**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** | Capstone Project |
| **Assessment Title:** | *Development and Optimisation of Convolutional Neural Networks (CNNs) to predict the nutrition and sustainability scores of foods from crowd sourced images* |
| **Supervisor Name:** | **David Gonzalez** |
| **Student Full Name:** | Cormac McElhinney |
| **Student Number:** | sbs23102 |
| **Assessment Due Date:** | 23rd Feb .24 |
| **Date of Submission:** |  |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

cormac mcelhinney / Student No.: sbs23102 Word count 3287 (15000 words) excl. supporting content https://moodle.cct.ie/course/view.php?id=2649

**Development and Optimisation of Convolutional Neural Networks (CNNs) to predict the nutrition and sustainability scores of foods from crowd sourced images**

2023

Contents

[Table of Figures 3](#_Toc150289525)

[Abstract 4](#_Toc150289526)

[Introduction 5](#_Toc150289527)

[Materials and Methods 5](#_Toc150289528)

[Results and discussion 21](#_Toc150289529)

[Selection of distributed data processing environment **Error! Bookmark not defined.**](#_Toc150289530)

[EDA 21](#_Toc150289531)

[Sentiment Classification and scoring **Error! Bookmark not defined.**](#_Toc150289532)

[Forecasting 22](#_Toc150289533)

[ARIMA and SARIMA 22](#_Toc150289534)

[Forecasting using gradient boosting: LightGBM 22](#_Toc150289535)

[Recurrent Neural Network (RNN) model and hyperparameter tuning 22](#_Toc150289536)

[Dynamic dashboard development 24](#_Toc150289537)

[Forecasting discussion and associated health warning 27](#_Toc150289538)

[Comparison of Mongo DB and MySQL using YCSB 27](#_Toc150289539)

[Conclusion 28](#_Toc150289540)

[Appendix 28](#_Toc150289541)

[Bibliography 35](#_Toc150289542)

# 

# Table of Figures

[Figure 1 Sentiment Classification Frame work (AlzamzamiHoda and El Saddik, 2020) 22](#_Toc150291542)

[Figure 2 Static graph of the results of an Auto-Sarima (ARIMA) results where 1,4 and 12 weeks forecast are captured. 22](#_Toc150291543)

[Figure 3 Output of a hyper-parameter tuned RNN prediction model, forecasting 1, 4 and 12 weeks in advance, with MAE of 0.084 and MSE of 0.007 24](#_Toc150291544)

[Figure 4 Illustration of a dash created interactive dashboard displaying the output from an ARIMA forecasting task including highlighting the training, testing and predicted sentiment score ranging from 4-12 weeks, including confidence intervals. 26](#_Toc150291545)

[Figure 5 Distribution of number (count) of tweets over time from the selected dataset. 28](#_Toc150291546)

[Figure 6 Distribution of number (count) of tweets over time from the selected dataset. 29](#_Toc150291547)

[Figure 7 Total number of tweets per day of the week. 29](#_Toc150291548)

[Figure 8 Total number of tweets per hour of the day. 30](#_Toc150291549)

[Figure 9 Distribution of user tweets of the top 10 most frequent posters of tweets per 15 min intervals of each hour each day. 31](#_Toc150291550)

[Figure 10 Distribution of sentiment classification (positive, negative and neutral) by week of analysis from tweets in the dataset. 32](#_Toc150291551)

[Figure 11 Distribution of sentiment classification (positive, negative and neutral) by week of analysis from tweets in the dataset. 33](#_Toc150291552)

[Figure 12 Distribution of sentiment classification (positive, negative and neutral) by week of analysis from tweets in the dataset. 33](#_Toc150291553)

[Figure 13 Forecasted sentiment from twitter using an hyperparameter tuned RNN model. 34](#_Toc150291554)

[Figure 14 Performance of database type (MongoDB Vs MySQL) when appraised using YCSB across 6 workloads. Runtime was the only significantly different variable of the 12 variables tested. 34](#_Toc150291555)

# Abstract

This research explores the use of Convolutional Neural Networks (CNNs) and big data processing techniques for the automated classification of food products based on publicly-sourced unprocessed images. With the increasing importance of big data in various domains, including food chains and agriculture, this study addresses the need for efficient and accurate methods for the classification of foods from images. This starts with a comprehensive literature review, examining the current state of the art in artificial neural networks, specifically CNNs, for food product classification based on image data. Simultaneously, a large-scale dataset of food product information was collected from a related online source, namely Open Food Facts. A CNN architecture specialised for food product classification was developed, with a focus on optimising model architecture and hyperparameters. The study also integrates state-of-the-art big data processing and storage techniques into the research workflow, aligning with the trends identified in the literature. The research question addressed is the effectiveness of this integrated approach compared to traditional manual methods for food classification. The study emphasizes the importance of high-quality and extensive datasets, highlighting the challenges in recognizing visually complex and diverse food images. Results show that the developed CNN model achieves a high accuracy rate during training and less so for the validated results, indicating the potential for overfitting. To mitigate this, hyperparameter tuning was conducted, with a focus on learning rate and dropout rate. The findings emphasize the delicate balance between model complexity and efficiency, with various techniques explored to enhance the model's performance.

# Introduction

# Big data (BD) is a critical technology for future research in food chains, agriculture, and other sectors of the economy (Sonka 2014). BD is defined as “a conglomeration of the booming volume of heterogeneous data sets, which is so huge and intricate that processing it becomes difficult, using the existing database management tools” (Subudhi et al. 2019, p.2). It can be understood as the processing and analysis of large data sets obtained from various sources such as online user interactions, consumer-generated content, commercial transactions, sensor devices, monitoring systems or any other consumer tracking tools (Li et al. 2019). Bid data also refers to the massive amounts of digital information about human activities, which are generated by a wide range of high-throughput tools and technologies (Marchetti 2016).

# Traditional and manual methods for assessing the nutritional profiles or ecological impact of foods is labour intensive and can be inaccurate as they are reliant on someone scribing the information from the product in the shop for classification in a system for further analysis (Lohala, Alsadoon et al. 2021). To assess whether a food product conforms with the respective legislation, it must be categories by food type i.e., is the food a cheese or a beverage for example and by sub-type. Recent research by (Ciocca, Napoletano and Schettini 2017) identified how supervised machine-learning systems were proposed to solve a similar problem, based on visual features of food images, to overcome the limitation of traditional food recognition and classification approaches. Machine-learning techniques used in these classifications include Support Vector Machine (SVM), K-Nearest Neighbours (K-NN) and Artificial Neural Network (ANN) (Ciocca, Napoletano and Schettini 2017) . However, the performance of these machine learning techniques is not perfect because they include several automated learning steps to train the model and this would require manual features engineering to predict the actions from the datasets correctly (Lohala, Alsadoon et al. 2021). The latest trend in this domain is the application of Convolutional Neural Network (CNN) techniques that are based on Deep Learning to handle the highly non-linear relationships between the image class and image datasets, making it very suitable to use CNN for food classification based on images (Murphy 2016).

## Research Objectives

1. Investigate the characteristics and challenges of crowd-sourced image data for food nutrition prediction of three food categories.
2. Develop CNN models using crowd sourced data and images (existing images from open source) to predict the nutrition status of three categories of food.
3. Evaluate the performance of the optimised CNN models in predicting nutrition scores for three categories of foods and present it as a tool to improve compliance.

## Technical Objectives

1. Use data preprocessing techniques to clean, standardise, and augment the image dataset and associated data.
2. Systematically fine-tune the CNN model's hyperparameters, such as learning rates, batch sizes, and layer configurations, to optimize its performance using Grid search CV.
3. Incorporate techniques to help interpret and explain why the CNN makes specific nutrition score predictions.

2. Relevance Communication partners (CP) are defined by Kent-Walsh and McNaughton (2005) as people who either compose the AAC user's personal life (family and friends) or retain an education or care nature towards the individual (e.g.: teachers, health care professionals). However, the authors attest that very little attention has been given to the improvement of AAC learning methods for CP intervention programs. A solution focused on improving the learning methods specifically for Lámh language communication partners may be proposed considering the fragility of this field according to the authors presented in the literature review. Such a solution would not extinguish the practicality and the portability that Wilkinson and Hennig (2007) highlight as the main advantages of KWS languages, once the assisting tool would not be introduced in the direct relation between the user and their communication partner. 2 Instead, a data analytics supporting alternative could be applied at the preparation of the communication partners, in a stage of the learning process that Kent-Walsh and McNaughton (2005) classify as "Controlled Practice and Feedback". The latter consists in a controlled environment with only the communication partner, the Lámh user and an instructor. In this context, if a Deep Neural Network model is able to classify Lámh signs in real time, this model can be used in the "Controlled Practice and Feedback" stage, giving the communication partner the possibility to train the learnt signs more often. Hence, the CP acquires confidence and knowledge to be able to use the language in a natural environment, directly with the AAC/Lámh user. Therefore, the proposal of a solution focused on facilitating the learning methods for Lámh communication partners using Deep Learning for sign recognition would promote more democratic, accessible and playful learning opportunities for those involved in the Lámh user’s life. Furthermore, the consistent implementation and use of this communication support can also provide a wide variety of benefits to one's inclusion in society. 3. Contribution The novelty of this work is the application of machine learning functions and tools to classify Lámh language signs. The main contributions of this work are firstly the data acquisition methodology – since there is no existing database of Lámh signs for machine learning classification, after the best Machine Learning models are chosen, they were evaluated in actual real time detection applications – and secondly the comparison of Convolution Neural Network, Long Short Term Memory and Support Vector Machine for 3 the classification of these signs, defining the most appropriate method for the proposed context.

# Literature Review

According to the World Health Organisation (WHO), Worldwide obesity has nearly tripled since 1975, in fact, in 2016, more than 1.9 billion adults, 18 years and older were overweight and of these over 650 million were obese of which 39% of adults aged 18 years and over were overweight and 13% were obese (WHO, 2023). Most of the world's population live in countries where overweight and obesity kills more people than underweight and 39 million children under the age of 5 were overweight or obese in 2020. Obesity is preventable and better communication and education on foods and diets are fundamental in shaping people’s choices and by making the choice of healthier foods and regular physical activity the easiest choice (the choice that is the most accessible, available and affordable) may help prevent obesity (Biermayr-Jenzano, 2019).

The use of computer vision approaches and methods incorporating digital images could prove a useful and powerful real-time communication tool to help inform consumers about the nutrition profile of foods albeit on highly marketed foods. Additionally, food regulators could also use such an approach to ensure compliance with the law. Food information to consumers Regulation 1169/2011 is quite broad in it that it permits a wide range of marketing of food products, however the proposed approach could also be used to identify breaches in such legislation i.e. illegal health or nutrition claims. The approach of using machine vision for classification or prediction has already been used in various fields, such in support of medical diagnosis (SreedhaNair and Maity, 2023, Mahajan et al., 2023), tracking criminals using facial recognition tools (Yang et al., 2022, J and Suresh L, 2023) and using food images to classify products in food processing facilities (Zhu et al., 2021).

### Systematic literature searching methodology

Identifying food in images is thought to be challenging and even difficult because food images often look like each other and sometime don’t have many features that are humanly discernible (Chaitanya Shetty and Chiplunkar, 2023). With the aforementioned context as background, a systematic literature review was conducted using various literature retrieval techniques. The authors performed searches in Science Direct, Pubmed, IEEE and Google scholar databases. The search query for the specific topic with binary logic operators (string with Boolean Combinations) was used by the author and is as follows: ("Classification OR prediction") AND ("Food OR Agri") AND ("Images") AND ("Convolutional Neural Network") AND (“nutrition OR health OR safety OR sustainability”). Google scholar, science direct, IEEE explore, and pub med returned 790, 4386, 117 and 79 results, respectively. These were sorted by relevance and the publications chosen for a more in-depth review can be found in the bibliography.

### Image classification and Recognition

However, food is a challenging problem in image classification due to its visually complex intricacy and the variations in mixing different ingredients of different shapes and sizes in different cuisines (Hassanien, 2019). Food images reportedly lack definitive spatial patterns, often evident in images showing scenes or objects and may also be taken and various distances and resolution. The varying colours, forms, and textures of the ingredients typically define the essence of the cuisine. Therefore, despite the availability of large numbers of applications and algorithms, the task of food recognition necessitates the development of models capable of exploiting local features within images. The most recent food recognition models are based on deep convolutional neural network architectures. However, they remain computationally expensive (Hassanien, 2019).

Image recognition requires more computing power compared to classifying text-based data as a comparison. However, for people to take advantage of food recognition models, they should be able to use them on affordable devices. Nowadays, many of the budget-friendly smartphones with sufficient computing power can manage high-quality image data. Therefore, the models discussed in this paper could in theory be deployed on higher end smartphones (Islam et al., 2018).

A typical approach to sourcing, developing and deploying a CNN model using food image data to classify the nutrition score or class is shown in Fig. 1. This framework includes sourcing of good quality images, preprocessing them before segmentation and feature extraction steps. Following this, dimension reduction is carried out in advance of the classification step. Further prediction can be applied to the nutrition profile of the foods if necessary.

A diagram of food processing

Description automatically generated

Fig 1: Architecture of image-based food-recognition systems for dietary assessment. Source (Dalakleidi et al., 2022)

### Development of CNN models

Within the framework outlined in Fig. 1 several key areas must be considered when developing such a method and these include but are not limited to the following; Hyperparameter tuning, data augmentation, regularisation methods, learning rate schedules and normalisation (Buhl, 2023). There are many approaches used by authors in this field of study who introduce a CNN approach for food classification, object detection, semantic segmentation or using RNN approaches for prediction where images are sources from online sources (Buhl, 2023). For example, Park et al. (2019) assessed their model's accuracy using different pre-trained CNN image models like AlexNet, Resnet, GoogleNet, and K-foodNet and found that KfoodNet demonstrated the highest accuracy.

In a separate study, Lohala et al. (2021b) presented a CNN algorithm aimed at enhancing the accuracy of food image predictions. Their model featured several key components, including a modified loss function and a flexible approach to feature extraction involving the toggling of convolutional layers, pooling layers, and a fully connected layer on top of all layers. The algorithm's performance was evaluated based on two metrics: total execution time for speed assessment and probability scores for correctness evaluation. Notably, the classification accuracy for fast-food images witnessed a notable 5% improvement, and the processing time was reduced significantly.

Liu et al. (2016) designed a CNN by combining two Inception modules with some extra convolutional layers, and they connected them with an additional max pooling layer. This network ended up having twenty-two layers with various settings. They also applied a dropout technique, where 70% of the data was temporarily excluded during training to improve the model's performance. They assessed this approach on different sets of food images. One set called UEC-256 contained 256 categories and a total of 28,375 images. Using their method, they achieved a top-1 accuracy of 54.7% and a top-5 accuracy of 81.5% on this dataset. They also tried it on UEC100, which had 100 categories, and achieved even better results, with a top 1 accuracy of 76.3% and a top five accuracy of 94.6%. Finally, they assessed it on another dataset called Food-101, where they got a top-1 accuracy of 77.4% and a top-5 accuracy of 93.7%. They found that by including bounding boxes in their approach, they were able to improve the top-1 accuracy for the UEC-256 dataset to 63.8%.

### NLP and other feature extraction techniques

Taking a more complex approach to this area of research, Yunus et al. (2018) used Natural language processing (NLP) due to the diverse nature of the data they collected from various online repositories. To guarantee the data quality, these authors carefully selected what they perceived to be trustworthy websites for webscraping and maintaining data integrity. Separately, a cascaded neural network was implemented to enhance the accuracy of the AlexNet pretrained model by over 12% on the Kaggle database called Food 101 (Sun, 2021). This model utilized a more advanced pre-trained network, offering the potential for further improvements, especially when trained on a larger number of classes.

A SIFT method was utilized by (He et al., 2014) for feature extraction using backpropagation neural networks (BPNN) and this out performed the k-dimensional trees approach in terms of accuracy score. This improvement was achieved by automating feature extraction using the CNN. Switching to dataset specifics, one research group used the food-11 dataset on kaggle and achieved an accuracy of 92.86% with the Inception V3 pre-trained CNN model . However, their proposed approach surpassed this, achieving an accuracy of 97.00% for twenty classes and 96.52% for twenty-five classes (Minija and Emmanuel, 2017). Notably, this approach incorporated web scraping to provide essential details about specific food items in addition to the food images. Mezgec and Koroušić Seljak (2017) introduced a groundbreaking architecture, known as deep CNN NutriNet, designed for the identification and recognition of food and beverage images. This architecture underwent training using a dataset containing 225,953 images across 520 different classes of foods and drinks. Remarkably, it achieved an impressive accuracy rate of 87%. A further image classification study conducted in Bengali of food images used the VGG16 model for transfer learning and achieved 98% for both the F1 score and accuracy (Deng et al., 2009). This approach proved to be more practical for real-world applications, particularly when dealing with a larger number of classes. Xiao et al. (2021) proposed a CNN-based food image recognition model where a combined use of a jumping convolution layer and a traditional convolutional layer was used to minimize the calculation parameters. Compared to the experimental outcomes of previous deep learning networks, the suggested method has a beneficial impact and reduced training time needed. Additionally, researchers investigated the influence of colour features on accuracy found that while colour features had limited impact, the proposed approach, featuring a pre-trained model, delivered high accuracy and enriched the analysis by providing detailed information about individual food items through (Akbar et al., 2017).

### Multi-step CNNs

Finally, (Jiang et al., 2020) introduced a multi-step model to identify food images using deep CNN and candidate regions. Initially, they used the Region Proposal Network (RPN) to generate numerous regions from input images. Subsequently, each proposal region was classified into specific food groups, and these regions were then mapped onto a feature map. The model generated a dietary evaluation report, considering factors such as fat content, protein, total calorie count, and carbohydrates. Critical evaluation of the state of the art: In recent years, CNNs used to identify food in pictures has become increasingly popular, following the broader trend of using neural networks (ChaitanyaShetty and Chiplunkar, 2023). These CNNs have proven to be quite skilled at recognising different foods in images, even when those foods look very similar

There is a growing number of studies aimed at making CNNs even better at classifying foods, employing various methods and techniques (Park et al., 2019) and researchers are also leveraging pre-trained models like AlexNet, ResNet, GoogleNet, and K-foodNet to improve food recognition (Park et al., 2019). Some authors are even using language related tricks, like Natural Language Processing (NLP), to make data collection smoother and enhance data reliability and validity, thus advancing the capabilities of CNN-based food recognition systems (Yunus et al., 2018). A constant theme in the research discussed in this literature review is the substantial and consistent improvement in the performance of CNNs for recognising food in images (Lohala et al., 2021a). Techniques like "dropout" and adjusting pre-trained models have emerged as crucial strategies to make these models work even better (Liu et al., 2016).

### Geographical bias and ethnic foods

Most of the research carried out over the past 14 years has generally focused on food images from Asian (Table 1). More recently, the focus of research appears to be using multi-ethnic food images, most likely due to the availability of food images and the accessibility to data analytics tools and computing power. The models described in Table 1 should be considered in the context of bias and limitation when being compared. For example, a dataset of images from typical Chinese food produced in China may not be applicable to use to train a CNN classification model of Chinese food produced for an Irish consumer, in Ireland as the recipes, methods of production and presentation may be different and may result in poor accuracy of the final results when the model is applied in the real world.

**TABLE 1. Publicly available food image datasets used since 2009 as inputs into image-based food recognition systems**

| **Name** | **Year** | **Food categories, *n*** | **Images, *n*** | **Cuisine** | **Reference** |
| --- | --- | --- | --- | --- | --- |
| Pittsburgh Fast-food Image Dataset (PFID) | 2009 | 61 | 1089 | Fast food | (Chen et al., 2009) |
| UEC-Food 100 | 2012 | 100 | 10,000 | Japanese | (MatsudaHoashi and Yanai, 2012) |
| NTU-FOOD | 2012 | 50 | 5000 | Multiethnic | (Chen et al., 2012) |
| UNICT-FD889 | 2014 | 889 | 3583 | Multiethnic | (FarinellaAllegra and Stanco, 2014) |
| Food-101 | 2014 | 101 | 101,000 | Multiethnic | (BossardGuillaumin and Van Gool, 2014) |
| UEC-Food 256 | 2014 | 256 | 31,397 | Multiethnic | (Kawano and Yanai, 2014) |
| Ambient Kitchen | 2014 | 12 | 1800 | Multiethnic | (Pham and Thuy, 2014) |
| UPMC Food-101 | 2015 | 101 | 90,840 | Multiethnic | (Wang et al., 2015) |
| Dishes | 2015 | 3832 | 117,504 | Multiethnic | (HerranzXu and Jiang, 2015) |
| Menu-Match | 2015 | 41 | 646 | Asian, Italian | (Joshi et al., 2017) |
| FooDD | 2015 | 23 | 3000 | Multiethnic | (PouladzadehYassine and Shirmohammadi, 2015) |
| UNIMIB 2015 | 2015 | 15 | 2000 | Italian | (CioccaNapoletano and Schettini, 2015) |
| Instagram800K | 2016 | 43 | 808,964 | Multiethnic | (RichHaddadi and Hospedales, 2016) |
| UNICT-FD1200 | 2016 | 1200 | 4754 | Multiethnic | (FarinellaAllegra and Stanco, 2014) |
| UNIMIB 2016 | 2016 | 73 | 1027 | Italian | (CioccaNapoletano and Schettini, 2016) |
| EgocentricFood | 2016 | 9 | 5038 | Multiethnic | (Bolanos and Radeva, 2016) |
| VIREO Food-172 | 2016 | 172 | 110,241 | Chinese | (Chen and Ngo, 2016) |
| FOOD-5K | 2016 | 2 | 5000 | Multiethnic | (SinglaYuan and Ebrahimi, 2016) |
| FOOD-11 | 2016 | 11 | 16,643 | Multiethnic | (SinglaYuan and Ebrahimi, 2016) |
| NTUA-Food 2017 | 2017 | 82 | 3248 | Multiethnic | (Kogias et al., 2018) |
| ECUST Food Dataset | 2017 | 19 | 2978 | Multiethnic | (Liang and Li, 2017) |
| Madima 2017 | 2017 | 21 | 21,807 | Central European | (Allegra et al., 2017) |
| ChineseFoodNet | 2017 | 208 | 192,000 | Chinese | (Chen et al., 2017) |
| Eating Occasion Image to Food Energy | 2020 | 21 | 96 | Multiethnic | (He et al., 2020) |
| ChinaFood-100 | 2021 | 100 | 10,074 | Chinese | (Ma et al., 2021) |
| VIPER-FoodNet | 2021 | 82 | 14,991 | Multiethnic | (Mao et al., 2021) |

### Computational considerations when developing and using CNNs

It is important to note that recognising food in images comes with its own set of challenges, mainly because food pictures can be quite diverse and visually complex, making it tricky to tell similar looking foods apart (ChaitanyaShetty and Chiplunkar, 2023). That is why having high-quality and extensive datasets are crucial for training CNNs effectively in food classification (Minija and Emmanuel, 2017). One surprising research paper by Tusień et al. (2022) highlighted that only a marginal difference in time would be saved when a certain GPUs were used while comparing with the performance of standard CPUs. Apart from performance and data challenges, there is a need to acknowledge the complexity and efficiency aspects of using CNNs for food image classification (Xiao et al., 2021) compared with human analysis and categorisation. Some studies recommend using complex architectures with many layers and features, while others explore ways to reduce the computational demands and speed up training (Liu et al., 2016; Xiao et al., 2021) so no clear approach is evidently better to date and it difficult to compare them as they are different. It is clear that the choices made in designing these models play a significant role in finding the right balance between accuracy and computational efficiency (Xiao et al., 2021). There is a potential for integrating big data processing technologies such as Hadoop and Apache Spark to handle larger datasets and improve computational efficiency, although the specific techniques for food image classification remain an area for further investigation (Istephan and Siadat, 2015) and it an area that is continuously evolving due to the continuous placement of new products on the market. In practical terms, researchers are also considering affordability and speed when developing food recognition systems, highlighting the relevance and importance of this research field (Islam et al., 2018).

### Hyperparameter tuning of CNNs

One of the key steps in developing, testing and validating a model is tuning the hyperparameters in a CNN. There are so many options and if the model is very complex this step can be time consuming (i.e. can take days to tune each attempt) to get right using a mid-powered computer. However, the hyperparameter settings that help fine-tune the network to work optimally are generally vitally important if developing a model which will be accurate and robust. These hyperparameters, such as the Learning Rate (LR), the number of neurons, the optimizer, activation functions (AF), batch size, and epochs, play a crucial role in making accurate predictions (KaurKumar and Gupta, 2023) . But there's no one-size-fits-all solution for determining the perfect hyperparameters because different datasets have different requirments. So, the key is to find the right set of hyperparameters for your specific dataset. To do this, a developer will need to go through a process of hyperparameter tuning, which involves finding the ideal combination of hyperparameters for the dataset of interest. The following is a limited overview of the key CNN hyperparameters (Chen et al., 2017) that can be tuned when developing a CNN and a non-exhaustive list of similar approaches with associated hyperparameter tuning is shown in table 2.

**Table 2: High level overview of typical Convolutional Neural Network (CNN) hyperparameters than are optimised during CNN development using various approaches\* (Chen et al., 2017)**

|  |  |
| --- | --- |
| **Activation function** | Introduces non-linearity in the output of the neuron—for example, ReLU, Softmax, Sigmoid, Tanh, and Leaky ReLU. |
| **Total layers** | The number of layers is added in between the input and output layer until the test error stops improving |
| **Epochs** | The number of epochs determines the number of times a network weight is updated |
| **Batch Size** | It defines the sub-samples to train the model with sizes such as 32, 64, 128, etc |
| **Learning rate** | The LR determines how frequently the optimization algorithm updates the weight—for example, Adam, RMSProp, SGD, Adagrad, and AdaDelta |
| **Dropout** | Dropout is a method of regularization used to prevent overfitting. It is used between 20 to 50 percent of neurons depending upon the type of the problem. |

*\*There are various approaches to conducting this hyperparameter tuning including manual search (very laborious), Grid search, Random search, and Bayesian optimization.*

An non-exhaustive list is shown in Table 3 below highlighting the vast amount of research being undertaken with CNN models and in food image recognition in recent years. A diverse array of models and their associated hyperparameters used in food image analysis shows the complexity and variability in this task. The outcomes of these models vary considerably, with some authors achieving improved accuracy, reduced processing time, and the creation of new food datasets. Other key considerations in this review include optimisation of learning rates, layer architectures, and loss functions to enhance performance in this area of research.

For instance, in a study authored by Shekar and Dagnew (2019) a grid search approach to optimise the hyperparameters for the Random Forest (RF) tree in classifying cancer disease from images. This approach helped to identify the hyperparameters that maximised accuracy and minimised errors. Table 3 offers a summary of 17 recent studies using CNN where hyperparameters were discussed in more detail. The CNN architectures and object detection models found in the literature, including their specific hyperparameter settings is also captured highlighting the interest and development in this area of research.

Following the development, training, testing and validation of a potential food recognition models it is technically feasible to launch this on various platforms, including smartphones (Islam et al., 2018) or other mobile devices thus enhancing dissemination, accessibility and usability. However, this approach comes with its own set of hardware and deployment considerations that require consideration by the end user.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model type** | **Hyperparameters** | **Outcome** | **Source** |
| Faster R-CNN and CNN | LR 0.001, Momentum 0.9, Weight attenuation 0.0005, 60 epochs | Proposed model outperformed others. | (Wei and Wang, 2020) |
| Xception Model | LR 0.001 to 0.0001, Dropout 0.5, Batch size 16, 100 epochs, added layers | Multi-task CNN had better accuracy. | (Situju et al., 2019) |
| Proposed CNN-based model | Used jump convolution, global pooling instead of fully connected layer | Faster food recognition and reduced computation time. | (Xiao et al., 2021) |
| Clarifai developed CNN | Images processed with Clarifai CNN | AI algorithm performed well, especially on FOOD-5K dataset. | (Jia et al., 2019) |
| R-CNN and deep CNN | Multiple regions, training:testing 80:20, momentum 0.9 | Created a new food dataset from FOOD-101. | (Jiang et al., 2020) |
| Pre-trained models (ISIA Food-500) | Optimizer SGD, batch size 80, momentum 0.9, LR 10^-2 | Presented ISIA Food-500 dataset and compared with others. | (Min et al., 2022) |
| Deep CNN | Softmax, modified loss function | Improved accuracy and processing time. | (Lohala et al., 2021a) |
| Proposed RUMTL architecture | Softmax for single-label task, sigmoid for multi-label task, batch size 20, LR 2e-4, decay 0.2 | Introduced a new dataset for single and multi-label classification. | (AguilarBolaños and Radeva, 2019) |
| Mask R-CNN | Optimizer SGD, LR 0.001, momentum 0.9, batch size 128, 200000 epochs | Created a new dataset of Chinese food dishes. | Li (LiXu and Yuan, 2020) |
| U-Net | ReLu activation, cross-entropy loss, optimizer SGD, LR 0.001, momentum 0.9, batch size 5, weight decay 0.0005, 50 epochs | Focused on raw material and background detection. | Son (Son et al., 2021) |
| Voting-based ensemble (CNNs) | SGDM optimizer, batch size 64, LR 10^-3, various epochs | Achieved highest accuracy using ensemble approach. | (Tasci, 2020) |
| VGG, ResNet, Inception, WISeR | Used default parameters | Inception V3 outperformed others. | (Ma et al., 2021) |
| DenseNet121, ResNet50 (DenseFood) | Initial LR 0.01, changed to 0.001 with cosine decay, center loss + SoftMax CE | DenseFood model achieved highest accuracy. | (MetwalliShen and Wu, 2020) |
| ResNet50 | SGD optimizer, batch size 16, initial LR 0.001, dropout 0.5 | Achieved highest accuracy of 91.34%. | (Won, 2020) |
| VGG-16 | LR 0.001, Batch size 64, various loss functions tested | Multi-task CNN improved correlation coefficient. | (Chen et al., 2021) |
| Modified EfficientNetB0 | Swish activation, dropout layer | Achieved an accuracy of 92.33%. | (Kumari and Singh, 2019) |
| Modified EfficientNetB0 | Swish activation function, Used dropout layer | Achieved an accuracy of 92.33% | (TaiThanh and Hung, 2022) |

**Table 3: Summary of recent (2019 onwards) research studies using CNN and hyperparameter tuning for predicting outputs using food related images.**

## State of the art literature review summary

Preventing obesity through informed dietary choices is a difficult task facing many public health bodies across the developed world. Computer vision techniques, particularly Convolutional Neural Networks (CNNs), offers a useful solution for real-time food nutrition communication and enhanced regulatory compliance as discussed in this review. CNNs' versatility is evident in various applications, including medical diagnosis, criminal tracking and data analytics. However, recognizing complex food items in images remains challenging and demands high-quality data and extensive datasets. Hyperparameter tuning is a crucial step, impacting factors such as learning rates and batch sizes to optimize model performance and this can be achieved with little effort using semi-automated approaches such as grid search CV. Research in this field has shown remarkable progress over recent years, with continual improvements in CNNs' food recognition capabilities, scope and model performances. Nevertheless, tackling data biases, especially concerning geography and ethnicity, is key and must not be overlooked. Researchers continue to add to the already large field of research from many sources (Peer-reviewed literature, blogs, websites, social media etc.) as highlighted in this area exploring more innovative methods, thus there is an opportunity to offer more effective food safety regulation through these evolving systems.

## Primary research methodology

During the in-depth interviews the facilitator (i.e. the author) and the participants can delve deeper into their experiences, perspectives, and opinions. The interviews will be semi-structured using the questionnaire as a prompting framework, but the questions will be broad enough to promote discussion and explore specific topics or opportunities/pain points in detail. Post interview, it is envisaged that the data will be summarised, and the results will lead to a more comprehensive understanding of the research topic and opportunities for the end users of any proposed research artifact. Using the two distinct populations, each consisting of five experienced individuals may introduce bias from experience, however this can be explored during the interview. The purpose of these interviews is to gain accurate, relevant and reliable insights into the applicability and practical implications of the proposed research output and its usefulness, particularly within the domain of food safety regulation. The two groups selected for these interviews are current food safety regulators and food safety inspectors. The author has chosen to use in-depth interviews because this method offers an accurate, relevant and reliable understanding of the perspectives and experiences of these participants. By engaging in one-on-one conversations with representatives from both groups and building rapport, the authors aims to explore their perceptions of food safety, regulatory challenges, and the effectiveness of existing practices in terms of nutritional labelling. In-depth interviews are well-suited for capturing detailed qualitative data, allowing participants to express their thoughts, concerns, and recommendations in their way. This approach will help uncover rich, context-specific information that may not be accessible through other data collection methods such as surveys. In addition, this research methodology offers the advantage of personalised interaction, creating a conducive environment for participants to share their expertise openly. However, the authors acknowledges the potential limitations, such as the possibility of individual bias and the relatively small sample size i.e. n=5. One key bias consideration that will need to be managed is that the author is a senior regulator and this may have an impact on biasing the output of other regulators and inspectorate. This point will be disarmed early in the discussion (i.e. at first contact) as it will be made clear that the author will be asking the questions as a MSc student from CCT college and not as a regulator and the responses will be used solely for the purpose of the research and not for regulatory follow up. This proposal was intentionally designed not to include sensitive data as outlined by the EU Commission (2021).

## Sampling strategy

### Sampling of the Primary data

For the primary research component of the proposed project in-depth interviews will be used on two distinct populations of (regulators/inspectors), each consisting of n=5 observations. The type of sampling is non-probability self-selection as the author will select experienced participants for this research. The reason for this is that it is associated with complex regulation and is open to interpretation thus increasing the bias and noise from respondents. Junior candidates may not appreciate or consider the nuances of this complex area.

This approach will be used to generate primarily qualitative information, to gain trust from the interviewees and ensure a clear understating of the questions. This in-depth analysis will assist the author in understanding the challenges faced by both the regulators and the inspectorate. The intention is to conduct these interviews prior to conducting the analysis so the author can understand the so called ‘pain points’ and challenges faced by the ‘people on the ground’ and plan how to ensure any potential benefits or opportunities can be maximised in the research development and reporting. The learning from these interviews will also be used in the critique of any research findings following the development and implementation of the DA models and research outputs. In terms of ethical considerations, it will be made clear from the outset that the information will be used primarily for this M.Sc. project and any information relating to personal data or the identification of any inspector or regulator or any identifiable natural person will not be published in the final thesis nor will it enter the public domain in line with GDPR (EU Commission, 2021) . A key part of setting the scene with participants, the author will explain to research participants what the research is about, what their participation in your project will entail and any risks that may be involved in line with EU Commission (2021) guidance on data protection and ethics. Following this the authors will aim to encourage the participants to relax and aim to quickly build up rapport as this area of enforcement can be a point of conflict in some jurisdictions. It is not foreseen that this data entails higher risk (EU Commission, 2021) data processing as it is based on open source non-personal data specifically used to be shared and modelled.

The style of the interview will be relaxed and open using the TED (Tell, Explain and Describe) approach to encourage discussion and use open questions allowing the participants to express their view while keeping them focused on the five main themes, namely:

1. The current situation with using food labels in the day to day work for enforcement/compliance
2. Challenges with the current approach
3. Opportunities for improvements / developments
4. The future direction of this area and expected trends in their professional fields
5. How the participants think/feel the author’s proposed approach (CNN predication) could assist current work

It is intentional that theme five comes last as the author doesn’t intend to bias the participants with a possible solution before exploring their opinions, challenges and ideas.

### Sampling of the secondary data

In the context of this research proposal, the author will implement a sampling strategy using the Open Food Facts database of food images. This approach aims to include all available samples from the database, totalling approximately 100,000 images from three distinct food categories. The objective is to ensure that our dataset is as balanced as possible, with an equal representation of food categories across the entire population. The type of sampling employed in this research proposal is a combination of non-probabilistic and convenience sampling methods. In the context of this specific dataset, this approach can be considered as census sampling. It entails selecting data that is readily available, accessible, licenced under open database licence (ODBL) and studying the complete population of food categories without taking a sub-sample.

A key advantage of this approach ensures that the research supported by high computing capacity can analyse every available food image, offering a thorough representation of the dataset for our research. As the underlying dataset increases (in the open-source nature) the model can be retrained to capture new information and products. The reliance on this approach (non-probabilistic convenience sampling) will likely introduce bias and limit the applicability of the research findings to the three food groups due to potential overrepresentation of certain food categories or sources compared to more general food images. This will be controlled for as much as possible in the design of the research project.

# Conclusion

This report outlined a research project proposal using Convolutional Neural Networks (CNNs) and their optimisation for predicting the nutritional content of foods through crowd-sourced images. The research objectives aim to tackle the challenges in using crowd-sourced image data for food nutrition prediction while highlighting the practical applications of CNN models. The potential output of the proposed research will aim to assist Environmental Health Officers (EHOs) in their regulatory roles and also consumers in making informed dietary choices using images. The current state-of-the-art review highlights the potential of using CNNs in recognising and predicting various food images, though it's not without challenges and computational demands. The importance of fine-tuning hyperparameters and the use of extensive datasets is also addressed. Additionally, CNN models can be deployed on various platforms, such as smartphones, making them accessible and user-friendly which could be useful to the end user. Sampling strategies for primary and secondary data were discussed, considering legal, ethical, validity and bias considerations (Commission, 2021). The primary research methodology involves in-depth interviews with current food safety regulators and inspectors to gain insights into the practical implications of the proposed research output and its utility in the food safety regulation domain. While potential limitations are acknowledged, like bias and a small sample size, this report proposes a way for a thorough exploration of CNNs in nutritional profiling for foods, considering both the opportunities and challenges in this problem domain.

# Materials and Methods

Add sampling strategy here

# Results and discussion

Primary data discussion

## EDA

Figure 1 Sentiment Classification Frame work (AlzamzamiHoda and El Saddik, 2020)

## Forecasting

### ARIMA and SARIMA

A graph of a graph

Description automatically generated with medium confidence

Figure 2 Static graph of the results of an Auto-Sarima (ARIMA) results where 1,4 and 12 weeks forecast are captured.

### Forecasting using gradient boosting: LightGBM

### Recurrent Neural Network (RNN) model and hyperparameter tuning

To further investigate the impact of a deep learning approach the authors used a hyperparameter-tuned Recurrent Neural Network (RNN) model, developed to capture temporal patterns in the sequential (date) data. This section analyses the details of the RNN model configuration, hyperparameter tuning, and the forecasted (1 week, 4 weeks and 12 weeks) results. This RNN used a sequence of long Short Term memory (LSTM) layers to capture the patter of the time series data. The RNN was designed using hyperparameter tuning including hidden units (10, 50, 100), learning rate 0.001,0.01 and 0.1, optimizer function (Adam, SGD and RMSPROP), activation function (RELU and Tanh) and batch size (32, 64 and 128), with loss being measured using MSE.

Hidden Units: The number of hidden units in each LSTM layer is key hyperparameter and because of this the author checked various configurations (10, 50, and 100 units per layer). By increasing the number of units the model's capacity to capture intricate patterns in the time series can be improved, however, it also increases the model's complexity and the risk of overfitting. The key here is to strike the right balance hence hyperparameter tuning approach with a wide approach. In this case 10 hidden units were selected using the hyperparameter approach.

Learning Rate: These values: 0.001, 0.01, and 0.1 were tested to establish the best learning rate. A lower learning rate call allows the model to converge more accurately but the drawback of this is that it might require more training time. On the other hand, the higher learning rate accelerates convergence but runs the risk of overshooting the optimal model. In this case a learning rate of 0.01 was selected using the hyperparameter approach.

Optimizer Function: Although previous advice was to generally use ADAM the author evaluated three applicable optimisers for this approach: Adam, SGD (Stochastic Gradient Descent), and RMSprop. As previously informed Adam combines the advantages of both AdaGrad and RMSprop and is widely favoured for its effectiveness in various tasks. SGD and RMSprop are alternatives that can provide stable convergence in specific scenarios. In this case the RMSprop optimizer was selected using the hyperparameter approach. The simplicity of RMSprop’s approach compared with Adam’s approach may have factored into its selection resulting in the lower MAE overall as the dataset was fairly uncomplex to begin with.

Activation Function: This defines the transformation applied to the data within the LSTM cells. ReLU (Rectified Linear Unit) and Tanh (Hyperbolic Tangent) were used in this turning. ReLU introduces non-linearity to the model, while Tanh is known for its squashing effect (Shen et al., 2022). The choice between these functions can influence the model's capacity to capture different types of patterns within time series data. In this case the ‘tanh’ was selected using the hyperparameter approach. Than was likely the best hyperparameter since it’s a centred activation (TheMightyBarbarian, 2019) function whereas Relu is not centred. Tanh is generally advantageous where the data is zero centred (i.e. -1 to +1) as this sentiment data was.

Batch Size: This is an important hyperparameter to also tune because it sets the number of samples processed at each of training steps. Three batch sizes: 32, 64, and 128 were explored. A smaller batch size offers a more stochastic learning process and can help avoid getting stuck in local minima, however, a larger batch size may provide computational efficiency (Omar, 2022). In this case a batch size of 128 was selected by the hyperparameter approach.

In addition, the Loss Function Mean Squared Error (MSE) to determine the best performing conditions and was used as it quantifies the average squared difference between the predicted and actual values. It is a common choice (Yang, 2022) for regression based prediction and was suited to our goal of forecasting numerical values in the time series.

A graph showing a graph of a graph

Description automatically generated with medium confidence

Figure 3 Output of a hyper-parameter tuned RNN prediction model, forecasting 1, 4 and 12 weeks in advance, with MAE of 0.084 and MSE of 0.007

## Dynamic dashboard development

With key stakeholders in mind this dashboard in Fig 4 was developed to share the insights learned from carrying out an auto-SARIMA which returned the most accurate method as a standard ARIMA as there was limited seasonal features identified. This simplicity of the dashboard design was intentional to ensure stakeholders could understand the impact of extending the forecast ahead at the three time points. This dashboard focuses on four key elements - training data, testing data, forecasted sentiment and relevant prediction confidence intervals (CI) providing users with the important and relevant information such as the increase in error (CI) as the forecast time frame increases. This dashboard is designed to be responsive, adapting to different user needs and the automatically selected SARIMA model (ARIMA) has effectively forecasted sentiment trends. The Mean Absolute Error (MAE) and Mean Squared Error (MSE) provide a comprehensive view of forecast accuracy which stakeholders can use to understand how well the model performs, with associated health warning.

A screen shot of a graph

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Figure 4 Illustration of a dash created interactive dashboard displaying the output from an ARIMA forecasting task including highlighting the training, testing and predicted sentiment score ranging from 4-12 weeks, including confidence intervals.

## Forecasting discussion and associated health warning

On a general note, forecasting or predicting future events, especially human behaviour or expression can be a challenging task. This is especially true when trying to predict events significantly further into the future than the time frame of the available data. In this analysis authors attempted to forecast 1, 4 and 12 weeks in advance using a dataset which only spanned 12 weeks and this by design is inherently limited for several reasons. Firstly, short time series data may not and did not capture any long-term trends, seasonality, or other significant patterns that could affect the forecasts over extended horizons. Secondly, the inherent uncertainty in forecasted tends can grow with the forecast horizon as illustrated by the increasing confidence intervals (Fig.4) displayed over time with short-term variability increasing errors over longer periods. Thirdly, the performance of complex models like RNN or gradient boosting can deteriorate as they overfit to the limited data, potentially leading to unreliable predictions. These advanced modelling techniques may provide insights for the short term, however, when relying on them for extended forecasts the authors which to warn that these results come with a high degree of uncertainty and should be complemented with other sources of information.

## 

# Conclusion

# Appendix

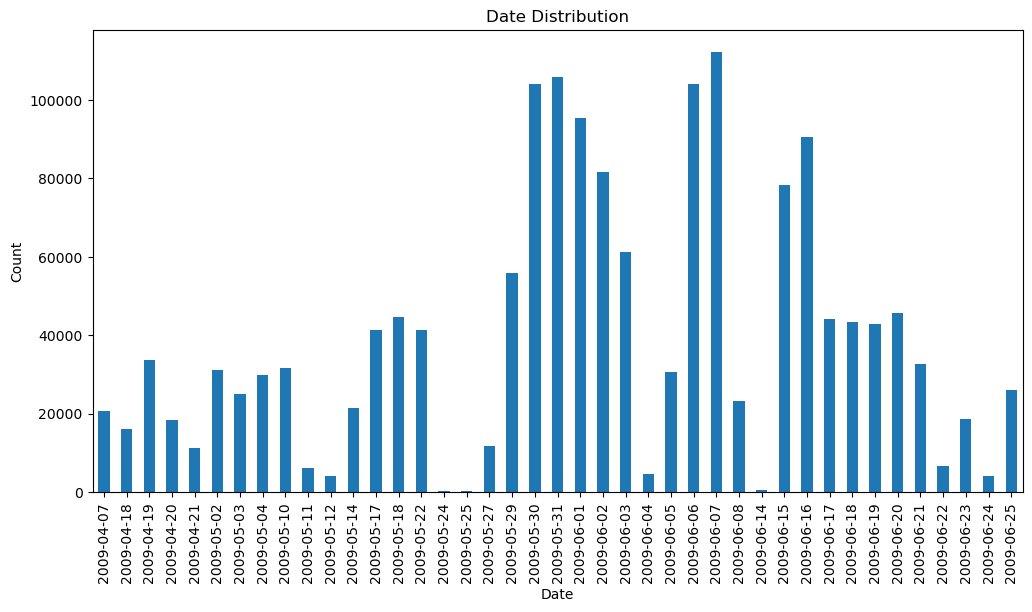
**Error! Reference source not found.**

Figure 5 Distribution of number (count) of tweets over time from the selected dataset.

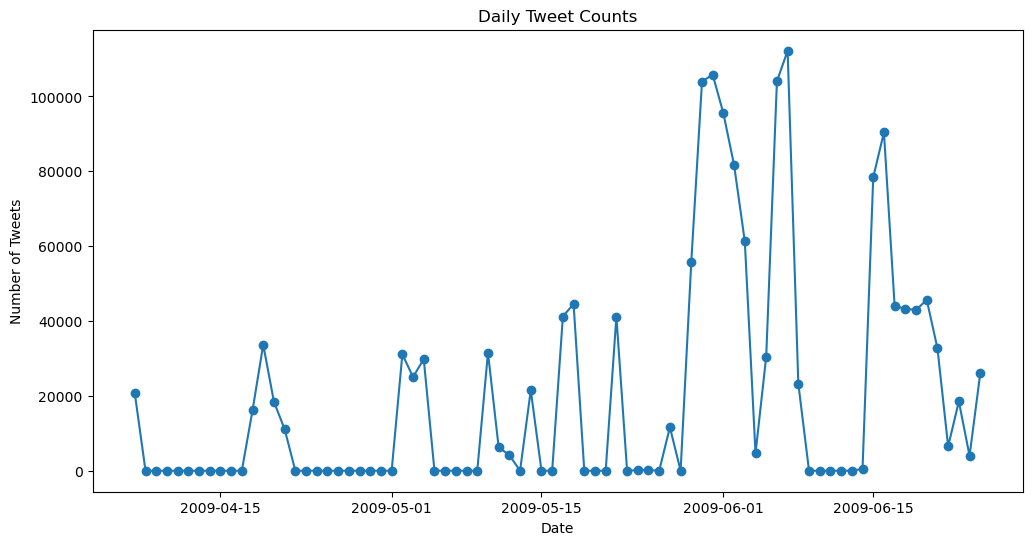


Figure 6 Distribution of number (count) of tweets over time from the selected dataset.

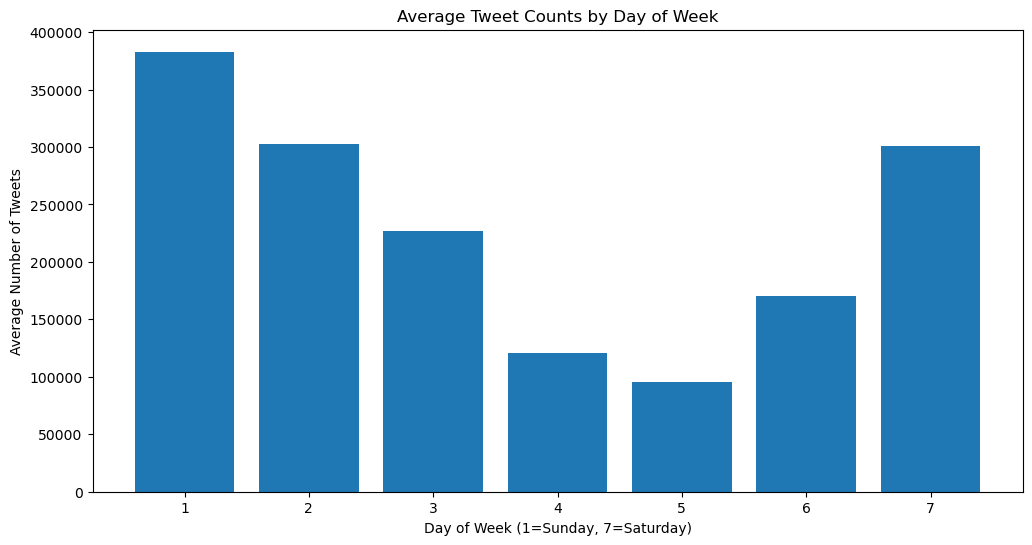


Figure 7 Total number of tweets per day of the week.

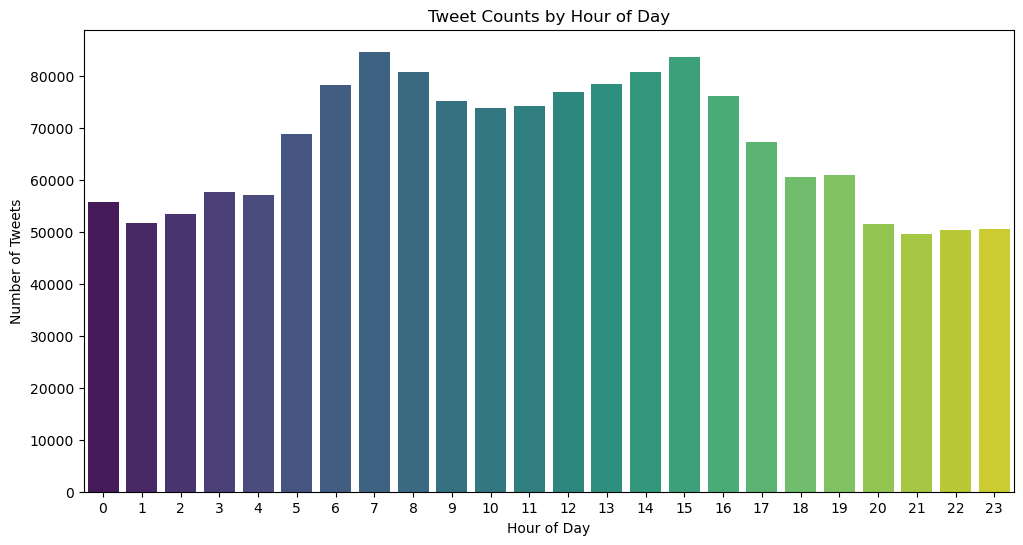


Figure 8 Total number of tweets per hour of the day.

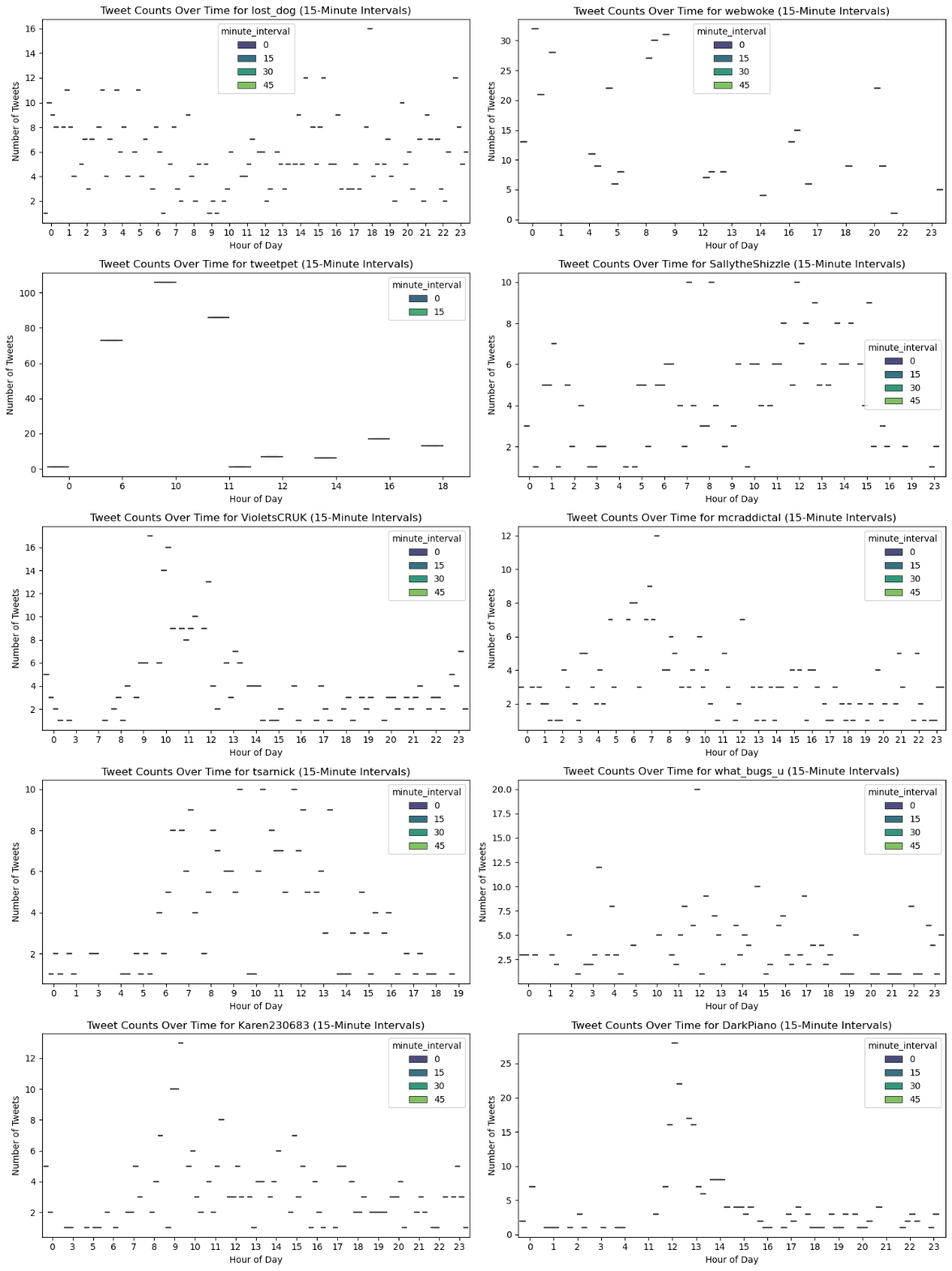


Figure 9 Distribution of user tweets of the top 10 most frequent posters of tweets per 15 min intervals of each hour each day.

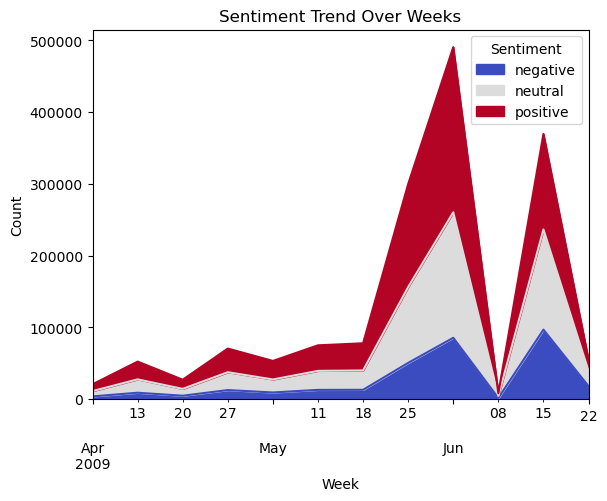


Figure 10 Distribution of sentiment classification (positive, negative and neutral) by week of analysis from tweets in the dataset.

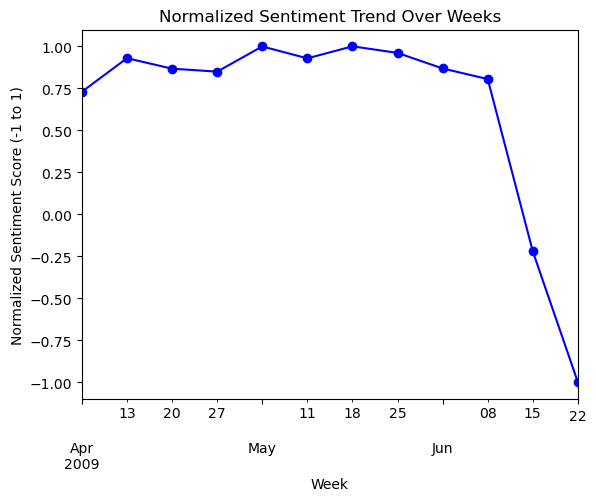
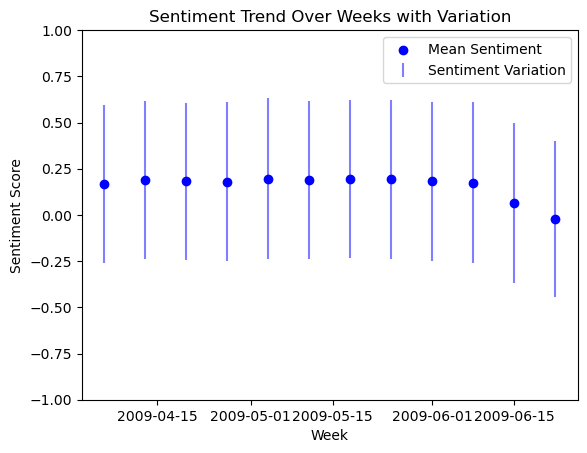


Figure 11 Distribution of sentiment classification (positive, negative and neutral) by week of analysis from tweets in the dataset.

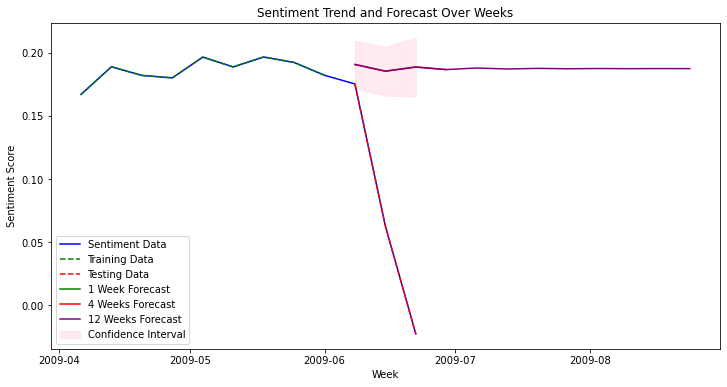


Figure 12 Distribution of sentiment classification (positive, negative and neutral) by week of analysis from tweets in the dataset.

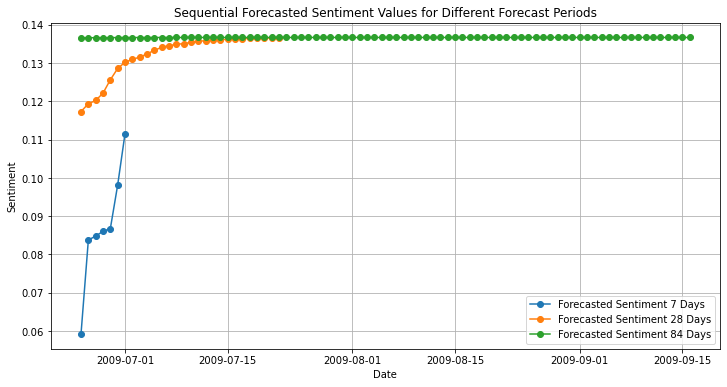


Figure 13 Forecasted sentiment from twitter using an hyperparameter tuned RNN model.

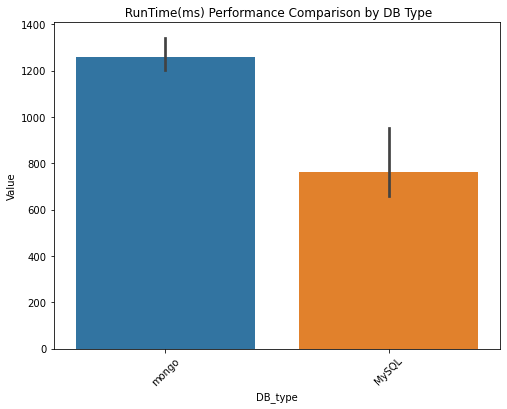


Figure 14 Performance of database type (MongoDB Vs MySQL) when appraised using YCSB across 6 workloads. Runtime was the only significantly different variable of the 12 variables tested.

# Bibliography

AGUILAR, E., BOLAÑOS, M. & RADEVA, P. 2019. Regularized uncertainty-based multi-task learning model for food analysis. *Journal of Visual Communication and Image Representation,* 60**,** 360-370.

AKBAR, S., PEIKARI, M., SALAMA, S., NOFECH-MOZES, S. & MARTEL, A. Transitioning between convolutional and fully connected layers in neural networks. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3, 2017. Springer, 143-150.

ALLEGRA, D., ANTHIMOPOULOS, M., DEHAIS, J., LU, Y., STANCO, F., FARINELLA, G. M. & MOUGIAKAKOU, S. A multimedia database for automatic meal assessment systems. New Trends in Image Analysis and Processing–ICIAP 2017: ICIAP International Workshops, WBICV, SSPandBE, 3AS, RGBD, NIVAR, IWBAAS, and MADiMa 2017, Catania, Italy, September 11-15, 2017, Revised Selected Papers 19, 2017. Springer, 471-478.

BIERMAYR-JENZANO, P. 2019. Food Systems, obesity, and gender in Latin America. *FAO-LAC. Recuperado de* [*http://ebrary*](http://ebrary)*. ifpri. org/utils/getfile/collection/p15738coll2/id/133322/filename/133534. pd*.

BOLANOS, M. & RADEVA, P. Simultaneous food localization and recognition. 2016 23rd International conference on pattern recognition (ICPR), 2016. IEEE, 3140-3145.

BOSSARD, L., GUILLAUMIN, M. & VAN GOOL, L. Food-101–mining discriminative components with random forests. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13, 2014. Springer, 446-461.

BUHL, N. 2023. Convolutional Neural Networks (CNN) Overview. *4 min read* [Online].

CHAITANYA, A., SHETTY, J. & CHIPLUNKAR, P. 2023. Food Image Classification and Data Extraction Using Convolutional Neural Network and Web Crawlers. *Procedia Computer Science,* 218**,** 143-152.

CHEN, J. & NGO, C.-W. Deep-based ingredient recognition for cooking recipe retrieval. Proceedings of the 24th ACM international conference on Multimedia, 2016. 32-41.

CHEN, J., ZHU, B., NGO, C. W., CHUA, T. S. & JIANG, Y. G. 2021. A Study of Multi-Task and Region-Wise Deep Learning for Food Ingredient Recognition. *IEEE Transactions on Image Processing,* 30**,** 1514-1526.

CHEN, M.-Y., YANG, Y.-H., HO, C.-J., WANG, S.-H., LIU, S.-M., CHANG, E., YEH, C.-H. & OUHYOUNG, M. 2012. Automatic chinese food identification and quantity estimation. *SIGGRAPH Asia 2012 Technical Briefs.*

CHEN, M., DHINGRA, K., WU, W., YANG, L., SUKTHANKAR, R. & YANG, J. PFID: Pittsburgh fast-food image dataset. 2009 16th IEEE International Conference on Image Processing (ICIP), 2009. IEEE, 289-292.

CHEN, X., ZHU, Y., ZHOU, H., DIAO, L. & WANG, D. 2017. Chinesefoodnet: A large-scale image dataset for chinese food recognition. *arXiv preprint arXiv:1705.02743*.

CIOCCA, G., NAPOLETANO, P. & SCHETTINI, R. Food recognition and leftover estimation for daily diet monitoring. New Trends in Image Analysis and Processing--ICIAP 2015 Workshops: ICIAP 2015 International Workshops, BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, September 7-8, 2015, Proceedings 18, 2015. Springer, 334-341.

CIOCCA, G., NAPOLETANO, P. & SCHETTINI, R. 2016. Food recognition: a new dataset, experiments, and results. *IEEE journal of biomedical and health informatics,* 21**,** 588-598.

COMMISSION, E. E. 2021. Ethics and data protection. European Commission.

DALAKLEIDI, K. V., PAPADELLI, M., KAPOLOS, I. & PAPADIMITRIOU, K. 2022. Applying Image-Based Food-Recognition Systems on Dietary Assessment: A Systematic Review. *Advances in Nutrition,* 13**,** 2590-2619.

DENG, J., DONG, W., SOCHER, R., LI, L.-J., LI, K. & FEI-FEI, L. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition, 2009. Ieee, 248-255.

FARINELLA, G. M., ALLEGRA, D. & STANCO, F. A benchmark dataset to study the representation of food images. European Conference on Computer Vision, 2014. Springer, 584-599.

HASSANIEN, A. E. 2019. *Machine learning paradigms: Theory and application*, Springer.

HE, J., SHAO, Z., WRIGHT, J., KERR, D., BOUSHEY, C. & ZHU, F. Multi-task image-based dietary assessment for food recognition and portion size estimation. 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2020. IEEE, 49-54.

HE, Y., XU, C., KHANNA, N., BOUSHEY, C. J. & DELP, E. J. Analysis of food images: Features and classification. 2014 IEEE international conference on image processing (ICIP), 2014. IEEE, 2744-2748.

HERRANZ, L., XU, R. & JIANG, S. A probabilistic model for food image recognition in restaurants. 2015 IEEE International Conference on Multimedia and Expo (ICME), 2015. IEEE, 1-6.

ISLAM, M. T., SIDDIQUE, B. M. N. K., RAHMAN, S. & JABID, T. Food Image Classification with Convolutional Neural Network. 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 21-24 Oct. 2018 2018. 257-262.

J, A. & SURESH L, P. 2023. A novel fast hybrid face recognition approach using convolutional Kernel extreme learning machine with HOG feature extractor. *Measurement: Sensors,* 30**,** 100907.

JIA, W., LI, Y., QU, R., BARANOWSKI, T., BURKE, L. E., ZHANG, H., BAI, Y., MANCINO, J. M., XU, G., MAO, Z.-H. & SUN, M. 2019. Automatic food detection in egocentric images using artificial intelligence technology. *Public Health Nutrition,* 22**,** 1168-1179.

JIANG, L., QIU, B., LIU, X., HUANG, C. & LIN, K. 2020. DeepFood: Food Image Analysis and Dietary Assessment via Deep Model. *IEEE Access,* 8**,** 47477-47489.

JOSHI, N. S., KHULLAR, S., SAPONAS, T. S., MORRIS, D. & BEIJBOM, O. 2017. Restaurant-specific food logging from images. Google Patents.

KAUR, R., KUMAR, R. & GUPTA, M. 2023. Deep neural network for food image classification and nutrient identification: A systematic review. *Reviews in Endocrine and Metabolic Disorders,* 24**,** 633-653.

KAWANO, Y. & YANAI, K. Foodcam-256: a large-scale real-time mobile food recognitionsystem employing high-dimensional features and compression of classifier weights. Proceedings of the 22nd ACM international conference on Multimedia, 2014. 761-762.

KOGIAS, K., ANDREADIS, I., DALAKLEIDI, K. & NIKITA, K. S. A two-level food classification system for people with diabetes mellitus using convolutional neural networks. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018. IEEE, 2603-2606.

KUMARI, M. & SINGH, T. Food Image to Cooking Instructions Conversion Through Compressed Embeddings Using Deep Learning. 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW), 8-12 April 2019 2019. 81-84.

LI, Y., XU, X. & YUAN, C. 2020. Enhanced Mask R-CNN for Chinese Food Image Detection. *Mathematical Problems in Engineering,* 2020**,** 6253827.

LIANG, Y. & LI, J. 2017. Computer vision-based food calorie estimation: dataset, method, and experiment. *arXiv preprint arXiv:1705.07632*.

LIU, C., CAO, Y., LUO, Y., CHEN, G., VOKKARANE, V. & MA, Y. Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. Inclusive Smart Cities and Digital Health: 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings 14, 2016. Springer, 37-48.

LOHALA, S., ALSADOON, A., PRASAD, P., ALI, R. S. & ALTAAY, A. J. 2021a. A novel deep learning neural network for fast-food image classification and prediction using modified loss function. *Multimedia Tools and Applications,* 80**,** 25453-25476.

LOHALA, S., ALSADOON, A., PRASAD, P. W. C., ALI, R. S. & ALTAAY, A. J. 2021b. A novel deep learning neural network for fast-food image classification and prediction using modified loss function. *Multimedia Tools and Applications,* 80**,** 25453-25476.

MA, P., LAU, C. P., YU, N., LI, A., LIU, P., WANG, Q. & SHENG, J. 2021. Image-based nutrient estimation for Chinese dishes using deep learning. *Food Research International,* 147**,** 110437.

MAHAJAN, N., HOLZWANGER, E., BROWN, J. G. & BERZIN, T. M. 2023. Deploying automated machine learning for computer vision projects: a brief introduction for endoscopists. *VideoGIE,* 8**,** 249-251.

MAO, R., HE, J., SHAO, Z., YARLAGADDA, S. K. & ZHU, F. Visual aware hierarchy based food recognition. International conference on pattern recognition, 2021. Springer, 571-598.

MATSUDA, Y., HOASHI, H. & YANAI, K. Recognition of multiple-food images by detecting candidate regions. 2012 IEEE International Conference on Multimedia and Expo, 2012. IEEE, 25-30.

METWALLI, A. S., SHEN, W. & WU, C. Q. Food Image Recognition Based on Densely Connected Convolutional Neural Networks. 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 19-21 Feb. 2020 2020. 027-032.

MEZGEC, S. & KOROUŠIĆ SELJAK, B. 2017. NutriNet: a deep learning food and drink image recognition system for dietary assessment. *Nutrients,* 9**,** 657.

MIN, W. Q., LIU, L. H., LIU, Y. X., LUO, M. J. & JIANG, S. Q. 2022. A Survey on Food Image Recognition. *Jisuanji Xuebao/Chinese Journal of Computers,* 45**,** 542-566.

MINIJA, S. J. & EMMANUEL, W. S. Food image classification using sphere shaped—Support vector machine. 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017. IEEE, 109-113.

OMAR. 2022. *Is batching a way to avoid local minima?* [Online]. Available: <https://stats.stackexchange.com/questions/565101/is-batching-a-way-to-avoid-local-minima> [Accessed 27/10/2023].

PARK, S.-J., PALVANOV, A., LEE, C.-H., JEONG, N., CHO, Y.-I. & LEE, H.-J. 2019. The development of food image detection and recognition model of Korean food for mobile dietary management. *Nutrition research and practice,* 13**,** 521-528.

PHAM, C. & THUY, N. T. T. Fresh food recognition using feature fusion. 2014 International Conference on Advanced Technologies for Communications (ATC 2014), 2014. IEEE, 298-302.

POULADZADEH, P., YASSINE, A. & SHIRMOHAMMADI, S. Foodd: food detection dataset for calorie measurement using food images. New Trends in Image Analysis and Processing--ICIAP 2015 Workshops: ICIAP 2015 International Workshops, BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, September 7-8, 2015, Proceedings 18, 2015. Springer, 441-448.

RICH, J., HADDADI, H. & HOSPEDALES, T. M. Towards bottom-up analysis of social food. Proceedings of the 6th international conference on digital health conference, 2016. 111-120.

SHEKAR, B. H. & DAGNEW, G. Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data. 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), 25-28 Feb. 2019 2019. 1-8.

SHEN, S.-L., ZHANG, N., ZHOU, A. & YIN, Z.-Y. 2022. Enhancement of neural networks with an alternative activation function tanhLU. *Expert Systems with Applications,* 199**,** 117181.

SINGLA, A., YUAN, L. & EBRAHIMI, T. Food/non-food image classification and food categorization using pre-trained googlenet model. Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, 2016. 3-11.

SITUJU, S. F., TAKIMOTO, H., SATO, S., YAMAUCHI, H., KANAGAWA, A. & LAWI, A. 2019. Food constituent estimation for lifestyle disease prevention by multi-task CNN. *Applied Artificial Intelligence,* 33**,** 732-746.

SON, G.-J., KWAK, D.-H., PARK, M.-K., KIM, Y.-D. & JUNG, H.-C. 2021. U-Net-Based Foreign Object Detection Method Using Effective Image Acquisition System: A Case of Almond and Green Onion Flake Food Process. *Sustainability,* 13**,** 13834.

SREEDHA, B., NAIR, P. R. & MAITY, R. 2023. Non-invasive early diagnosis of jaundice with computer vision. *Procedia Computer Science,* 218**,** 1321-1334.

SUN, E. Small-scale image recognition based on Cascaded Convolutional Neural Network. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021. IEEE, 2737-2741.

TAI, T. T., THANH, D. N. H. & HUNG, N. Q. 2022. A Dish Recognition Framework Using Transfer Learning. *IEEE Access,* 10**,** 7793-7799.

TASCI, E. 2020. Voting combinations-based ensemble of fine-tuned convolutional neural networks for food image recognition. *Multimedia Tools and Applications,* 79**,** 30397-30418.

THEMIGHTYBARBARIAN. 2019. *How does a zero centered activation functions like tanh helps in gradient decent?* [Online]. Data Science: Stack Exchange Available: <https://datascience.stackexchange.com/questions/43603/how-does-a-zero-centered-activation-functions-like-tanh-helps-in-gradient-decent> [Accessed].

TUSIEŃ, E., WILKE, A., WOŹNA, J. & CZARNUL, P. Food Classification from Images Using a Neural Network Based Approach with NVIDIA Volta and Pascal GPUs. International Conference on Computer Information Systems and Industrial Management, 2022. Springer, 269-283.

WANG, X., KUMAR, D., THOME, N., CORD, M. & PRECIOSO, F. Recipe recognition with large multimodal food dataset. 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2015. IEEE, 1-6.

WEI, P. & WANG, B. 2020. Food image classification and image retrieval based on visual features and machine learning. *Multimedia Systems***,** 1-12.

WHO. 2023. *Obiesity Fact Sheet* [Online]. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>: WHO. Available: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> [Accessed 2023].

WON, C. S. 2020. Multi-Scale CNN for Fine-Grained Image Recognition. *IEEE Access,* 8**,** 116663-116674.

XIAO, L., LAN, T., XU, D., GAO, W. & LI, C. 2021. A simplified CNNs visual perception learning network algorithm for foods recognition. *Computers & Electrical Engineering,* 92**,** 107152.

YANG, B., WU, J., IKEDA, K., HATTORI, G., SUGANO, M., IWASAWA, Y. & MATSUO, Y. 2022. Face-mask-aware Facial Expression Recognition based on Face Parsing and Vision Transformer. *Pattern Recognition Letters,* 164**,** 173-182.

YANG, D. 2022. Correlogram, predictability error growth, and bounds of mean square error of solar irradiance forecasts. *Renewable and Sustainable Energy Reviews,* 167**,** 112736.

YUNUS, R., ARIF, O., AFZAL, H., AMJAD, M. F., ABBAS, H., BOKHARI, H. N., HAIDER, S. T., ZAFAR, N. & NAWAZ, R. 2018. A framework to estimate the nutritional value of food in real time using deep learning techniques. *IEEE Access,* 7**,** 2643-2652.

ZHU, L., SPACHOS, P., PENSINI, E. & PLATANIOTIS, K. N. 2021. Deep learning and machine vision for food processing: A survey. *Current Research in Food Science,* 4**,** 233-249.