assigns to the image, together with the hit probability.

The evaluation process for the Inception-V3 model followed a similar methodology. A notebook is available on GitHub (Kappa, 2024) was used, containing Python code that allowed downloading the Food-101 dataset and training the pre-existing implementation of the Inception-V3 model. The platform used was once again Google Colab (Google, 2023), and the technical characteristics for training this model were described above.

The trial for the Inception-V3 model was also carried out on Google Colab, taking advantage of the existing implementation on the notebook. The Tensor Flow library (Campesato, 2019) was used to train the model. The best model was then extracted after training to make predictions, and two images were used as input for testing. The code returns the image duly identified with the class that the model predicted.

Results and Discussion

The process of training the ResNet-50 model took a long time, and it was difficult to meet the time limit for using the free resources provided. Given that the maximum usage time is 4-6 h, training had to be carried out in several phases. Each epoch took approximately 45 min to train, and the model was trained for 42 epochs. Whenever the execution time ran out, it was necessary to save the file of the model trained so far and start training again from that point when resources were made available again.

Figure (15) shows a training loss of 0.5944, which means that during training, the model had an average error of 59.44%. The training accuracy shows that the model made 83.77% of the predictions in the training set. The test loss indicated that the model was, on average, making an error of 0.5303. Finally, the test accuracy value was 0.8627, which means that the model correctly classifies approximately 86.27% of the images in this dataset.

Figure (16) shows a graph comparing training loss and test loss. It can be concluded that over the course of training the model, there was a positive trend in the training loss and test loss curves. As the epochs increased, there was a consistent reduction in both training loss and test loss. This shows that the model is learning effectively on the training images and, similarly, on the test images.

As training was carried out in phases, for the reasons explained above, this means that in some epochs, it is necessary to use the model trained in the previous epoch and start training again from there. In Fig. (16), it can be observed that between epochs 17 and 18, there was a sharp drop in both metrics. This happens for the reason explained above: The model trained in epoch 17 ended up with a training loss of 2.4298, and when it was restarted in the following epoch, it dropped to 1.6715.

Table 3: Food-101 dataset classes

Food-101 dataset classes		
Apple-pie baby-back-ribs baklava beef-carpaccio beef-tartare beet-salad	edamame eggs-benedict escargots falafel filet-mignon fish-and-chips	Omelette onion rings oysters' pad-thai paella pancakes
Beignets bibimbap bread pudding breakfast-burrito bruschetta Caesar-salad cannoli caprese-salad carrot cake ceviche cheese plate cheesecake chicken curry chicken quesadilla chicken wings chocolate cake chocolate mousse churros clam-chowder club-sandwich crab-cakes crème-brulee croque-madame cupcakes deviled eggs donuts dumplings	Foie-gras french-fries French-onion-soup French-toast fried-calamari fried-rice frozen-yogurt garlic-bread gnocchi Greek-salad grilled-cheese-sandwich grilled salmon guacamole gyoza hamburger hot-and-sour-soup hot-dog huevos-rancheros hummus ice-cream lasagna lobster-bisque lobster-roll-sandwich macaroni-and-cheese macarons miso-soup mussel's nachos	Panna-cotta Peking-duck pho pizza pork-chop poutine prime-rib pulled-pork-sandwich ramen ravioli red-velvet-cake risotto samosa sashimi scallops seaweed-salad shrimp-and-grits spaghetti-bolognese spaghetti-carbonara spring rolls steak strawberry-shortcake sushi tacos takoyaki tiramisu tuna-tartare waffles

DOI: 10.3844/jcssp.2025.851.868

Fig. 15: Results of the metrics in epoch 42 of the ResNet-50 model

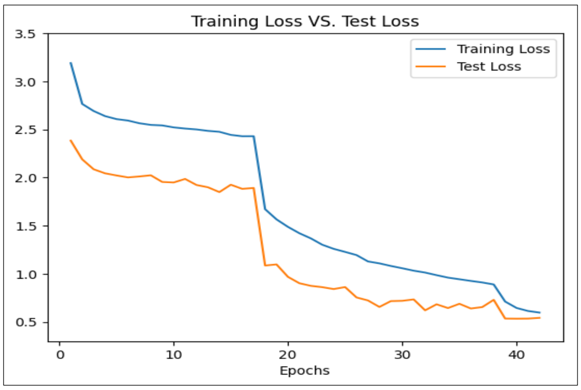

Fig. 16: Results of the ResNet-50 training process for the training loss and test loss metrics

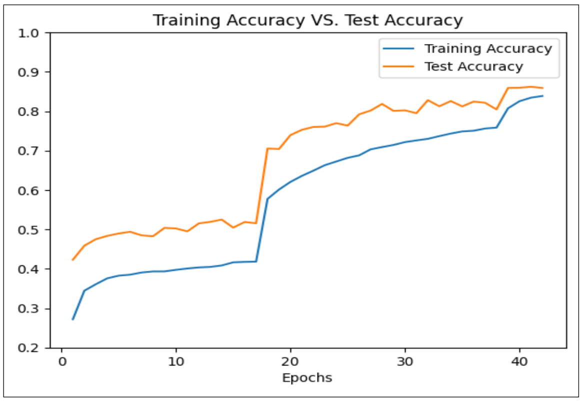

Fig. 17: Results of the ResNet-50 training process for the training accuracy and test accuracy metrics

It can also be seen from Figs. (16-17) from epoch 40 onwards, the decrease in loss and the increase in Accuracy were not significant. For this reason, it was decided to finish training the model in epoch 42. In this way, it is possible to prevent the model from overfitting, i.e., its tendency to adapt too much to the training images.

Figure (17) shows a graph comparing training accuracy and test accuracy. Over the epochs, there is a consistent evolution in both training accuracy and test accuracy. This indicates a constant improvement in the model's performance with the training data, as well as a corresponding ability to efficiently apply the knowledge acquired to new test data. This positive trend suggests robust and progressive Learning, contributing to confidence in the model's ability to make accurate predictions. As in the previous Fig. (16), the same thing happens with the accuracy values in epoch 17. The values improved substantially in the following epoch because the training was done in stages.

In addition to presenting the training results, trials were carried out to confirm the model's effectiveness. Figure (18) shows an example of a test carried out on an image of risotto, which is one of the classes included in the dataset. The trained ResNet-50 model correctly identified it with a percentage of 99.76%.

Figure (19) shows another example of a trial carried out on a pasta image, which is not a class included in the dataset. The result was "spaghetti_bolognese", with a hit percentage of 61.98%. This means that it didn't get the classification completely right since the dataset doesn't contain this class.

In the Inception-V3 model, the training process also took a long time, around 30 min per epoch. There was also the challenge of using the platform for a period (as with the ResNet-50 model). The resources are free, but they are limited in the time they can be used. As the dataset used has 101,000 images, both models take time to train, so on average, it is only possible to train 10 epochs per time of use.

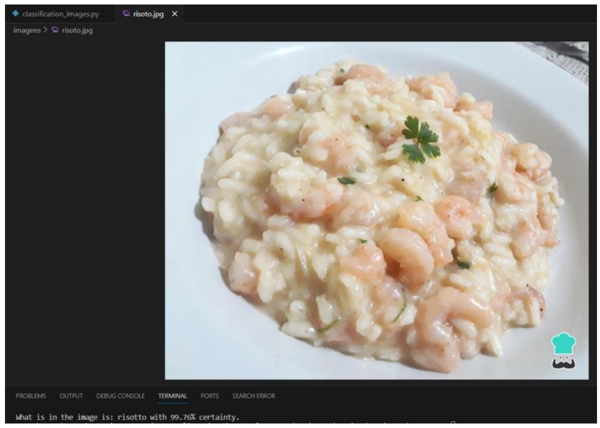

Fig. 18: Image of risotto (class included in the dataset) classified with Resnet-50

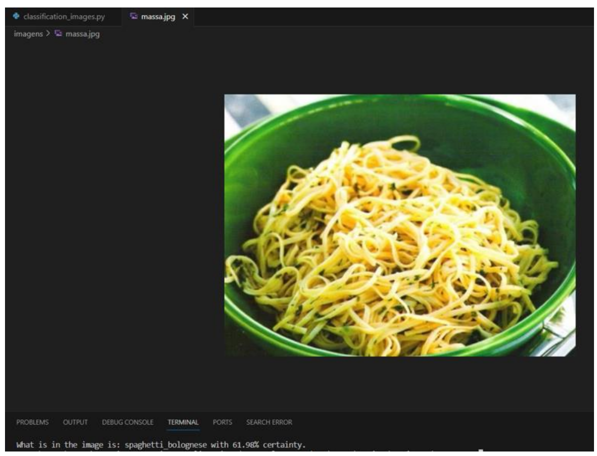

Fig. 19: Image of pasta (class not included in the dataset) classified with Resnet-50

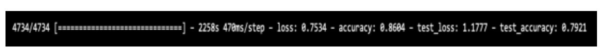

Fig. 20: Results of the metrics in epoch 28 of the InceptionV3 model

DOI: 10.3844/jcssp.2025.851.868

The Inception-V3 model was trained for 32 epochs. However, the model with the best results was the one corresponding to epoch 28. The results shown in Fig. (20) reveal a training loss of 0.7534, indicating that, on average, the model had an error of 75.34% during training. The training accuracy metric revealed that the model was correct in approximately 86.04% of the predictions in the training set. In terms of test loss, the model showed an average error of 1.1777 when tested with images from the test set. Finally, test accuracy revealed an accuracy of 79.21%, which indicates that the model correctly classifies around 79.21% of the images in the test set.

The Figs. (21-22) show graphs comparing the loss in the training and testing processes and the Accuracy in the same processes. The graph in Fig. (21) shows that there are variations in test loss between some epochs, while training loss shows a more stable curve. This indicates that as the model was trained, the loss decreased steadily. It can be seen in Figs. (21-22), test loss did not decrease, test accuracy did not increase, and test accuracy did not change significantly from epoch 28 onwards. Therefore, the model was no longer trained to avoid overfitting.

Regarding accuracy in both the training and testing process, Fig. (22) shows a steady increase in training accuracy. However, there are visible variations in the evolution of test accuracy. These variations may be because training must be carried out in phases. As a result, the loss in the following epoch may vary from the previous epoch.

Trials were also conducted to confirm the effectiveness of Inception-V3, using the same images used to assess the previous model. Figure (23) shows an example of a test on an image of risotto, which is one of the classes included in the dataset. Inception-V3 predicted the correct class.

Fig. 21: Results of the inception-V3 training process for the training loss and test loss metrics

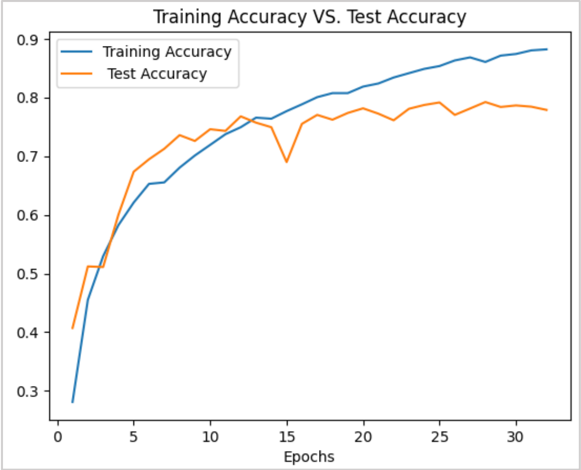

Fig. 22: Results of the inception-V3 training process for the training accuracy and test accuracy metrics

Fig. 23: Image of risotto (class included in the dataset) classified with inception-V3

Fig. 24: Image of pasta (class not included in the dataset) classified with inception-V3

In conclusion, it can be said that both models correctly identified the foods in the images whose classes exist in the dataset. In classes that are not present in the dataset, the predictions were different. The most suitable model to use in future work is the one with the highest test accuracy. The experiments that were conducted showed that the best results were obtained using the ResNet-50 model.

In the example shown in Fig. (24), an image was submitted as an input whose food should be identified as pasta. However, as with the previous model, the prediction was wrong. The prediction made by the Inception-V3 is the class "seaweed-salad". A possible reason for the model not classifying the image correctly is that the pasta class does not exist in the dataset, so it would not be possible to identify it.

Conclusion

Combating food waste is not just a choice; it is a moral obligation and a race against the clock, where every meal saved contributes to overcoming significant challenges with profound economic, environmental, and social implications. Recognizing the urgency of addressing these issues is crucial, especially given the considerable loss of food along the supply chain, accentuated by the behaviors adopted by retailers and consumers. The use of information technologies, such as computer vision and artificial intelligence, shows promise in identifying and analyzing patterns that can direct strategies to effectively reduce these losses.

The main contributions of this study are:

1) A clear definition of the problem related to food waste and an analysis of how computer vision technologies can help to minimize this issue
2) A review of the state-of-the-art and related works
3) The identification of the most promising computer vision techniques, namely Inception-V3 and ResNet-50, together with the Food-101 dataset
4) A performance evaluation to compare Inception-V3 and ResNet-50

The performance evaluation of the Inception-V3 and ResNet-50 models revealed several challenges that will have to be addressed in future work. One of the most important is the inadequacy of the dataset used in the context of Portuguese gastronomy. This obstacle is particularly relevant since the application scenario is an institutional canteen, where the dishes served are typical of Portuguese meals and simpler than the food categories found in the dataset used. This could compromise the effectiveness of the classification model. Therefore, it will be necessary to improve and complement the dataset or even create a new one. It may also be necessary to evaluate other CNN models using that dataset.

In the scope of future work, it is planned to develop a prototype with a targeted Technology Readiness Level (TRL) of 3-5 (TRL 3-experimental proof of concept, TRL 5-technology validated in a relevant environment, industrially relevant in the case of key enabling technologies). This solution can be expanded and implemented in various institutional canteen settings, addressing food waste challenges in these environments.

During the implementation of the prototype, other challenges may arise related to adapting computer vision techniques to the specific characteristics of the food and environments found in an institutional canteen. These particularities may include variations in lighting, the arrangement of food on trays, variations in the presentation style of dishes, and other specific characteristics of the canteen environment. Other challenges that may arise in the future implementation of the prototype concern the hardware to be used to ensure that the proposed solution is viable, effective, and efficient. Integrating the computer vision system with other existing technologies or processes in the canteen is also a challenge. It would be desirable to have a cohesive and harmonious solution that can be easily incorporated into the operational routine. In addition, other challenges are issues such as the speed of capturing the photos in the process of depositing the tray after consuming the meal, the resolution of the images, and the decision about computing in the cloud or on specialized local hardware.

Funding Information

João M. L. P. Caldeira and Vasco N. G. J. Soares: Acknowledge that this study is funded by FCT/MECI through national funds and when applicable co-funded EU funds under UID/50008: Instituto de Telecomunicações.

Author's Contributions

Ana Correia and Clara Aidos: Conceptualization, methodology, investigation, written original draft preparation.

João M. L. P. Caldeira and Vasco N. G. J. Soares: Test, formal analysis, writing-review and editing, funding acquisition.

All authors have read and agreed to the published version of the manuscript.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Andrew, A., & Santoso, H. (2022). Compare VGG19, ResNet50, Inception-V3 for Review Food Rating. *SinkrOn*, *6*(2), 845–494. https://doi.org/10.33395/sinkron.v7i2.11383

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *Journal of Big Data*, *8*(1), 1–74. https://doi.org/10.1186/s40537-021-00444-8

b-on. (2023). Biblioteca do Conhecimento Online. *Accessed*. https://www.b-on.pt/

Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101–Mining Discriminative Components with Random Forests. *Computer Vision – ECCV 2014*, *8694*, 446–461. https://doi.org/10.1007/978-3-319-10599-4-29

COCO. (2024). *Common Objects in Context*. https://cocodataset.org/#home

Campesato, O. (2019). *Appendix: TF 2, Keras, and Advanced Topics*. 169–224. https://doi.org/10.1515/9781683924616-007

Carvalho, D. (2023). *Deep Learning: CNN e aplicações*. https://carvalhodouglas1999.medium.com/deep

Chaudhari, M. S., Majhi, S., & Jain, S. (2023). CNN-Attention-DNN Design for CFO Estimation of Non-Pilot-Assisted OFDM System. *IEEE Communications Letters*, *27*(2), 551–555. https://doi.org/10.1109/lcomm.2022.3228732

Chen, J., & Ngo, C.-W. (2023). Vireo-Food 172 Dataset - Deepbased Ingredient Recognition for Cooking Recipe Retrieval. *In Proceedings of the 24th ACM International Conference on Multimedia*, 32–41. https://doi.org/10.1145/2964284.2964315

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1800–1807). https://doi.org/10.1109/cvpr.2017.195

Commission, E. (2023a). European Food Loss and Waste Prevention Hub. *Accessed*. https://ec.europa.eu/food/safety/food_waste/eufood-loss-waste-prevention-hub/

Commission, E. (2023b). *European Citizens' Panel on Food Waste*.

Commission, E. (2023c). *Frequently Asked Questions: Reducing Food Waste in the EU*.

Czekała, W., Drozdowski, J., & Łabiak, P. (2023). Modern Technologies for Waste Management: A Review. *Applied Sciences*, *13*(15), 8847–9045. https://doi.org/10.3390/app13158847

Developers, G. (2024). *Classificação: Acurácia | Machine Learning*. https://developers.google.com/machine

Fang, B., Yu, J., Chen, Zhonghao, Osman, Ahmed I, Farghali, M., Ihara, Ikko, Hamza, Essam H, Rooney, David W, & Yap, Pow-Seng. (2023). Artificial Intelligence for Waste Management in Smart Cities: A Review. *Environmental Chemistry Letters*, *21*(4), 1959–1989. https://doi.org/10.1007/s10311-023-01604-3

Ferreira, M., Monteiro Pereira, R. M., & Pereira, A. J. C. (2022). Techno-Economic Assessment of Hydrogen Production Using Solar Energy. *Renewable Energy and Power Quality Journal*, *20*, 751–756. https://doi.org/10.24084/repqj20.422

Food and Agriculture Organization of the United Nations. (2019). *The State of Food and Agriculture 2019*. https://doi.org/10.18356/32f21f8c-en

Gabinete de Planeamento Politicas e Administração Geral, "National Strategy and Action Plan to Combat Food Waste. (2017).

Géron, A. (2022). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*.

Google. (2023). *Google Colab*. https://pytorch.org/docs/stable/index.html

Gustavsson, J., Cederberg, C., Sonesson, U., Otterdijk, R. V., & Meybeck, A. (2011). Global Food Losses and Food Waste: Extent, Causes and Prevention, Rome. *Food and Agriculture Organisation of the United Nations*, *365*(1554).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778). https://doi.org/10.1109/cvpr.2016.90

Health, E., & Agency, D. E. (2024). Fighting Food Waste in the EU: New Call for Proposals to Help Stakeholders Take Action. *Accessed*. https://hadea.ec.europa.eu/news/fighting-food-waste-eunew-call-proposals-help-stakeholders-take-action-2022

Herick-Asmani. (2023). *Food-101-classification-using-ResNet50/Food101challenge.ipynb," Github*. https://github.com/HerickAsmani/Food-101-classification-using-ResNet-50/blob/master/Food101challenge.ipynb

Hosseinnia Shavaki, F., & Ebrahimi Ghahnavieh, A. (2023). Applications of Deep Learning into Supply Chain Management: A Systematic Literature Review and a Framework for Future Research. *Artificial Intelligence Review*, *56*(5), 4447–4489. https://doi.org/10.1007/s10462-022-10289-z

Howard, G. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *Apr*.

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2018). Squeezeand-Excitation Networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141.

Huang, J., Zhang, H., Xiao, X., Huang, J., Xie, J., Zhang, Liang, Hu, H., Dai, S., Li, M., & Xu, Y. (2021). Method for C/n Ratio Estimation Using Mask r-Cnn and a Depth Camera for Organic Fraction of Municipal Solid Wastes. *International Journal of Agricultural and Biological Engineering*, *14*(5), 222–229. https://doi.org/10.25165/j.ijabe.20211405.6382

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708. https://doi.org/10.1109/cvpr.2017.243

Instituto Nacional de Estatística. (2022). *"Desperdício Alimentar - Resultados e Perspetivas*.

Kitro. (2024). Simplifying Food Waste Management. *Accessed*. https://www.kitro.ch/

Kappa. (2024). *Food_Classifier.ipynb - Colaboratory*. https://colab.research.google.com/github/theimgclist/examples/blob/MultiClassTF2.0/community/en/multi-class-cla ssification/food-classifier.ipynb#scrollTo=JBs1U7hZkp1 U

Kateřina, M., & Adriana, Š. (2023). 'Ensure Sustainable Consumption and Production Patterns.' *Oxford Academic*, 948–859. https://doi.org/10.1093/law/9780192885173.003.0013

Kawano, Y., & Yanai, K. (2023). *FoodCam - 256*.

Keras. (2023). *Deep Learning for humans*. https://keras.io/

Konstantakopoulos, F. S., Georga, E. I., & Fotiadis, D. I. (2024). A Review of Image-Based Food Recognition and Volume Estimation Artificial Intelligence Systems. *IEEE Reviews in Biomedical Engineering*, *17*, 136–152. https://doi.org/10.1109/rbme.2023.3283149

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kumar, T. B., Prashar, D., Vaidya, G., Kumar, V., Kumar, S. D., & Sammy, F. (2022). A Novel Model to Detect and Classify Fresh and Damaged Fruits to Reduce Food Waste Using a Deep Learning Technique. *Journal of Food Quality*, *2022*, 1–8. https://doi.org/10.1155/2022/4661108

Lehn, F. (2024). Interpreting Training/Validation Accuracy and Loss. *Medium*. https://medium.com/@frederik.vl/interpretingtraining-validation-accuracy-and-loss-cf16f0d5329f

Lin, M., Chen, Q., & Yan, S. (2013). Network in Network. *International Conference on Learning Representations*, 1–10.

Lubura, J., Pezo, L., Sandu, M. A., Voronova, V., Donsì, F., Žlabur, Jana Šic, Ribić, B., Peter, A., Šurić, J., Brandić, I., Klõga, M., Ostojić, S., Pataro, G., Virsta, A., Daraban, O., Micić, D., Đurović, S., Feo, G. D., Procentese, A., & Voća, N. (2022). Food Recognition and Food Waste Estimation Using Convolutional Neural Network. *Electronics*, *11*(22), 3746–3836. https://doi.org/https://doi.org/10.3390/electronics11223746

Martin-Rios, C., Hofmann, A., & Mackenzie, N. (2020). Sustainability-Oriented Innovations in Food Waste Management Technology. *Sustainability*, *13*(1), 210–436. https://doi.org/10.3390/su13010210

Matsuda, Y., & Yanai, K. (2023). Dataset Food - 100. *Accessed*. http://foodcam.mobi/dataset100.html

Maurício, J., Domingues, I., & Bernardino, J. (2023). Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. *Applied Sciences*, *13*(9), 5521–5821. https://doi.org/10.3390/app13095521

Merdas, H. M., & Mousa, A. H. (2023). Food Sales Prediction Model Using Machine Learning Techniques. *International Journal of Electrical and Computer Engineering (IJECE)*, *13*(6), 6578–6585. https://doi.org/10.11591/ijece.v13i6.pp6578-6585

Naik, P. A. (2020). Intelligent Food Recommendation System Using Machine Learning. *Nutrition*, *5*(8).

Onyeaka, H., Tamasiga, P., Nwauzoma, U. M., Miri, T., Juliet, U. C., Nwaiwu, O., & Akinsemolu, A. A. (2023). Using Artificial Intelligence to Tackle Food Waste and Enhance the Circular Economy: Maximising Resource Efficiency and Minimising Environmental Impact: A Review. *Sustainability*, *15*(13), 10482–12849. https://doi.org/10.3390/su151310482

Orbisk. (2024). Automatically Reduce Food Waste in Your Kitchen. *Accessed*. https://orbisk.com/

Python. Org., (2023). *Welcome to Python.org*. https://www.python.org/

Pandey, V. K., Srivastava, S., Dash, K. K., Singh, R., Mukarram, S. A., Kovács, B., & Harsányi, E. (2023). Machine Learning Algorithms and Fundamentals as Emerging Safety Tools in Preservation of Fruits and Vegetables: A Review. *Processes*, *11*(6), 1720–1862. https://doi.org/10.3390/pr11061720

Praveen, K B., Kumar, P., J, P., G, P., & J, M. (2020). Inventory Management using Machine Learning. *International Journal of Engineering Research and Technology*, *9*(06), 866–869. https://doi.org/10.17577/ijertv9is060661

Prechelt, L. (1998). Early Stopping - But When? *Neural Networks: Tricks of the Trade*, *1524*, 69–55. https://doi.org/10.1007/3-540-49430-8_3

Prisma. (2023). Preferred Reporting Items for Systematic Reviews and Meta-Analyses. *Accessed*. http://www.prismastatement.org/?AspxAutoDetectCookieSupport=1

PyTorch. (2023). *PyTorch 2.1 Documentation*. https://pytorch.org/docs/stable/index.html

Sebastian, K., Balakrishnan, D., Koshy, S. M., Sara Sabu, S., & Amrutha, P. (2023). Application of Artificial Intelligence in Horticulture: A Review. *Ecology, Environment and Conservation*, *29*(Suppl), 259–268. https://doi.org/10.53550/eec.2023.v29i04s.041

Sekar, S. (2023). *Waste Classification Data*.

Shazia, A., Xuan, T. Z., Chuah, J. H., Usman, J., Qian, P., & Lai, K. W. (2021). A Comparative Study of Multiple Neural Network for Detection of COVID-19 on Chest X-Ray. *EURASIP Journal on Advances in Signal Processing*, *2021*(1), 1–16. https://doi.org/10.1186/s13634-021-00755-1

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. In *Journal of Big Data* (Vol. 6, Issue 1, pp. 1–48). https://doi.org/10.1186/s40537-019-0197-0

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science and Engineering*.

Sood, S., & Singh, H. (2021). Computer Vision and Machine Learning based approaches for Food Security: A Review. *Multimedia Tools and Applications*, *80*(18), 27973–27999. https://doi.org/10.1007/s11042-021-11036-2

Stenmarck, Å., Jensen, C. M., Quested, T., & Moates, G. (2016). *Estimates of European Food Waste Levels*. https://doi.org/10.13140/RG.2.1.4658.4721

Stöckli, S., Niklaus, E., & Dorn, M. (2018). Call for Testing Interventions to Prevent Consumer Food Waste. *Resources, Conservation and Recycling*, *136*, 445–462. https://doi.org/10.1016/j.resconrec.2018.03.029

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1), 2818–2826. https://doi.org/10.1609/aaai.v31i1.11231

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9). https://doi.org/10.1109/cvpr.2015.7298594

Storup, K. (2016). *Relatório Especial - Luta Contra o Desperdício Alimentar: Uma Oportunidade Para a UE Melhorar a Eficiência Dos Recursos na cadeia de abastecimento alimentar*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826. https://doi.org/10.1109/cvpr.2016.308

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scasling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning*, 6105–6114.

Union, E. E. (2023a). Eurostat. *Accessed*. https://ec.europa.eu/eurostat

Urry, D. (2023). The Cost of Food Waste. *Earth First. Accessed*. https://www.foodunfolded.com/article/the-cost-of-foodwaste

Union, E. E. (2023b). Food Waste and food Waste Prevention - Estimates. *Accessed*. https://ec.europa.eu/eurostat/statisticsexplained/index

Vijay, M. (2024). What is the Difference Between Training Loss Validation Loss and Evaluation Loss? *Medium*. https://medium.com/@penpencil.blr/what-is-thedifference-between-training-loss-validation-loss-andevaluation-loss-c169ddeccd59

Wang, C.-Y., Mark Liao, H.-Y., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., & Yeh, I.-H. (2020). CSPNet: A New Backbone that can Enhance Learning Capability of CNN. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1571–1580. https://doi.org/10.1109/cvprw50498.2020.00203

Wansink, B., & van Ittersum, K. (2013). Portion Size Me: Plate-Size Induced Consumption Norms and Win-Win Solutions for Reducing Food Intake and Waste. *Journal of Experimental Psychology: Applied*, *19*(4), 320–332. https://doi.org/10.1037/a0035053

Wang, X., Bouzembrak, Y., Lansink, A. O., & van der Fels-Klerx, H. J. (2022). Application of machine learning to the monitoring and prediction of food safety: A review. *Comprehensive Reviews in Food Science and Food Safety*, *21*(1), 416–434. https://doi.org/10.1111/1541-4337.12868

Wang, S.-S., Lin, P., Wang, C.-C., Lin, Y.-C., & Tung, C.-W. (2023). Machine Learning for Predicting Chemical Migration from Food Packaging Materials to Foods. *Food and Chemical Toxicology*, *178*, 113942–113960. https://doi.org/10.1016/j.fct.2023.113942

Winnow. (2024). Commercial Food Waste Solutions. *Accessed*. https://www.winnowsolutions.com/

Xie, S., Girshick, R., Dollar, P., Tu, Z., & He, K. (2017). Aggregated Residual Transformations for Deep Neural Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5987–5995. https://doi.org/10.1109/cvpr.2017.634

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. *Computer Vision – ECCV 2014*, *8689*, 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

Zahisham, Z., Lee, C. P., & Lim, K. M. (2020). Food Recognition with ResNet-50. *020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 1–5. https://doi.org/10.1109/IICAIET49801.2020.9257825