Manuscript Submitted	4 Nov 2019
Accepted	26 Nov 2019
Published	31 Dec 2019

Classification of Food Nutrients Composition using Deep Learning

Abdul Salam Bin Abdul Aziz
Faculty of Science, Technology, Engineering & Mathematics,
International University of Malaya-Wales,
salamlimlom@gmail.com

Riyaz Ahamed Ariyaluran Habeeb
Faculty of Science, Technology, Engineering & Mathematics,
International University of Malaya-Wales,
riyaz@iumw.edu.my

Abstract

Deep Learning is the technique that uses multiple layers of a neural network to automatically distinguish patterns to learn to make predictions. A problem that humans face on a daily basis is how to make a conscious decision regarding our daily food consumption that is nutritious and healthy. By having a tool that helps facilitate the decision making process of what type of food to eat by showing useful nutritional information to us immediately would greatly improve our lives. By critically analysing prominent research papers that relate to deep learning techniques to classify food and their nutrients composition, we decided upon the suitable Deep Learning algorithm to classify food nutrients composition as well as the appropriate image dataset to be used. Therefore in this paper we propose the classification of food nutrients composition utilizing deep learning techniques. The proposed framework uses convolutional neural networks (CNN) as a basis of recognising images of food and classifying the food into their corresponding nutrients composition such as fats, carbohydrates, proteins and more. As part of our future work, we shall use the proposed framework to conduct the training and implementation of the deep learning model to make predictions on food nutrients. The chosen dataset shall be used to train the model where patterns and characteristics of the food images are distinguished over multiple passes of the neural network. Once the model has been trained, then new food images may be introduced to make a prediction from the context that have been learned from before.

Keywords: *Deep Learning, Artificial Neural Network, Convolutional Neural Network, Food Nutrients Classification, Nutrition Composition.*

1. Introduction

Artificial Intelligence (AI) is a very broad concept with many distinguishing components, tools and methods that can be used to achieve Artificial Intelligence. One promising aspect of AI is Deep Learning where a computer algorithm is fed with data that it then uses to automatically learn patterns from, without the need for human intervention. Deep learning is a subfield of Machine Learning where the differential factor of deep learning compared to machine learning is the use of multiple layers in the neural networks. The ability to delegate key decision making processes to computer systems in any form of industry or business, opens up an infinite amount of possibilities for us humans to progress and prosper. By encouraging computers to decide for us, in a short amount of time, a particular course of action, from a mountain of data that may need years of manual labour for us to sift through, leaves us open to focus on more important aspects of our work. Not to mention how cost effective, efficient and convenient doing so might bring us. Figure 1 below we propose an example of the basic concept of deep learning and their attributes.

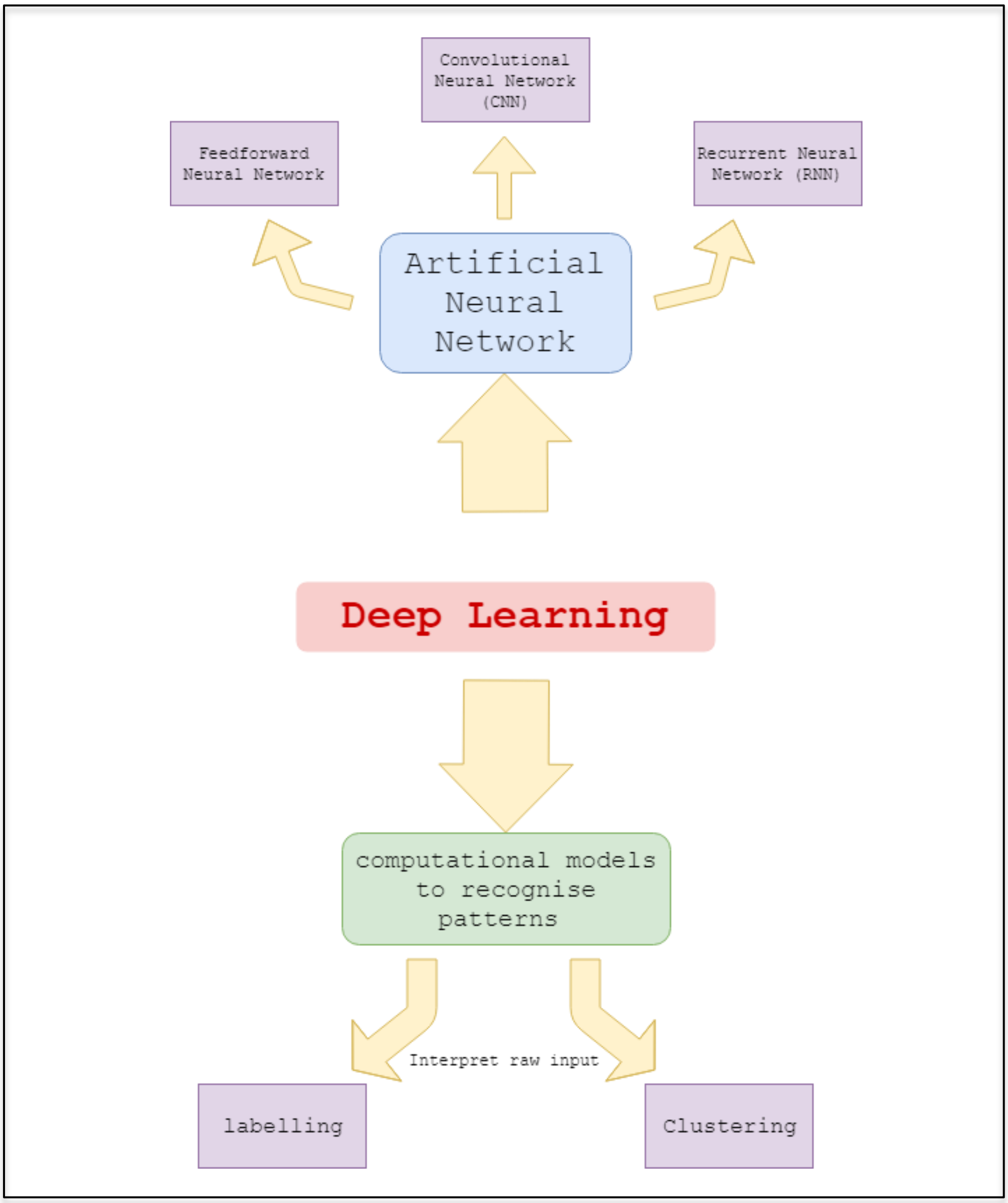


Figure 1: Example of the deep learning concept

In this paper, the research aim is to develop a system that identifies and classifies food items according to their nutritional compositions, such as fats and carbohydrates, using deep learning techniques. More and more people all around the world are suffering from chronic diseases that are associated with having poor dietary habits and an inactive lifestyle. People should pay more attention on monitoring daily caloric intake according to the recommended daily food allowance that is outlined by the health authorities. Having a comprehensive tool to monitor the nutritional composition of their meal, according to the food pyramid, is crucial for the public to have an awareness of what they consume. It is important for us to identify and classify our daily meals into its nutrient components, (e.g. carbohydrates, fats, protein and etc.) so that we understand the effects they have on our body as well as make a conscious decision of how much nutrients we eat. From what has gathered, there is no particular food intake monitoring system that considers the composition of the food, but rather there is an existing system that only monitors the calories of the food. This work is proposed by Parisa Pouladzadeh et al., where a system that classifies and recognises food items to obtain the calorific values of that item (Pouladzadeh et al., 2016). This system lacks in features whereby it does not take into consideration the type of nutrition the food is made of. This is a limitation for the user since knowing the particular type of nutrients composition in their food is much more important compared to just knowing the number of calories of the dish. Therefore there is a need for a system that addresses these concerns regarding people's nutritional requirements and to further improve their health.

The significance of having this kind of system is that we need to do even more to raise awareness on the negative effects of overconsumption of food or having an imbalanced diet. To help increase this health awareness, there needs to be a tool for the public to monitor and make a well informed decision regarding their daily food intake. Therefore it is important for us to determine the major features that are required to build a deep learning based application for food nutritional composition classification, as well as the appropriate deep learning techniques to most effectively classify the nutritional composition of any particular food.

The main contributions of this research is the development of a deep learning based application for food nutritional composition classification, as well as evaluating the proposed application with openly available datasets.

The paper is structured as follows; Section 2 will discuss and compare the existing research work regarding the topic of deep learning systems for the classification of food and their nutrition composition. Section 3 will focus on discussing the properties of the datasets which contains the different food dishes that are to be used to train the deep learning neural network of the proposed system. Section 4 details the major features of the proposed system, including the proposed framework of the deep learning based application. Finally, this paper concludes by summarising the work and highlighting the future direction.

2. Related Work

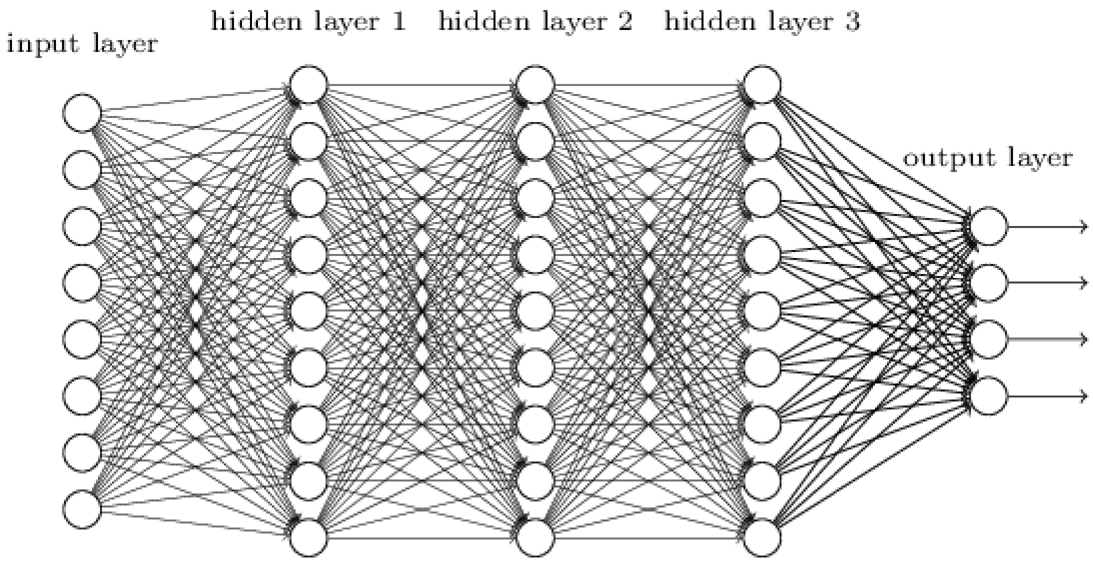
The following section is where critical analysis on the topic of food classification using deep learning techniques is conducted. Scientific research articles from published journals related to the topic is sourced and its contents reviewed so that we establish a better understanding on the topic at hand.

2.1 Monitoring of Food Nutrients Composition

Food nutrients composition refers to the chemical components that make up any kind of food that can be consumed. In general, we classify nutrients composition in food into the following major categories; carbohydrates, proteins, fats, vitamins and minerals with the addition of fibres. It is important for us human beings to monitor our food nutrition composition of what we eat to fully understand the effects they bring to our body.

2.2 Deep Learning Artificial Neural Networks

Within the scope of Machine Learning, Deep Learning concerns algorithms that takes inspiration from the complex structure of the human brain to help solve various real world problems. These algorithms have been coined as the term Artificial Neural Networks. Neural networks are computational models that have been designed to recognise patterns in data by interpreting raw sensor inputs through labeling and clustering. Figure 2 shows an example of an artificial neural network that consists of many layers of neurons or nodes. The general aim of data scientists to do research in the development of deep learning is to progress towards computers learning to naturally think like a human. The reason on why deep learning in used for this kind of application is because of the high accuracy of recognition which is essential. We have found that there are various types of artificial neural networks that have been established to compute and solve various problems. Such neural network models include Feedforward Neural Network, Convolutional Neural Network (CNN), Radial basis function Neural Network, and Recurrent Neural Network (RNN) to name a few (Jordan, & Mitchell, 2015). Figure 2 below shows an example of a graphical representation of a basic Artificial Neural Network (ANN).



DEEPNN.PNG. CONTROL SYSTEMS TECHNOLOGY GROUP WIKI
EINDHOVEN UNIVERSITY OF TECHNOLOGY (TU/E)
DEPARTMENT OF MECHANICAL ENGINEERING

Figure 2: Example of an Artificial Neural Network Concept Graphical Representation

2.3 Existing Research

According to Pouladzadeh et al. (2016), in their work they have proposed an application of deep learning for food classification and recognition. The issue highlighted is how to calculate the food and energy consumption of consumers effectively and provide them with strategies that help them fight preventable diseases such as obesity. They had used a Food Frequency Questionnaire (FFQ) as their methodology to obtain data from prospective users on their daily diet. FFQ is a survey of the level of consumption of a small number of food products (Maes, & Vereecken, 2003). The image bank they used to train their model was a pub;ic web service called FoodLog. Other than that, the framework that the authors use are what they call they Deep Belief Network which is used in their Android application. The advantages that they highlighted in their research is that their method could identify food portions very precisely within 3 seconds, and that their results show that they have 99 percent accuracy in their single food portions of their algorithm. The disadvantages, on the other hand, using their algorithm, the calorie estimate is determined on the basis of the food image captured by the consumer, which is the only known variable, and the food image captured from a closer range has a greater food dimension, and vice versa has led to an incorrect calorie estimation. As a result, demonstrating the use of novel combinations of graph-cutting segmentation and deep-learning neural networks as a means of accurate classification and recognition of food items (Pouladzadeh et al., 2016).

Furthermore, in the work provided by Min et al.(2017), the authors proposed to model both the recipe attributes and the multimodal content information within a unified recipe modelling framework. The proposed framework is intended to capture the underlying rich correlations between different recipe knowledge and to expand the scenarios of conventional recipe-oriented problems to three application problems, which are; multimodal cuisine classification, attribute-augmented cross-modal recipe image retrieval, and ingredient and attribute inference from food images. The issue they had been attempting to address is that food varies in many ways that contribute to many different food tastes as well as variations, and so the challenge is to identify cuisine and extract recipes from food photos. The framework that the authors established in their work is to propose a MultiModal MultiTask Deep Belief Network (M3TDBN) model to address the issue of low dependency between visual content and textual components, as well as cooperation between attributes. Apart from that, they gathered real-world food data in a dataset called Yummly-28k to assist in their work.The benefit found in this research is that by integrating the culinary parameters into their algorithm, the baseline results surpassed 5% to 10%. Further enhancements of a 16 percent growth in output from the baseline are then observed when multiple algorithm passes are implemented. One deficiency in this work was that more evaluation metrics were needed to improve their outcome. This resulted in a model that explores multimodality content and multi-attribute information in the food domain to success (Min et al., 2017).

Separately, in the research conducted by the same authors (Min et al., 2018), their objective is to carry out the first cross-regional study of recipes using ingredients, food images and their attributes such as cooking and meal. The problem they wanted to address is that recipes of different cuisines shared on the web are an indication of culinary cultures in different countries and, thus, an examination of these recipes can contribute to a deep understanding of food from a cultural point of view. To solve this, they propose a framework for the study of the culinary culture in order to discover the patterns of the ingredients bases and visualize them in order to enable various applications which they called Bayesian Cuisine-Course Topic Model (BC2TM). Their research methodology consists of conducting systematic analyses of the real-world recipe dataset called Yummly-66K. Here they present with three applications including multi-modal cuisine summarization, cuisine-course pattern analysis, and cuisine recommendation. The advantage of their approach is that the implementation of cooking or course knowledge contributes to improved performance. The disadvantage that was identified is that there were limited evaluation of multiple and cross-modal recipe. They have achieved their goal of unveiling the diversity of culinary arts from different regions they sample (Min et al., 2018).

Moreover, Wang, Min, Li, & Jiang, (2016) have proposed a model to learn the semantics of the dish and the identity of the restaurant. Their problem was that it is difficult for tourists to perceive the identity of restaurants and dishes separately in their food choices. The algorithm the authors came up with is the Partially Asymmetric Multi-Task Convolutional Neural Network (PAMT-CNN) that includes the dish pathway and the restaurant pathway to learn the dish semantics and the restaurant identity, respectively. The benefit of their model is that using two distinct types of information in separate pathways in their algorithm increases the efficiency and accuracy of all multi-task deep model approaches compared to other models already available. Apart from that, interactions between two pathways with lower layers allow for more robust layer features in the model. One disadvantage to their approach was that only the accuracy was calculated, where other metrics were needed to b e analysed to prove their findings. The PAMT-CNN model used to address the problem of simultaneous dish and restaurant recognition from food images was successful (Wang, Min, Li, & Jiang, 2016).

In addition, EMMANUEL & MINIJA (2018), they suggested a method that determines the calorie content of each food item from the user evaluation of dietary images in three stages. which are; segmentation, feature extraction, and food recognition.The concern was that the calorie value of people's food consumption needs to be monitored in order to reduce the risk of obesity, heart problems and diabetes. Therefore, the challenge is in determining the calorie value is the identification of food items in the food picture category. Food items should be separated from nutrition and non-food items in order to determine the exact calorie count. The algorithm that is proposed in their research is the Whale Optimization Algorithm (WOA), which performs classification of the food items from the image. Attributes such as colour, shape, and texture have been extracted from the segmented images to create a vector database of attributes. From that the dietary assessment system finds the calorie value of the recognized food items obtained from the WLM-NN classifier. The steps in the methodology is as follows; Preprocessing, Segmentation of the food image, Feature extraction, and lastly Classification Estimation of calorie value.The benefit of their approach is that the overall segmentation and classifier model obtained the best MAA metric with a value of 0.9643 in their algorithm and outperformed the other existing models. The shortage in their study is that there were limited variety in images used in the evaluation. The result of the study, the authors have achieved their goal of assessing the calorie value is the recognition of the food items from the group of food images to high success (EMMANUEL & MINIJA, 2018).

Furthermore, in the study by Cioccaa, Napoletanoa, & Schettinia, (2018), their goal is to explore the use of features based on the Convolutional Neural Network for the purpose of food recognition and retrieval. Their problem as they explained, is that there needs to be a way to help patients keep track of their food consumption in a more user-friendly manner that allows for more detailed regular dietary monitoring. The approach used is to first create a comprehensive food database that combines multiple already available food datasets called Food-475, and secondly, to analyse and categorise datasets using CNN-based features related on their food-domain representativeness. They use a Residual Network with 50 layers as a reference architecture to learn food-domain features. The advantage they have accomplished is that the higher the importance of the food-domain representativeness of the database, the more features they have acquired, which greatly improves the precision of the retrieval relative to other approaches. This allows for more relevant images to be returned in the first position. The downside to this approach is that the evaluation techniques in terms of metrics need to be improved. They concluded by presenting an assessment of CNN-base functionality for food classification and retrieval, as well as introducing a new Food Database, Food-475, which significantly improves on previous iterations (Ciocca, Napoletano, & Schettini, 2018).

Moreover, according to Mori et al (2012), they introduced a machine learning approach to the problem of text processing of recipes, aimed at translating the text of recipes into a workflow. The question they have found is that providing a multitude of recipes is not always a good thing, as consumer-made recipe text uses a variety of phrases and writing styles that discourage consumers from discovering new cooking processes for a meal. The methodology that they use in their work is divided into three natural language processing (NLP) tasks: word recognition, named entity recognition, and syntactic analysis. The benefit of their strategy is that they achieve high precision in word recognition as a result of increased data training. Apart from that the model that they produced is robust.The downside of the model is that the variations between the text structure of the sentence and the general text used to train the NLP modules cause problems such as linguistic anomalies and predicate-argument structure connections. The authors have achieved their goal of presenting a machine learning approach to recipe text processing aiming at converting a recipe text into a workflow (Mori, Sasada, Yamakata, & Yoshino, 2012).

Besides that, In the research conducted by Bolaños, Valdivia, & Radeva (2018), they propose to explore the problem of image-based food menu recognition. The problem here is if given an image, to ascertain the specific menu item corresponding to the restaurant where it was taken to be able to match the image to the menu item, it would be easier to find the exact nutritional information of the food or any other data stored by the restaurant owners. On a broader scope the problem they want to help solve is the lack of basic knowledge or awareness in most people regarding their eating habits which leads to a risk in coronary heart diseases. To combat that, the authors proposed a model that uses two different inputs, one in the form of an image, an the other in the form of text, to be fed into a three different convolutional neural networks (CNN) with a penultimate layer called a InceptionResNetV2. Then, by using LogMeal's API4, that predict the dish, food group (or family) and the ingredients detected in the image. The advantage of this is that the machine learns a language model, considering a large number of possible names associated with the images given. On the other hand, the downside is that it does not take into account the handling of dishes with foreign names which cannot be easily learned by their language model. The conclusion of the study is that it is possible to build a food restaurant menu recognition model for any restaurant, without the need to have a separate model for each restaurant or restaurant pair (Bolaños, Valdivia, & Radeva, 2018).

Other than that, according to the research conducted by Liu et al. (2016), they has formulated a deep learning food image recognition algorithm to tackle the problem of recording dietary caloric intake in order to manage weight loss. The problem highlighted stems from the fact that personal dietary assessment relies primarily on memory to recall previous meals. They came up with an algorithm which they call DeepFood which recognises food images through the use of deep learning techniques notably the Convolutional Neural Networks (CNN). Their proposed approach uses the datasets UEC Food-100 and 256 as their repository of food information. The advantage of the method was that it exceeded all other approaches and achieved high classification accuracy. The downside of the approach on the other hand is that the training of the large model with the heavy datasets lead to a lengthy process time. This all resulted in an algorithm that uses CNN to effectively assess dietary intake of users through capture images of their food (Liu et al., 2016).

According to the study by Riloff, Chiang, Julia, & Jun'ichi, (2018), The authors proposed the development of a mobile application framework that recognises the name of the food and returns the corresponding details on the plate. The problem arises from the fact that due to awareness in fitness in Japan, people care more about the ingredients and their nutritional information that make up what they eat. Another challenge is that when foreigners come to Japan, it is quite hard to choose food because of the language barrier. Therefore the authors built an application that recognises the name of a food in a sentence of a menu, and returns an image of the dish that is corresponded to it as well as their ingredients, cooking method and nutritional information. The framework that they used for their system is the Inception-ResNet-V3 Architecture with the ETH Food-101, UEC FOOD 100 and UEC FOOD 256 as their datasets. The benefit of their model is that it has less training time and more reliable food recognition compared to existing models. The disadvantage of the model is that when introduced with incomplete data of food, results in a partial mapping of the network (Riloff, Chiang, Julia, & Jun'ichi, 2018).

In a separate study that was done by Chokr and Elbassuoni (2017), they propose to help make it easy for users to calculate information regarding the food they eat. This is due to the fact that more and more people have a goal of maintaining a healthy diet and would need to track their calories in their diet, but entering in the calorie information is tedious and not an efficient use of time. The methodology that they use for their approach is the Mathworks Image Processing Toolbox and the Principal Component Analysis (PCA) method. The dataset available to them is the Pittsburgh fast-food image dataset. The advantage of their model is that they have been able to distinguish between two different food products and precisely measure the amount of calories in the food. The disadvantage on the other hand is that it is only able to predict the calories of one of the dishes at a time. The result is that they have succeeded in predicting the calorie amount of certain foods based upon their images provided (Chokr & Elbassuoni, 2017).

2.4 Tabulation of Findings of Existing Research

Table 1 below shows the extracted objectives and findings of each of the related work

Table 1: The objectives of each research article and their subsequent findings respectively.

Objective	Findings	Citation
To propose an application of deep learning for food classification and recognition.	Using a combination of graph cut segmentation and deep learning neural networks, results in accurately classifying and recognising food items	(Pouladzadeh et al., 2016)

To model both the recipe attributes and the multimodal content information within a unified recipe modelling framework.	They proposed a MultiModal MultiTask Deep Belief Network (M3TDBN) model to address the problem of weak dependence between visual content and textual ingredients, as well as the collaboration among different attributes	(Min et al., 2017)
To perform the first cross-region recipe analysis by jointly using the recipe ingredients, food images and attributes.	They presented three applications including Multi-modal Cuisine Summarization, Cuisine-course Pattern Analysis, and cuisine recommendation to reveal the diversity of culinary arts	(Min et al., 2018)
To propose a model to learn the dish semantics and the restaurant identity.	The use of the PAMT-CNN model used to address the problem of simultaneous dish and restaurant recognition from food images.	(Wang, Min, Li, & Jiang, 2016)
To propose a system that finds the calorie value of each food item from the image for dietary assessment	By using the Whale Optimization Algorithm (WOA), which performs classification of the food items from the image, they managed to find the calorie value of each food item in three steps which are; segmentation, feature extraction, and food recognition	(EMMANUEL & MINIJA, 2018)
To investigate the use of Convolutional Neural Network based features for the purpose of food recognition and retrieval.	They used a Residual Network with 50 layers as a reference architecture to learn food-domain features where the more the features learned greatly improves the precision of the retrieval. This allows for more relevant images to be returned in the first positions.	(Ciocca, Napoletano, & Schettini, 2018)

To propose a machine learning approach to recipe text processing problem aiming at converting a recipe text to a workflow.	They divided the task into three natural language processing (NLP) tasks: word recognition, named entity recognition, and syntactic analysis to achieve high accuracy in word recognition for the recipe text processing.	(Mori, Sasada, Yamakata, & Yoshino, 2012)
To explore the problem of image-based food menu recognition.	The result is that it is possible to build a food restaurant menu recognition model for any restaurant, without the need of having a separated model per restaurant or restaurant pairs.	(Bolaños, Valdivia, & Radeva, 2018)
To propose a deep learning food image recognition algorithm to solve the challenge of documenting dietary caloric intake	They came up with an algorithm which they call DeepFood which recognises food images through the use of Convolutional Neural Networks (CNN). This is effective at assessing dietary intake of users	(Liu et al., 2016)
To build the application to recognize the food name in the sentence and return back the food image corresponding to the same name with food ingredients, cooking method, and nutrition.	They designed a model that shows what kind of food, how the food looks like, as well as its nutritional information, from just selecting the name of the dish in the mobile app.	(Riloff, Chiang, Julia, & Jun'ichi, 2018)
To propose to help make it easy for users to calculate information regarding the food they eat.	Success in predicting the calorie amount of certain foods based upon their images provided.	(Chokr & Elbassuoni, 2017)

The following Table 2 shows the algorithms and their advantages and disadvantages of each related work discussed.

Table 2: The algorithms used in each research paper including their advantages and/or disadvantages of their use.

Algorithm	Advantage	Disadvantage	Citation
Deep Belief Network in Android Application	99% accuracy in recognising the food portions in 3 seconds.	Determining the dimension of the food portion based on the image captured is challenging.	(Pouladzadeh et al., 2016)
MultiModal MultiTask Deep Belief Network (M3TDBN) model	Better performance compared to the best baseline, and outperforms it by 16%. The course and cuisine attributes enforce each other in a multitask fashion.	Evaluation metrics need to be improve.	(Min et al., 2017)
Generative process of Bayesian Cuisine-Course Topic Model (BC2TM)	Introduction of the cuisine or course information leads to the improvement of performance.	Limited evaluation of multiple and cross-modal recipe.	(Min et al., 2018)
Partially Asymmetric Multi-Task Convolutional Neural Network (PAMT-CNN)	Outperforms all other baselines in dish recognition and restaurant recognition. Has 1% improvement compared with the best baselines. Enables more robust lower layer features.	Only the accuracy was calculated. Other evaluation metrics need to be analysed to prove the findings.	(Wang, Min, Li, & Jiang, 2016)
Whale Optimization Algorithm (WOA)	The best MAA metric with the value of 0.9643 and has outperformed the other existing models	Limited varieties of images were used for evaluation.	(EMMANUEL & MINIJA, 2018)
Convolutional Neural Networks (CNN)	More the features learned, greatly improve the precision. Allows more relevant images to be returned in the first positions.	Evaluation metrics need to improve.	(Ciocca, Napoletano, & Schettini, 2018)

Natural language processing (NLP) tasks: word recognition, named entity recognition, and syntactic analysis.	Highly accurate and robust.	Differences between sentence structure or dictionary example sentences, causes problems. Problems of some linguistic phenomena arose.	(Mori, Sasada, Yamakata, & Yoshino, 2012)
InceptionResNetV2 CNN	Learns huge amount of names and associates them to their corresponding pictures. It takes a completely new restaurant's menu and picture of the menu items and find the correct menu item.	Cannot easily learn dishes with exotic names.	(Bolaños, Valdivia, & Radeva, 2018)
Convolutional Neural Networks (CNN)	Outperforms all existing approach. High classification accuracy.	Training a large model requires a large amount of time.	(Liu et al., 2016)
Inception-ResNet-V3	Provides less training time. More accurate at recognising the food.	Incomplete data of food, results in a partial mapping of the network.	(Riloff, Chiang, Julia, & Jun'ichi, 2018)
Principal Component Analysis (PCA) method.	Able to distinguish between the two different food. Accurately predict the amount of calories	Can only predict the amount of calories for one food item at a time.	(Chokr & Elbassuoni, 2017)

Table 3 as shown below shows the technology framework of each related work with the datasets that they used for their respective models.

Table 3: The technology frameworks used in each research and the datasets used respectively.

Technology Framework	Datasets	Citation
Deep Belief Network	FoodLog	(Pouladzadeh et al., 2016)
MultiModal MultiTask Deep Belief Network (M3TDBN)	Yummly-28k	(Min et al., 2017)
Cuisine-Course Topic Modelling and Visualization Bayesian Cuisine-Course Topic Model (BC2TM)	Yummly-28k Yummly-66k	(Min et al., 2018)
Partially Asymmetric Multi-Task Convolutional Neural Network (PAMT-CNN)	Dianping AlexNet	(Wang, Min, Li, & Jiang, 2016)
Wavelet kernel based Wu-and-Li Index Fuzzy clustering (CSW-WLIFC) Whale Levenberg Marquardt Neural Network (WLM-NN	UNIMIB2016	(EMMANUEL & MINIJA, 2018)
Average pooling (AVG-POOL)	ResNet-50	(Ciocca, Napoletano, & Schettini, 2018)
Natural language processing (NLP)	The Balanced Corpus of Contemporary Written Japanese (BCCWJ)	(Mori, Sasada, Yamakata, & Yoshino, 2012)
InceptionResNetV2 LogMeal's API Word Embedding matrix	Food-101 UEC Food256 Vireo-Food	(Bolaños, Valdivia, & Radeva, 2018)
Convolutional Neural Network (CNN)	UEC-100 UEC-256 DeepFoodCam Project Food-101	(Liu et al., 2016)
Inception-ResNet-V3 Architecture	Food-101 UEC FOOD 100 UEC FOOD 256	(Riloff, Chiang, Julia, & Jun'ichi, 2018)

2.5 Convolutional Neural Network (CNN) for Pattern Recognition Training

From the related work that have been reviewed, we have decided to choose Convolutional Neural Network (CNN) as the artificial neural network algorithm for the training to recognise and classify the food images into their nutrients composition. The reason for the choice of CNN as our technology framework is the highly accurate classification predictions that is generated from the input of datasets of food images. From the related work, it has been established that CNN has the highest performance neural network algorithm at handling recognition of patterns from image data as compared to other forms of neural networks that have been tested.

3. Food Datasets

The datasets are a crucial component to any deep learning model. This is because you need a large dataset of raw data to feed and teach to the model in order for the neural network to learn and identify the patterns that exist between the inputs of the data as well as to compare it to other inputs. The more the quantity of raw data is input into the system, then the more patterns the model can recognise, resulting in a higher accuracy and precision of the predictions. The following datasets have been selected to be used to teach the proposed deep learning model.

3.1 Food-101

This dataset is one of the most popular food images datasets to be used in machine learning algorithms. It consists of 101,000 images of food that have been collected from Foodspotting program conducted by the Federal Institute of Technology Zurich. It also includes 250 test images that have been manually reviewed including an additional 750 training images. This dataset is suitable for our application of the deep learning techniques.\

3.2 UEC-256

This dataset consists of 256-kind food photographs that comprise mostly of Japanese dishes. The benefit of this dataset is that each picture of the food has a bounding box that shows the location of the food item within the image. This makes it easier for the model to recognise and classify the food item from the image.

3.4 Yummly-28K

The Yummly-28K recipe dataset is a recipe forward multimodal food collection of data. It includes images of food with the addition of the name of the recipe, ingredients, cuisine, and course type included corresponding to the image. This dataset is useful to evaluate multimodal recipe retrieval., ingredient inference and cuisine classification.

4. Proposed Deep Learning Framework

The framework for our deep learning model is the guidelines that need to be followed in order to achieve an accurate prediction and classification of the food images that is input. Figure 3 shows our proposed framework that we intend to use to develop our nutrients composition classification model.

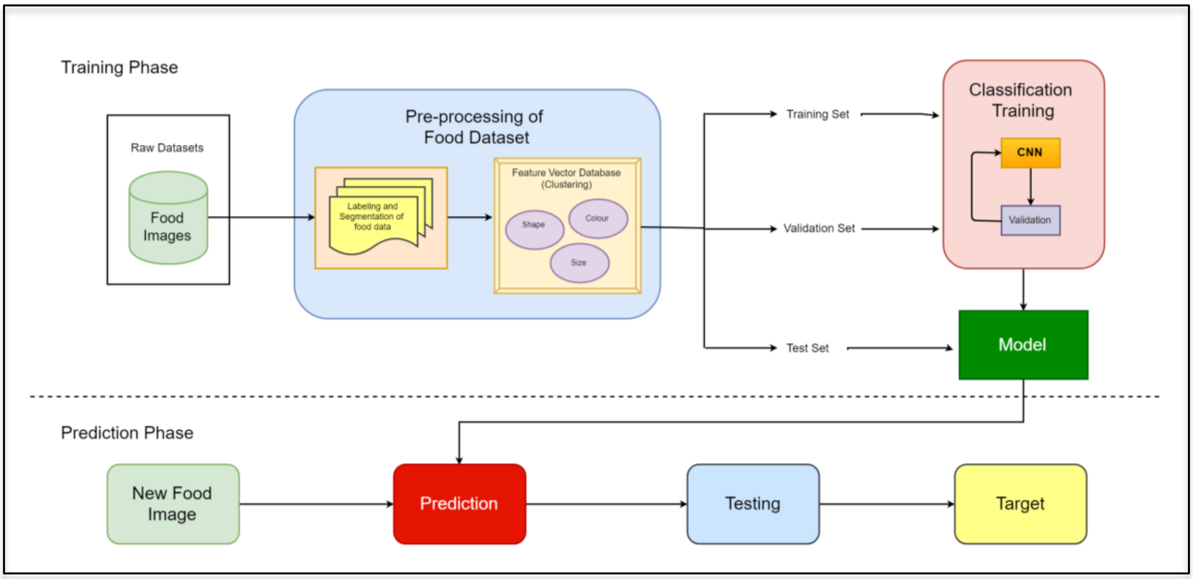


Figure 3: The proposed framework for deep learning food nutrients composition classification

4.1 Pre-processing of Food Dataset

We must first pre-process the food datasets by manually customising and fine tuning the data before it is pushed through the model itself to achieve our desired output. This is also known as perimeter optimisation. Once you have decided the appropriate datasets to use, the raw data within the datasets must be standardised according to our needs before being presented to the model to be taught. Raw data is not always perfectly organised and some may be incomplete, therefore you need to pre-process the datasets to avoid incompatibility and minimise errors.

For our case, to begin the data pre-processing phase, the datasets of food images and information is compiled and then manually labelled according to the type of data each of them represent, ie. Image file, text file or etc. Moving on, the data is segmented according to their respective data type. Only once those processes are completed can the clustering process begin. Here the data that has been labelled and segmented is separated and organised into the specific characteristics they possess. For images, the physical characteristics of the food is recognised such as their shapes, colour and etc. All these processes are done in the beginning to facilitate the training of the model later on.

After the datasets have been pre-processed, duplicate sets are made for training, validation, and testing respectively. The training set is used to be pushed through the convolutional neural network model to determine patterns. The validation set is used as the fixed variable and is compared with the trained model. The test set is used at the end of the training process to test the final output of the model.

4.2 Classification Training Model using Convolutional Neural Network (CNN)

Deep Learning primarily uses Artificial Neural Network (ANN) to create and train a deep learning model, The model or framework is the core feature to be used for our deep learning model. It is the brains of our system where the ANN is situated. The training and the classification of the food items is done within the neural network. There are many different types of ANNs but for our case, the type of

ANN that is to be used to create our model is Convolutional Neural Network (CNN). The most important aspect of the model is the use of the CNN within the classification training portion of the framework, to train and learn from the datasets introduced.

Convolutional Neural Network is a deep learning algorithm that primarily utilises the use of images as the main inputs in the network. CNN, or sometimes refers to as ConvNet, can take images and assign objects within the images with vectors such as weights or biases, that can be used to differentiate them with other images. Therefore the potential applications of such image recognition is boundless.

CNN manipulates the Convolutional Layers of the image to extract the weightage value of the objects present in the image and transforms into matrix of values that the machine can understand and train from. The mathematical model for Convolutional Layers is as follows:

$$\mathbf{x}_j^\ell = f\left(\sum_{i \in M_j} \mathbf{x}_i^{\ell-1} * \mathbf{k}_{ij}^\ell + b_j^\ell\right),$$

This mathematical model is derived from the backpropagation model to calculate the feature maps that are convolved from previous layers where the activation function is formed to output a feature map (Bouvrie, 2006).

The process of classification training and validation of the dataset is completed multiple times to increase the model's accuracy of the predictions. Each image taken from the dataset runs through the CNN and patterns are recognised by passing through the nodes or neurons of the neural network. The CNN consists of multiple layers of neurons with the addition of the input and output nodes situated at both the beginning and end of the network respectively. Having multiple nodes and layers inside the CNN increases the probability of success of recognising the patterns within the images of the food. Once the model has been trained and the image patterns can be recognised, the result of the process moves to the prediction phase of the framework.

4.3 Prediction

Now that a trained model has been established through the training phase, the prediction phase of the deep learning framework can be done. We can introduce new data to the trained model for it to recognise the patterns that exist, compare it to the patterns that it has learned, and predict what the data may be. In this case, the new raw data is in the form of a new image of a food dish which can be shot from a camera sensor. This image's patterns such as shape and colour, is compared to the patterns that the model has established during training, for their similarities. The model then may predict the type of food and their nutrient composition of the captured image from the similarities that exist between them.

4.4 Testing

To ensure that the deep learning model behaves according to the requirements that have been established, testing needs to be conducted. The test set that have been duplicated previously is used as a controlled variable for the model. An image from the test set can be introduced to the trained model to observe the predictions made from it. To ensure that a correct and accurate prediction is output by the model, the prediction of the new image is compared to the control prediction to establish if the prediction is accurate or not. This test is conducted multiple times with different images from the test set and new images and the results are tabulated. The accuracy factor is determined and calculated from these results in the form of a percentage with the higher value as being more accurate prediction. The accuracy of the predictions from the model is tested after each time a batch of image data is pushed through the CNN over 10 iterations. The goal is to achieve a high percentage of prediction accuracy. If the accuracy of predictions are higher than 75%, then the deep learning model is considered as a success and the model can be deployed in a web application to predict the food nutrients composition from images. However if the accuracy factor falls lower than this threshold, the process of training the neural network should be conducted once again.

5. Conclusion and Future Work

In this paper, we identified the major problem of people needing to identify and classify their food according to the nutrients composition so that they may have a clear understanding of their diet and make necessary decisions to improve their health. Furthermore, we have discussed in detail the related published studies relating to the topic of food image classification using deep learning techniques and extracted the many pros and cons of the techniques and algorithms that they have used. We have also determined the potential datasets that contain food images and information to be used in our deep learning model and the benefits they bring to the model. We propose a framework for the implementation of classification of food nutrients composition using deep learning. The use of Convolutional Neural Network is the basis of recognition and classification of food images to discern patterns from images of food.

For the future work, the model is to be developed and implemented according to the framework and datasets as stated. We shall use Convolution Neural Network as the algorithm to conduct the training and implementation of the deep learning model to make predictions on food nutrients composition. After data pre-processing, the chosen datasets, Food-101 UEC-256, and Yummly-28k, shall be used to train the model where patterns and characteristics of the food images are distinguished over multiple passes of the neural network. Once that is completed, testing is done to ensure the model achieves accuracy in prediction and fit for deployment in a food nutrients classifier web application.

References

Bolaños, M., Valdivia, M., & Radeva, P. (2018). Where and What Am I Eating? Image-Based Food Menu Recognition.

Bouvrie, J. (2006) Notes on Convolutional Neural Networks. *Center for Biological and Computational Learning Department of Brain and Cognitive Sciences Massachusetts Institute of Technology Cambridge, MA 02139*

Ciocca, G., Napoletano, P., & Schettini, R. (2018). CNN-based features for retrieval and classification of food images. *Computer Vision and Image Understanding, 176*, 70-77.

EMMANUEL, W. S., & MINIJA, S. J. (2018). Fuzzy clustering and Whale-based neural network to food recognition and calorie estimation for daily dietary assessment. *Sādhanā, 43*(5), 78.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. doi:10.1126/science.aaa8415

Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., & Ma, Y. (2016). Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. *Paper presented at the International Conference on Smart Homes and Health Telematics.*

Maes, C. A. V. a. L. (2003). A Belgian study on the reliability and relative validity of the Health Behaviour in School-Aged Children food-frequency questionnaire. *Cambridge University Press, 6(6).*

Min, W., Bao, B.-K., Mei, S., Zhu, Y., Rui, Y., & Jiang, S. (2018). You are what you eat: Exploring rich recipe information for cross-region food analysis. *IEEE Transactions on Multimedia, 20*(4), 950-964.

Min, W., Jiang, S., Sang, J., Wang, H., Liu, X., & Herranz, L. (2017). Being a supercook: Joint food attributes and multimodal content modeling for recipe retrieval and exploration. *IEEE Transactions on Multimedia, 19*(5), 1100-1113.

Mori, S., Sasada, T., Yamakata, Y., & Yoshino, K. (2012). A machine learning approach to recipe text processing. *Paper presented at the Proceedings of the 1st Cooking with Computer Workshop.*

Pouladzadeh, P., Kuhad, P., Peddi, S. V. B., Yassine, A., & Shirmohammadi, S. (2016). *F*ood calorie measurement using deep learning neural network. *Paper presented at the 2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings.*

Wang, H., Min, W., Li, X., & Jiang, S. (2016). Where and what to eat: Simultaneous restaurant and dish recognition from food image. *Paper presented at the Pacific Rim Conference on Multimedia.*

Chokr, M., & Elbassuoni, S. (2017). Calories prediction from food images. *Paper presented at the Twenty-Ninth IAAI Conference.*

Riloff, E., Chiang, D., Julia, H., & Jun'ichi, T. (2018). Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. *Paper presented at the Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.*