

SOLENT UNIVERSITY SOUTHAMPTON
FACULTY OF BUSINESS, LAW AND DIGITAL TECHNOLOGIES

Master's Project
Master's Thesis and Graduate Research

**MSc Applied AI and Data Science
2022**

Surya Gopu Pillai - Q15784037

***“Computer Vision and Speech
Recognition - Auto Recipe detection
and alert system using AI”***

SOLENT UNIVERSITY
FACULTY OF BUSINESS, LAW AND DIGITAL
TECHNOLOGIES

**MSc Applied AI and Data Science
Academic Year 2021-2022**

Author Name: Surya Gopu Pillai
Student ID: Q15784037

Computer Vision and Speech Recognition - Auto Recipe
detection and alert system using AI

Module: Dissertation (COM726)

Supervisor: Drishty Sobnath

September 2022

Full-Time Student

This report is submitted in partial fulfilment of the requirements of Southampton Solent University for the degree of MSc Applied AI and Data Science.

List of Contents:

1 Acronyms	5
2 Abstract	6
3 Acknowledgement	6
4 Chapter 1: Introduction and Background	7
4.1 Introduction	7
4.2 Background Theory and Justification	8
5 Research question, Aim and Objective	10
5.1 Research question	10
5.2 Aim	10
5.3 Objective.....	10
6 Chapter 2: Literature Review.....	10
7 Chapter 3: Neural Network for Image Classification	11
8 Algorithms for Image Classification	13
9 Proposed artefact and social impact	15
10 System concepts and methodology.....	16
11 Project Implementation.....	18
11.1 Data Collection	18
11.2 Data Preparation (Data Pre-Processing)	18
12 Speech Recognition	25
13 Flutter Application and Front End	26
14. Challenges and Limitations	30
15 Evaluation, Results, and Discussions	32
15.1 Evaluation	32
15.2 Results	33
15.3 Discussions.....	35
16 Conclusion.....	36
17 References	36
18 Bibliography	39
19 Appendix C	39

List of Figures:

Figure 1: Neural Network Architecture	12
Figure 2: Convolutional Neural Network Architecture	13
Figure 3: End-to-End Workflow	16
Figure 4: Importing all Libraries.....	19
Figure 5: Reading the dataset	19
Figure 6: Image Rescaling and normalization.....	20
Figure 7: Lables.txt file creation.....	20
Figure 8: Prediction Labels for an existing class	20
Figure 9: Neural Network Graph with Shared Inputs	21
Figure 10: Neural Network Graph with Multiple Inputs	22
Figure 11: Building a sequential model.....	22
Figure 12: Compiling the model	23
Figure 13: EarlyStopping, ModelCheckpoint, ReduceLRonPlateau	24
Figure 14: Model Fitting	24
Figure 15: Training and Test loss with Epochs	24
Figure 16: Checking the accuracy of the model	24
Figure 17: Converting Keras model to TemnsorFlow Lite	25
Figure 18: App.py - Image Scanning.....	26
Figure 19: Speech Recognition.....	26
Figure 20: Running the Flutter application	27
Figure 21: Splash Screen.....	28
Figure 22: Home Screen	28
Figure 23: Preview Page: Allergens detected.....	29
Figure 24: Preview Page: Allergens not detected	29
Figure 25: Preview Page - with speech recognition.....	30
Figure 26: Accuracy score of SVM, KNN, Random Forest and Decision Tree algorithm	32
Figure 27: Model comparison with confusion matrix	33
Figure 28: Model Comparison	35
Figure 29: Ethics application approval	39

Index of Tables:

Table 1: Results with the accuracy rate	34
---	----

1. ACRONYMS:

AI: Artificial Intelligence

API: Application Programming Interface

BoF: Bag of Features Model

CNN: Convolutional Neural Network

DNN: Deep Neural Network

GPU: Graphics Processing Unit

gTTS: Google Text-to-Speech

GUI: Graphical User Interface

iOS: iPhone Operating System

KNN: K-Nearest Neighbour

ML: Machine Learning

OCR: Optical Character Recognition

OOP: Object-Oriented Programming

PIL: Python Imaging Library

SDK: Software Development Kit

SVM: Support Vector Machine

TXT: TeXT (File Extension)

.jpeg: Joint Photographic Experts Group (File Extension)

.jpg: Joint Photographic Experts Group (File Extension)

.png: Portable Network Graphics

2. ABSTRACT:

Food allergies happen when the body's immune system responds oddly to a particular food. Although allergic reactions are frequently mild, they still pose a threat. Food intolerance and food allergy are not the same things. Symptoms of food intolerance can be brought on by issues with the same substances, including lactose, being digested. However, there are no allergic reactions. A large amount of food must be consumed than is necessary for an allergy to be triggered before experiencing symptoms of food intolerance. Contrary to allergies, a food intolerance never poses a life-threatening risk. Our project aims to create an Android-based (mobile-based) application that can identify British items accurately. The system will identify the allergens present in the ingredients used and will alert the user, it will also assist the user in knowing the calorie intake. The system is an extension for visually challenged people or people who require assistance to read and write. The major concern is the system configuration and training time as we are working with images, thus restricting our scope to eleven British items. The methodology includes creating an android / iOS application using flutter as a front end, classification of images using the CNN model, inclusion of speech functionality and checking for allergens, and determining calorie intake. We have limitations in building a complete application. Millions of images are required per class to build a model capable of predicting appropriately. It's very difficult to achieve this as the existing system configuration is very low compared to the requirements and the system requires GPU. Our future work will include creating a complete application that can accurately predict the British food items displayed in a food festival or an event. This application is beneficial to any people using this product to scan the food displayed in the food event or food festival. It will help the user in alerting any allergic components as well as the calorie count included in the dish.

3. ACKNOWLEDGEMENT:

The author would like to acknowledge all the sources used for data collection. Images are collected from imageSearchDownloadAll, Unsplash website, and google images. Edamam API is used for ingredients and calorie intake. The author has created an account to get the data from Edamam API.

4. Chapter 1: Introduction and Background

4.1 Introduction:

Human health depends on healthy eating (Salim et al., 2021). People become increasingly concerned with keeping track of their weight, eating healthy, and avoiding allergens, thus, there is an growing need for calorie counting. Consuming too many calories might harm you and cause a number of ailments (Shen et al., 2020). It is common for people to suffer from food allergies and intolerance/physical reaction to food, which affects a substantial level of people in society (8% children and 2% adults on average). It is not necessary that an individual should be directly impacted, many of us still have family, friends, or relatives who are impacted. World Allergy Organization (WAO) oversees the term used to identify allergies, but due to various complexity and lack of proper clarity, it becomes very difficult to understand. To be more peculiar, allergies mainly caused by food are thought to be the most difficult variety of all allergies. Food allergies are becoming more common, and eating is a daily routine of every individual, the commonness of food allergies is increasing fast. An individual with a food allergy always amends the contents in the food which is a bit uncomfortable for the ones who have been diagnosed recently with food allergy. Thus, their lives are more difficult; hence, people use the internet to research allergen-safe recipes, giving them access to a safe variety of foods. Allergens in food can be deadly at times. It is very difficult to maintain track of foods containing possible allergens considering the variety of components used and the availability of food products. To ensure an acceptable level of protection, some amendments to the components are required. Most online recipe websites are not labelled with allergen-safe components, also the allergic categories come with varied components not easy to remember. In order to make sure that all allergic components are overlooked, it can take time to research specific sensitive recipes on the website. Thus, we are introducing a system with the objective of automatically detecting components/ingredients, and allergens with calorie consumption based on the recipes. This approach uses the Edamam nutrition information database to offer further thorough information about the component's nutritional value used in the recipes.

Edamam is a food and grocery database API for searching food and diet related data. The Edamam Food and Grocery Database API is designed to assist application developers, web developers etc., and incorporate a quick and simple way to seek food database information into software or websites (Fang, 2022). Nutritional analysis API enables one to understand the nutritional information of any recipe by copy-pasting them. Food database API gives free accessibility to approximately 9,00,000 foods and over 6,80,000 unique UPC codes. Recipe Search API allows more than 2.3 million recipes searchable by diets, calories, and nutrient range.

Not all details are present in the database about possible allergens linked with every content, and to the best of my knowledge, a reference database exists that is used to label components with associated allergens which is a significant issue for the system as uploading a new recipe each time should detect them accurately. There should be no error by the system while doing this classification

as this could cause serious potential issues impacting users who suffer from food allergies. Hence, the solution we are proposing to this system can automatically detect food, ingredients, calorie intake, and alert message for allergic food.

4.2. Background Theory and Justification

During our research, it is found that deaths from severe allergic responses, also known as “anaphylaxis”, were caused by declined food during the past 20 years, this is an analysis from National Health Service (NHS) UK data. Thus, there is an increase in the hospital admission of patients with induced anaphylaxis. Scientists from Imperial College London have conducted an analysis and revealed that among school-aged children, cow’s milk is the commonest single source of deadly food induced allergic reactions which are published in the BMJ. However, anaphylaxis related deaths are very rare. In the UK, it is estimated that less than 10 fatalities are attributed to food allergies every year. 9% of people in the US are affected by anaphylaxis. Allergen labelling is improving over the years, specifically with the introduction of laws, as per business regulations allergen labelling is required by the company for any food which has contributed significantly to these changes. A food business must now label its product with details of 14 main food allergies mentioned by EU legislation. We are introducing a pre-screening system acting as a protective layer against allergies. This study demonstrates the process by which a dataset is created and used by machine learning to bifurcate cuisines and allergens.

Computer vision requires an image dataset. The model training processes take a long time when working with images. Images are converted to the numerical format. The accuracy rate depends upon the number of images supplied to every dish. Therefore, we must define our domain and we intend to concentrate on British food Festival. The scope of the project is focusing on eleven British items, but its scope is expandable in the future to cover a large area and create a full system.

Some of the dessert items are “Battenberg cake” and it was made as a wedding gift for Prince Louis of Battenberg and Princess Victoria. It was named after a member of the English royal family by utilizing his last name (Day, Day and profile, 2022). The next item is “Brandy Snaps”. In Britain, the brandy snaps are a favorite treat, golden flaky tubes that are crisp and filled with brandy whipped cream. It is a sweet dish that is typically served around Christmas and the brandy is an optional content (British Brandy Snaps, 2022). The next one is “Bread and Butter Pudding”, a traditional bread pudding in British cuisine. The fourth one is “British Scones”. Scones aren’t typically served for breakfast in the UK. They are offered tea as a quick bite in the morning with cream tea or with afternoon tea (Ozimek, 2022). The fifth British item is “Carrot Cake”. Carrot cake’s ancestry is not entirely known. It is said to have originated in England during World War II where housewives utilized sweet items to naturally sweeten their confections. It appears to be evolved from the carrot pudding of Medieval times, given that sugar and other sweeteners were scarce or expensive to come by in Britain and carrots had long been used as sugar substitutes (The History of Carrot Cake, 2022). The sixth one is “Chelsea bun”, a popular British snack. The buns are from the 18th century in the Chelsea neighbourhood of West London (Chelsea bun | Definition, Ingredients, & History, 2022). The next British dessert item is “Classic Shortbread

Cookies”. Shortbread’s origin can be traced to the 12th century when it was first known as “biscuit bread” - cookies prepared from leftover bread dough that was occasionally sweetened and baked until they turned into a tough dry rusk. This practice existed not just in Scotland but throughout the British Isles (Shortbread, 2022). The next one is “Strawberry Fool”, a traditional British dessert known as a fool initially invented in the late 16th century. Most fools are made with whipped cream instead of custard. Any fruits can be used in the fool. However, rhubarb, gooseberries, and strawberries are the most common choices (Strawberry Parfait (Strawberry Fool), 2022). The other item is “Summer Pudding”, sometimes known as summer fruit pudding is a traditional British dessert composed of sliced white bread piled with fruit and fruit juice in a large bowl. Its current name first appears in 1904, but earlier versions of the same recipe for “hydropathic pudding” and “Malvern pudding” date as far back as 1868 (Summer Pudding, 2022). The next one is “Swiss Roll” which is a sponge cake with a filling, typically a jam. Each slice exposes a spiral design when we cut across the end. In the UK and Canada, it is typically referred to as a “Swiss Roll”. The sponge cake is typically white, but there are chocolate and other flavour variations (Swiss Roll - CooksInfo, 2022). We have “Trifle” which is a cold dessert and “Scottish tablets” as another item. There are many varieties of desserts and British cuisines that we can embed with our system. However, the system needs billions of data to identify the dish accurately, hence with the aim to create a system that can predict the items appropriately, we have crated a dataset or classes which is a mixture of both desserts and other British cuisines as desserts appear similar and training the model to capture the differences is challenging since we are working with a restricted number of datasets. With the existing system configuration, this is very challenging. End users can get health-related information such as calorie intake for these desserts and if someone is allergic to any of the contents used in them.

Classification, Clustering, and Feature Selection

The effectiveness, stability, and utility of classification algorithms are increased by using a few features in the classification process. Therefore, selecting relevant features is very important. There are various options available to select operations including wrapper techniques, vector supporters, feature selection prototypes, and neural networks that are similar to our own method. The goal of selecting a clustering function is to maximize variety which can be done in two basic steps; first feature space creation selecting appropriate distance measurement; second, features are grouped to generate selection result in a clustering method, a representative feature of each class is labelled (most relevant feature). The feature set was hierarchically clustered in line with the correlations and the best feature (final feature) is selected using the packaging approach from each cluster. This strategy doesn’t require parameter adjustment, it lengthens the time and biased learning process in the introduction of the packaging method. Integrate sparse K Means and hierarchical clustering in a feature cluster framework. A set of functions are grouped first into a number of functional clusters. The representative features making the final subset of features are chosen from every cluster by Lasso type penalty factor, using the supervised learning environment’s agglomerative hierarchical clustering to partition feature set and finalizing the subset after removal of features far from others. Relativity of the subset, correlation effect

and selection of centroid as defining characteristics of each cluster's cluster can be measured using maximum data coefficient. The availability of high dimensional data on the internet is significantly high during the previous several years. Machine Learning systems are finding it difficult to handle numerous input features with great interests to researchers. Pre-Processing is required to apply machine learning techniques 5 efficiently. Feature selection is one of the most important and essential feature and crucial component of machine learning, this is the popular and significant methods of data preparation also referred as a variable selection characteristic.

5. Research Question, Aim, Objective

5.1. Research Question:

How can the feature of AI like computer vision and speech recognition benefit any individual to ensure they are having allergy safe cuisines in a food event?

5.2. Aim:

The aim is to create a system capable of recognizing British items displayed in a food event, which will automatically detect the components/ingredients used in the dish with the calorie intake and alert the user about any allergens. This system can be accessed by any individual including visually challenged people. The system helps to prevent health complications or sudden food reactions in a user.

5.3. Objective:

- a) Build a mobile interface that can be handy and easily accessible by every individual including visually challenged people.
- b) Usage of Computer vision in pattern recognition by scanning the image.
- c) Enabling speech recognition feature (Text to speech functionality)

6. Chapter 2: Literature Review:

(Liu et al., 2022) conducted a study providing computer-aided technical solutions that will enhance and increase the precision of current food intake measures. The author has used a deep learning-based algorithm to recognize food images as the main technological novelty. Significant research has shown that digital imaging accurately predicts dietary intake in a variety of settings and has many advantages over alternative techniques. The effective extraction of food information (such as food type, and portion size) from the images is still difficult and unsolved research subject. With few optimizations, the author has suggested a Convolution Neural Network (CNN)-based strategy. The results of experiments have shown the effectiveness of this method in solving the food image recognition issue.

(Anthimopoulos et al., 2014) conducted a study to estimate the carbohydrate content in the food served to a diabetic individual using computer vision-based food recognition. This study suggests a bag-of-features (BoF) model-based methodology for automatically identifying foods. The best-performing components in BoF design were identified and optimized, and related parameters were

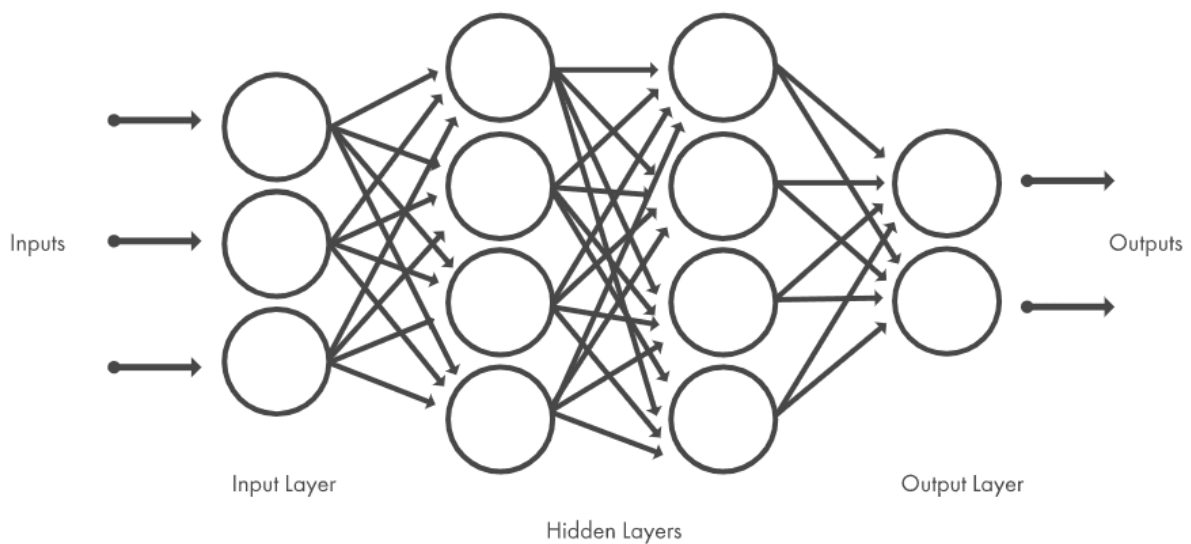
estimated through a technical estimation. 11-class visual datasets with almost 5000 images were produced for the design and evaluation of the prototype system. The system was able to classify the images with an accuracy of 78% demonstrating the viability of the suggested approach in a highly difficult image collection. (Johan and Rizal, 2021) conducted a study to create an allergen recognition system helpful for customers who are very cautious about the food they choose to eat taking into account that some ingredients can behave differently in different adults. The author aims to build an android based system to identify food ingredients to assist customers in learning about all allergens present in the recipe followed. The application is made by utilising the Boyer Moore algorithm for word matching and Optical Character Recognition (OCR) technique. The work demonstrated 90% accuracy with various scenario applications. One of the suggestions to this artefact was to add text to speech feature to help people who lack the ability to read and write. (Mishra et al., 2022) carried out a study to understand the impact of food allergies in terms of health. Every individual is allergic to different components. The allergic reactions range from minor hives or itching to anaphylaxis proving fatal. Such reactions can be avoided by knowing the food that contains allergens and simply avoiding them. The author has aimed to train an object detection model based on deep learning to recognize the presence of such items in a picture. The author presented the Allergen30 dataset containing over 6000 annotated photos of 30 frequently consumed foods that can cause an unpleasant reaction. (Ross, Bremer and Nielen, 2018) conducted a thorough overview of immunochemical food allergy assays and detectors in the critical assessment focusing on how user-friendliness when connected to smartphones. Food allergens are a major global health concern; thus, customers should be able to check their food for allergens whenever and wherever they are. It is critical to create quick, safe, and easy tests to be used with cell phones increasing user accessibility. This study has summarized traditional laboratory-based methods determining the potential to modernise the techniques by integrating with smartphones for food allergen detection such as enzyme-linked immunosorbent assay, flow cytometry, and surface plasmon resonance. (Jessica Kent, 2022) carried out a study using a deep learning algorithm to recognize allergic reactions in hospital patient safety records, assisting doctors in preventing medical errors and enhancing event surveillance. According to the study, up to 36% of patients report allergies due to drugs, 4% to 10% allergies to foods according to researchers. Because patients in healthcare settings are more likely to experience an allergic reaction, it's crucial that healthcare professionals can rapidly recognize and track these episodes. The goal was to create a deep learning technology recognising allergic reactions. Furthermore, when compared to keyword-search methodologies, the deep learning methodology detected 24.2% more cases of verified allergic reactions and bring down the number of cases that require manual review by 63.8%.

7. Chapter 3: Neural Network for Image classification:

Deep learning is a more sophisticated branch of machine learning that handles data processing using multiple layers of algorithms to create abstractions or to

mimic the thinking process. It is frequently used to comprehend spoken language and distinguish objects visually. Each layer transmits information to the one underneath it, with the output of one layer serving as the input of another. The input layer and the output layer are the first and the last layers respectively in a network. Hidden layers are the ones which exist between input and output. Typically, each layer is a straightforward algorithm with a single type of activation function.

Image recognition is the process of recognizing images and classifying them into many of the several predetermined different classes also known as image classification. As a result, image recognition software and apps can identify objects in an image and bifurcate one object from another. Computer vision uses image recognition as a foundation.



*Figure 1: Neural Network Architecture
(Image Recognition with Deep Neural Networks and its Use Cases, 2022)*

Deep Neural Network excels in recognising images. Computer programs called neural networks are designed to spot and identify patterns. The architecture consists of three different layers i.e., input layer, hidden layer, and output layer. The input layer receives a signal, the hidden layer processes it, and the output layer makes predictions using the input data. The interconnected nodes carry out the calculations. The system must understand the features in order to identify them. They must be trained to determine the object. Machine learning models are trained and learn the characteristics in a different way compared to Deep learning models. Because of this, various model training strategies have been developed. We use feature engineering to build a deep learning model to distinguish between any objects, for instance, a dog and a cat. We are unable to design precise characteristics that will work for every potential image while taking into account challenges like viewpoint dependent object variability, background clutter, lighting conditions and image distortion. Because of neural networks, an alternative approach should be possible. There is no need for manual feature extraction because neural networks learn features directly from the data they are

trained on (Image Recognition with Deep Neural Networks and its Use Cases, 2022). Neural network requires massive amount of labelled data. Thus, there is an increase in the training time.

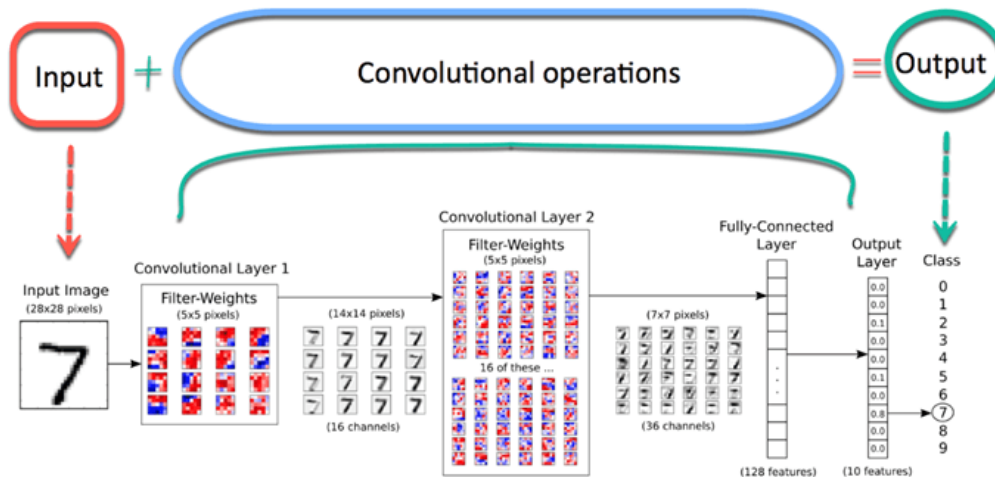


Figure 2: Convolution Neural Network Architecture (Johnson, 2022)

Convolutional Neural Networks were first called when Yann LeCun and his team created the LeNet (paper). It was primarily designed for the handwritten digit recognition problem between 1989 and 1998. During the convolution step, an input image is analysed and then given a label. The above image is the architecture of CNN. The input image is sent to the network. The convolution portion of the network then subjected the input image to an indefinite number of steps. The neural network can also anticipate which digit will appear in the image. The convolutional layer is the most important part of the model, which seeks to shrink the size of the image to speed up weight computations and increase generalization. The network retains the critical aspects of the image during the convolutional phase and filters out unimportant noise. For instance, the model is training itself to identify an elephant in an image of a mountain. When using a typical neural network, the model will give all the pixels including those from the mountain, a weight, which is not necessary and can lead to a network error. Instead, just the most pertinent pixels will be extracted using a mathematical method by a Keras convolutional neural network. Convolution is the name of this mathematical operation. The network may learn progressively complex information at each layer. The matrix is divided into smaller bits by convolution in order to discover the key components of each piece (Johnson, 2022).

8. Algorithms for image classification:

- **Support Vector Machine:**

SVM (Support Vector Machine) is a fantastic classification algorithm. It is a supervised learning technique utilized for classifying data into several categories. A set of labelled data is used to train SVM. SVM's primary benefit is that it can be applied to both classification and regression. SVM creates a decision boundary or a

hyperplane to distinguish different classes. SVM algorithm is used in Image classification and object detection. However, SVM has certain limitations while working with image classification. Producing clear output in image classification typically relies on the presumption that a pixel belongs to only one class. SVM doesn't work well, or it is unsuitable for large datasets. SVM cannot handle noisy data. With overlapping target classes and when the dataset has more sound, SVM doesn't operate well. The support vector machine will perform poorly when there are more attributes for each data point than training data specimens.

- **K-Nearest Neighbour:**

The K-Nearest Neighbour algorithm referred to as KNN or k-NN is a supervised learning classifier that employs proximity to produce classifications or predictions about the grouping of a single data point. A number k is selected by KNN algorithms as the nearest neighbour to the data point that needs to be categorized. Let us assume K to be 4, in this instance, KNN learns the four neighbours that are closest. KNN doesn't go into detail about the class-specific attributes. It identifies the variations of every pixel, but it is not capable of finding the features. KNN performs better with column data serving as attributes i.e., tabular data where each attribute defines a particular characteristic, however, in the case where images are required, pixel values do not define specific characteristics.

- **Random Forest Algorithm:**

A technique for classification and regression called random forests was first developed for the machine learning community. This technique is typically used for the classification of satellite and aerial images as well as the development of continuous field data sets, such as biomass and percent tree cover. Before training the model, the three primary hyperparameters for random forest algorithms must be established. The size of the node, number of trees, and sample size are the three hyperparameters. The random forest classifier can then be utilized to address classification or regression issues. In general, random forest operates well with tabular data, meaning, there isn't a strong qualitatively significant relationship between the features such as when the data is an image, or the observations are connected on a graph. Numerous rectangular partitions often do not accurately resemble these architectures. Random Forest is less expensive to compute and does not need a GPU to complete training. A decision tree can be interpreted differently yet more effectively by a random forest. For neural networks to be truly effective, far more data will be needed than the average person may have handy.

- **Decision Tree:**

The decision tree algorithm is frequently used in data mining. The input data set is categorized into specified classes using a decision tree, a tree-based classification technique. The image recognition system is trained to execute supervised machine learning using a decision tree technique. Algorithms built on decision trees play a significant role in the classification process. The primary benefit is that there is no assumption made on the distribution of the data and are typically quite quick to compute. The main benefit is, they are relatively simple to understand and depict non-linear data patterns. Additionally, they are quicker, especially when analysing

exploratory data. Decision trees provide the best accuracy score for tiny datasets. The decision tree is relatively instable in comparison to other decision predictors. A minor change in the data can have a significant impact on the decision tree structure, which can transmit a different outcome than what users would typically receive.

- **Neural Network:**

Deep neural networks have numerous layers, which makes it possible to models to perform more demanding computational tasks, i.e., performing many difficult and complicated operations simultaneously, which helps in learning more complex properties. CNNs (Convolutional Neural Networks) are frequently used for image classification in deep learning. The identification and extraction of features from images are made possible using deep learning. The ability of the deep learning approach is to carry out feature engineering on its own. In this method, an algorithm searches the data to find traits that correlate, combines them, and promotes faster learning without being specifically instructed to do so.

SVM, KNN, Random Forest algorithm, decision tree are machine learning algorithms which give constant result every time. However, Neural networks are artificial intelligence algorithms which has the scope of improvement and can give varying results every time on training the model by making some changes in the dataset and parameters as required.

9. Proposed artefact and societal impact:

We proposed a unique model using machine learning and artificial intelligence to detect recipes accurately. The proposed system will be an android application that will identify British items displayed in a food event with ingredients in it with an alert message if the food is allergic. The proposed system is also beneficial to visually challenged people as it will have text-to-speech functionality which will give voice output. We will be using flutter SDK (Software Development Kit), dart programming language, and Android Studio IDE to create the application and gTTS API for converting speech to text. Flutter is Google's free, open-source SDK (Software Development Kit) for creating cross-platform mobile applications. Flutter provides programmers with a lot of advantages that expedite the development and result in more streamlined mobile applications. We aim to create a society with allergen free food by providing an allergen information to the customers directly. If an individual has food allergy, it is very important that the individual has the required information to make safe food choices and thus a healthy environment.

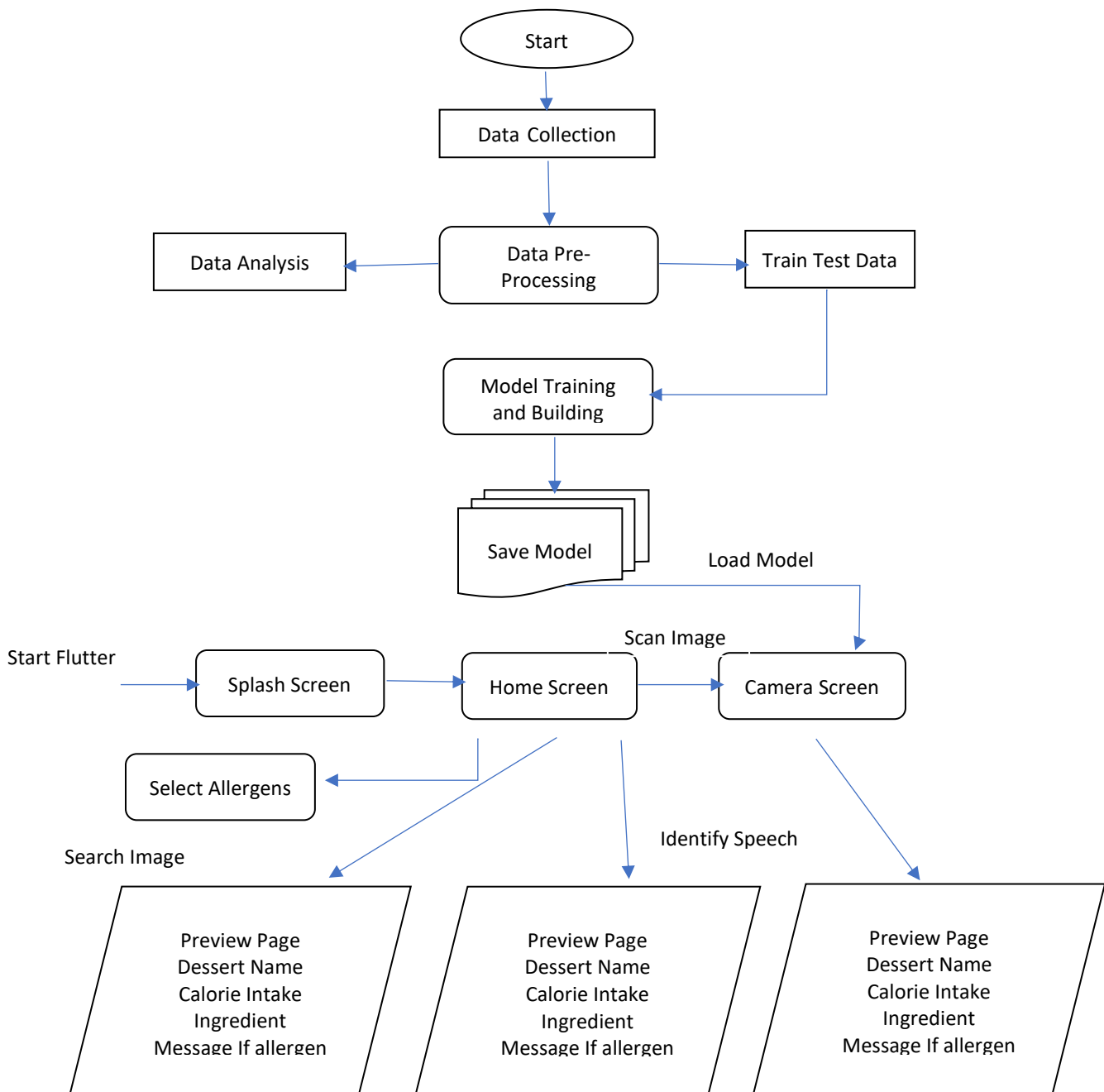


Figure 3: End-to-End Workflow

10. System Concept and methodology

The requirements for our proposed system include the following:

- Classification of images:

Determining the scope and finalizing the name of dishes or more specifically British items are defined. Every dish is defined as a class. Thus, we have eleven different classes. Our model should be able to classify all the eleven dishes defined in the scope. Image classification is a technique used to

categorize or forecast the class of a particular object in an image. The main goal of this technique is to identify the characteristics of an image appropriately.

- **Checking for allergens and determining calorie intake:**
All dish's ingredients are looked up using the Edamam API. It's crucial to check the calories and allergens in every meal before consuming it to maintain a healthy habit.
The intensity of allergens varies for every individual. There could be different recipes followed for every dish. Thus, it is crucial to check for allergens rather than generalizing the components. Knowing the probable allergy triggers could save our lives because an allergic reaction can be lethal if appropriate medical assistance is not received. Health problems can be caused by consuming too few or too many calories over time. By counting the food calories, we may establish how much potential energy it has. Just as important as the calories themselves is the material from which they are obtained.
- **Android/iOS application using flutter:**
Selecting an appropriate GUI (Graphical User Interface) as a front end is very important. There are various alternatives available to create the GUI. Flutter is considered to be the best option while developing a mobile application. It has many features, inbuilt libraries and packages that can create beautiful GUI. A single codebase may be used by developers to create native apps for both iOS and Android using Google's mobile UI framework, Flutter. Flutter uses Dart programming language which is quite similar to any other Object-Oriented Programming (OOP) language. The code is natively compiled and utilizes the GPU to produce a consistent user interface providing native access to platform APIs like GPS and Bluetooth. Moreover, with Flutter everything is possible with widgets, the foundation of the flutter app. In addition to being simple and resulting in shorter development cycles, widget building allows the reuse of previously created widgets across several applications. Additionally, widgets provide consistent rendering on android and iOS devices as well as consistent APIs for developers to access device native services (Srivastava, 2022).
- **Speech functionality**
The proposed system should be able to detect recipes and find ingredients from recipes and allergens from ingredients. The ability of a machine or program to recognize words spoken aloud and translate them into legible text is known as speech recognition, often known as speech-to-text. It is the capacity of a machine to hear spoken words and recognize them. Python's speech recognition software can then be used to translate the spoken words into text, ask a question, or respond. Even some equipment can be programmed to react to those spoken words. With the use of computer programmes that accept input from the microphone, process it and convert it into a useful form, we can perform voice recognition in python. SpeechRecognition is an inbuilt library available in python which can be easily installed through pip install SpeechRecognition

11. Project Implementation:

11.1. Data Collection:

The practice of obtaining and analysing information from a wide variety of sources is known as data collection. If we were to use the data that we collect to develop practical Artificial Intelligence (AI) and machine learning solutions, it must be collected and maintained in a way that makes sense for the particular business problem at hand. AI is a group of technology that are excellent at drawing conclusions and patterns from vast amounts of data. These observations and trends can be used by AI to forecast what causes outcomes. Even better, it can gradually learn to make better predictions. For computer vision technology, we require images as a dataset. Images should be in .jpg, .jpeg, .png format. Initially, 150-200 images were collected for every dish. Thus, we had a dataset of approximately 2000 images in total. However, with this number of images, the training accuracy rate is very less, and the system was unable to recognize the dish. Thus, the number of images is increased, and 700-800 images are collected for every dish. But there is a concern here. We have already used maximum images for every dish and there is a limited image available online. Also, the training time of the model depends upon GPU and system configuration. Working with images is very challenging in terms of data availability, data quality, and system configuration. The initial approach was to download images using the imageSearchDownloadAll application. However, the images should be collected manually and does not support by chrome. Image download should be done on mobile with limited images. Thus, we decided to add the “Download All Images” extension to chrome to download multiple images. A useful Google chrome browser plugin called Download All Images allows downloading several images from any website at once. The quality of images downloaded were not satisfactory, hence we are using only good quality images from there and other images are collected from Unsplash. Unsplash is a website or a photo discovery platform that is free to use. They provide high-definition images. Edamam nutrient database will be used to get information about components of the recipe and calorie intake.

11.2. Data Preparation (Data Pre-Processing):

Data Pre-Processing, which is a component of data preparation, refers to any sort of processing carried out on raw data to prepare it for another data processing action. Historically, it has been an important step in the data mining process. The quality of the data and the information that can be gleaned from it directly affect how well the model can learn, making data pre-processing an essential step in machine learning process. It is essential that we pre-process the data before supplying it to the model for this reason. Raw data i.e., data from the real world is inherently incomplete and cannot be used directly by the model. This could lead to some errors. For this reason, pre-processing of data is important before sending it through a model. Only the .jpg, .jpeg, and .png image formats are used; all other picture formats are eliminated in order to reduce data noise and improve precision

(i.e., webp). The model only accepts numerical data; hence the images are transformed to NumPy array format.

Image Resizing: Images collected during the data collection process vary in size, and hence they cannot be stacked in an array or used as a part of machine learning methods. By interpolating pixel color values, the resulting scaled image can begin a prolonged transition. Bicubic interpolation can be deployed for this because despite being more expensive than other interpolation approaches, it produces superior results. A cropped image with the same height and width will be the output.

Data normalization: Feature scales should be divided by 255 to obtain values ranging from 0 and 1. By taking into account keeping 99% of the differences, normalization is used for dimensionality reduction to speed up training without abandoning trivial details.

The datasets are divided into two groups for machine learning. The first subset referred to as training data is a section of the actual dataset which is used to train the machine learning model. The other subset is the testing dataset. We can reduce the effects of data discrepancies and gain a better understanding of the model's properties by using similar data for training and testing. A model is tested by making predictions against the test set after it has been trained using the training set. Data Pre-Processing starts by downloading all the required libraries and importing them.

```
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import tensorflow as tf
import cv2
import imageio
import pathlib
%matplotlib inline

import cv2
import imghdr
import os
```

Figure 4: Importing all Libraries

After importing all the libraries, the dataset should be read from the directory. Images with extension .jpg, .jpeg and .png are only considered.

```
data_dir = 'dataset'
image_exts = ['.jpeg', '.jpg', '.png']

for image_class in os.listdir(data_dir):
    for image in os.listdir(os.path.join(data_dir, image_class)):
        image_path = os.path.join(data_dir, image_class, image)
        try:
            img = cv2.imread(image_path)
            tip = imghdr.what(image_path)
            if tip not in image_exts:
                print('Image not in ext list {}'.format(image_path))
                os.remove(image_path)
        except Exception as e:
            print('Issue with image {}'.format(image_path))
            # os.remove(image_path)
```

Figure 5: Reading the dataset (Renotte, 2022)

PIL (Python Imaging Library) library is imported. Python Imaging Library is a free and open-source extra library that enhances support for accessing, modifying, and saving a wide range of image file formats. It works with Windows, Mac OS X, and Linux. Only specified image formats are allowed and removing other formats. This is done for both the train and test datasets. Train and test datasets are split into an 80-20 ratio.

All file paths that fit a certain pattern are returned using the glob (short form for global) function. We can use glob to search for a specific file pattern, or perhaps more usually, we can utilize wildcard characters to search for files where the filename matches a given pattern. In our case we are looking for all images with .jpg, .jpeg and .png extension.

The “ImageDataGenerator” class is imported that allows randomly rotating images across any angles between 0 and 360 degrees by supplying an integer value in the rotation range argument. Images are rescaled, target size is set for both train and test data with a good batch size of 32, and class mode as categorical is used as the parameters.

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator
rescaled = ImageDataGenerator(1/255)
trained_fed = rescaled.flow_from_directory('dataset/train', target_size = (128,128), batch_size = 32, class_mode='categorical')
test_fed = rescaled.flow_from_directory('dataset/test', target_size = (128,128), batch_size = 32, class_mode='categorical')
```

Figure 6: Image Rescaling and normalization

Labels.txt file is created to store the dish names or dataset index that we are using. This file is used while building the flutter application. It is used in the assets folder.

```
# trained_fed.class_indices

labels = '\n'.join(sorted(trained_fed.class_indices.keys()))
with open('labels.txt', 'w') as f:
    f.write(labels)
```

Figure 7: Labels.txt file creation

```
: trained_fed.class_indices
: {'Battenberg Cake': 0,
  'Brandy Snaps': 1,
  'Bread and Butter Pudding': 2,
  'British Scones': 3,
  'Carrot Cake': 4,
  'Chelsea Bun': 5,
  'Classic Shortbread Cookies': 6,
  'Strawberry Fool': 7,
  'Summer Pudding': 8,
  'Swiss Roll': 9}
```

Figure 8: Prediction Labels for an existing class

Deep learning models can be created using the Sequential model API by first creating an instance of the Sequential class, followed by the addition of model

layers.

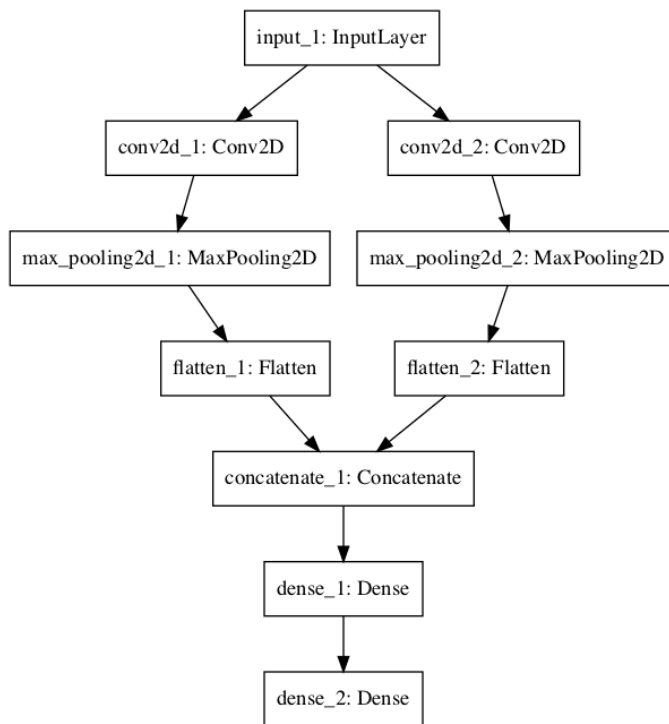


Figure 9: Neural Network Graph with Shared Inputs (Brownlee, J., 2022.)

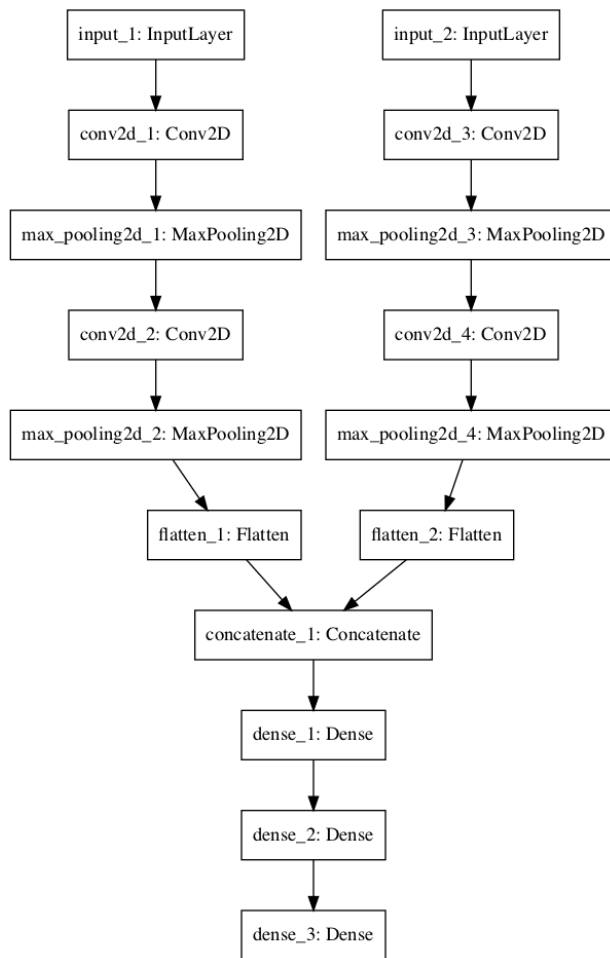


Figure 10: Neural Network Graph with Multiple Inputs (Brownlee, J., 2022.)

For a simple stack of layers where each layer has precisely one input tensor and one output tensor, a sequential model is sufficient. A sequential model can be created by passing a list of layers to the sequential constructor. After building the model, we can call the `summary()` method to display the contents.

```

model = tf.keras.models.Sequential([tf.keras.layers.Conv2D(16, (3,3), activation='relu', input_shape = (128,128,3)),
    tf.keras.layers.MaxPool2D(2,2),

    tf.keras.layers.Conv2D(32, (3,3), activation='relu'),
    tf.keras.layers.MaxPool2D(2,2),

    tf.keras.layers.Conv2D(64, (3,3), activation='relu'),
    tf.keras.layers.MaxPool2D(2,2),

    tf.keras.layers.Conv2D(128, (3,3), activation='relu'),
    tf.keras.layers.MaxPool2D(2,2),

    tf.keras.layers.Flatten(),

    tf.keras.layers.Dense(128, activation = 'relu'),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(10, activation = 'softmax')
])

```

Figure 11: Building a sequential model

Optimizers are used to increase speed and training performance for a particular model. There are various optimizers available. We are using Adam optimizer. A stochastic gradient descent technique called Adam optimization is based on an

adaptive estimate of first-order and second-order moments. Most of the time, Adam is the most effective adaptive optimizer. They are good with sparse data and the datasets used are ideal for adaptive learning rate. While compiling the model, we are using the loss function “categorical_crossentropy” which is used when there are two or more output labels in a multi-class classification model. One-hot category encoding values in the form of 0s and 1 are allocated to the output label. If the output label is in the integer form, Keras transforms it into category encoding. Metrics are defined as ‘accuracy’. A metric is a calculation that is used to assess how well the model is working. Metric functions are comparable to loss functions, with the exception that during model training, the outcomes of a metric’s evaluation are not taken into account.

```
: from tensorflow.keras.optimizers import Adam
model.compile(loss = 'categorical_crossentropy' , optimizer = 'Adam' , metrics = ['accuracy'])
```

Figure 12: Compiling the model

How well a machine learning model generalises to data that is similar to that on which it was trained is determined by how well it fits the data. Results from the model that is well fitted are more precise. A model that is overfitted too closely resembles the data sufficiently. We can use COMPILE to compile the model once we have finished writing all of the statements in it. We can use COMPILE to assist in debugging the code before running the model because it looks for format issues during compilation. Compilers translate human-written programming languages into binary code that can be executed by computer hardware. Compilers operate on vast, intricate, diverse, non-deterministic, and dynamic systems. Compiler optimization is challenging since there are so many potential improvements.

We are importing EarlyStopping. The number of training epochs to utilize can be challenging while developing a model. Too many epochs cause overfitting of the model and using a few epochs causes underfitting. Early Stopping is a technique enabling us to define an arbitrarily large number of epochs and terminates the training as soon as the model performance stops increasing. This is a regularisation method for a deep neural network that halts the training once parameter updates stop producing gains on a validation set.

ModelCheckpoint - When training the model, the ModelCheckpoint callback is used to save a model or set of weights at regular intervals. The models and weights can then be loaded later to continue training from the saved state.

ReduceLROnPlateau - Imported to reduce learning rate when a metric stop improving. Once learning reaches a plateau, the model frequently gains by decreasing the learning rate by a factor of 2-10. This callback keeps track of a quantity, and it slows down learning if there hasn’t been any improvement for a predetermined number of epochs. This can be used before applying the EarlyStopping callback if the loss is not reducing for a while. This could cause the loss to start decreasing at a slower pace of learning. It’s possible that loss will continue to increase.

```

from keras.callbacks import EarlyStopping, ModelCheckpoint, ReduceLRonPlateau
import h5py

early_stop = EarlyStopping(monitor = 'val_loss', patience = 6, restore_best_weights = True)
model_checkpoint = ModelCheckpoint(filepath = 'modelcheckpoints/my_tfkerasmodel1.hdf5', monitor = 'val_loss', save_best_only=True)
lr_rate = ReduceLRonPlateau(monitor = 'val_loss', patience= 3, factor = 0.1)

```

Figure 13: EarlyStopping, ModelCheckpoint, ReduceLRonPlateau

```

history = model.fit(trained_fed, shuffle=True, epochs=20, validation_data = test_fed, callbacks = [early_stop, model_checkpoint]

```

Figure 14: Model Fitting

```

# plotting results for loss

plt.plot(history.history['loss'], color = 'blue', label = 'train')
plt.plot(history.history['val_loss'], color = 'orange', label = 'test')
plt.grid(True)
plt.title("Train & Test Loss with epochs",fontsize=12)
plt.xlabel("Training epochs\n",fontsize=12)
plt.ylabel("Train and Test Loss", fontsize=12)
plt.show()

```

Figure 15: Training and Test Loss with Epochs

An epoch refers to one full iteration of the algorithm over the training dataset. An essential hyperparameter for the algorithm is this epoch's number. The amount of complete runs of the whole training dataset during the algorithm's training or learning process is specified in terms of epochs.

```

: accuracy = model.evaluate(test_fed, steps=len(test_fed), verbose=1)
print("Models accuracy")
print('%.2f'%(accuracy[1]*100))

```

Figure 16: Checking the accuracy of the model

The next step is converting the Keras model to TensorFlow lite. TensorFlow lite is a suite of tools that allow developers to run their trained models on mobile, embedded, and IoT devices and desktops, enabling on-device machine learning. Platforms including embedded Linux, Android IOS, and MCU are supported.


```

from tensorflow import keras
model = keras.models.load_model("my_model")
converter = tf.lite.TFLiteConverter.from_keras_model(model)
tflite_model = converter.convert()

# Save the model.
with open('model_unquant.tflite', 'wb') as f:
    f.write(tflite_model)

```

Figure 17: Converting Keras model to TensorFlow lite

The model is saved as “model_unquant.tflite” to be utilized in the flutter application.

12.Speech Recognition:

The ability of a machine or program to recognize words spoken aloud and translate them into legible text is known as speech recognition, often known as speech-to-text. Basic speech recognition software can only pick up words that are uttered clearly and has a small repertoire. More advanced software can handle diverse languages, accents, and natural speech. Research in computer science, linguistics, and computer engineering are all used in speech recognition. Speech recognition features are built into a lot of contemporary gadgets and text-focused software to facilitate easier or hands-free usage. Voice and speech recognition are two distinct technologies that shouldn't be mixed. In order to recognize words in spoken language speech recognition is used. Voice recognition uses biometric technologies to recognize a specific person by their voice (Lutkevich, 2022). When we scan the images, they are saved to the URL for the flask server. The speech recognition feature returns the image of the dish that is provided as input along with the calorie intake and the ingredient list. We are importing the speech recognition library from machine learning. While training the model, we are saving another copy of the model as “my_tfkerasmodel1.hdf5” which is used in the speech recognition portion i.e., app.py file.

```

@app.route('/', methods=["GET"])
def indexpage():
    return render_template('index.html')

@app.route('/upload', methods=["POST"])
def uploads():
    if(request.method == 'POST'):
        imagefile = request.files['image']
        filename = werkzeug.utils.secure_filename(imagefile.filename)
        imagefile.save("./uploadedimages/" + filename)

        classifier_model = tf.keras.models.load_model(r'my_tfkerasmodel1.hdf5')

        img = Image.open(imagefile).convert('RGB') # get RGB PNG image
        img = img.resize((128,128))

        img = np.array(img) # Convert to numpy array
        img = np.expand_dims(img, axis=0) # add an extra dimension for batch

        class_names = ['Beef Wellington', 'British Scones', 'Chelsea bun', 'Lancashire hotpot', 'Mince pies', 'Scottish Tablet', "She

```

Figure 18: App.py - Image Scanning (Build Developers, How to upload image from Flutter to Flask server, 2022)

```

p.route('/speech', methods=["POST"])
speechRec():
    if(request.method == 'POST'):

        audiofile = request.files['audio']

        # Create an instance of the Recognizer class
        recognizer = sr.Recognizer()

        # Create audio file instance from the original file
        audio_ex = sr.AudioFile(audiofile)
        # type(audio_ex)

        # Create audio data
        with audio_ex as source:
            audiodata = recognizer.record(audio_ex)
        # type(audiodata)

        # Extract text
        text = recognizer.recognize_google(audio_data=audiodata, language='en-US')

        # print(text)

        return jsonify(

```

Figure 19: Speech Recognition (Build Developers, How to upload image from Flutter to Flask server , 2022)

13.Flutter Application and Front End:

We are application to build our GUI (front end). The benefit of using Flutter is that we can easily customize our widgets. Flutter application takes less time to test as

they use single codebase. There are various advantages of using flutter. Flutter is open source, it is compatible, great delivery and performance with high productivity. Moreover, flutter is easy to learn and adapt. After checking the system configuration, the flutter setup should be completed. The very first step is designing how the application should look and creating a flow. Flutter has a stateless and stateful widget. There is communication between the screens with a stateful widget whereas a stateless widget never changes. Flutter allows us to build each screen and establish connectivity as per the flow created. Individual screens are created under the lib folder. The saved model “model_unquant.tflite” and “labels.txt” file is pasted under the assets folder. Always run or execute ngrok file or application before running our flutter application to scan the image. “Ngrok” is a cross-platform application that exposes local server ports to the internet. Run the application by typing the command “flutter run”. We are using rear camera to scan the images, which is only supported in mobile device. Hence, the application should be executed in mobile device.

```

30 final List<CameraDescription> cameras;
PROBLEMS 198 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER
NGER, id[1]=1, x[1]=482.0, y[1]=499.0, toolType[1]=TOOL_TYPE_FINGER, id[2]=2, x[2]=693.0, y[2]=535.0, toolT
E, metaState=0, flags=0x0, edgeFlags=0x0, pointerCount=3, historySize=0, eventTime=4889077340, downTime=488
D/ViewRootImpl[MainActivity](20325): dispatchPointerEvent handled=true, event=MotionEvent { action=ACTION_C
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS D:\take_image_input\take_image_input> flutter run
The plugin `tflite` uses a deprecated version of the Android embedding.
To avoid unexpected runtime failures, or future build failures, try to see if this plugin supports the Andr
future release of Flutter will remove these deprecated APIs.
If you are plugin author, take a look at the docs for migrating the plugin to the v2 embedding: https://flu
Launching lib/main.dart on CPH1969 in debug mode...
Running Gradle task 'assembleDebug'...
  
```

Figure 20: Running the Flutter application

When we open the application, there is a brief time or period during which the native app loads Flutter. By default, the native program shows a white splash screen during this period. This package automatically generates native code for the native splash screen’s background color and splash images for iOS, Android, and the web. We are using an image to be displayed on the splash screen.



Figure 21: Splash Screen

After a few seconds, the home screen is displayed. We have a camera icon to scan the images, and an icon for speech recognition, we have also included searching functionality.



Figure 22: Home Screen

On scanning an image, the system can identify the dish and alerts the user with calorie intake, ingredients used, and if any allergens are present in the system.

Please note that, not necessarily system will detect the image the very first time. Sometimes, we must scan the image 2-3 times in order to get the output as the data is fetched from Edamam API, there could be some connectivity issues established with the server at times. The system also displays the accuracy rate of the built model.

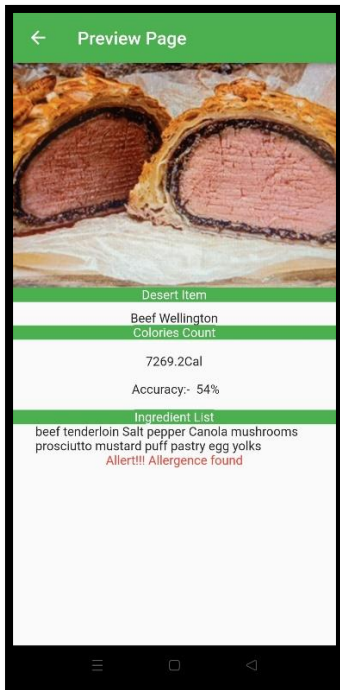


Figure 23: Preview Page: Allergens detected

Below is the snapshot where the application is identifying the dish, however, there are no allergens present. The name of the dish, calorie intake, and ingredients with accuracy rate are displayed.

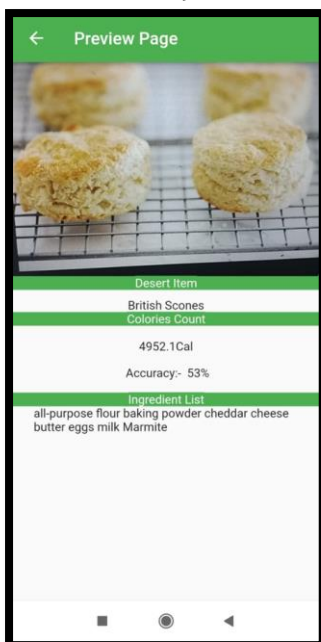


Figure 24: Preview Page: Allergens not detected

By using the speech recognition feature, when we input the dish name, the application is able to identify the dish and display the dish name as an output with calorie intake, ingredients and any allergens if present in the component

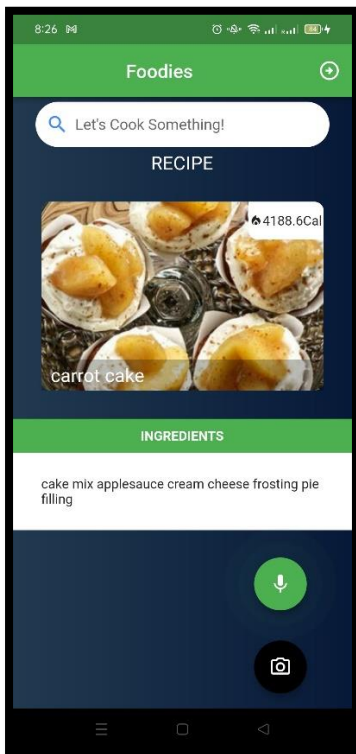


Figure 25: Preview Page - with speech recognition

14.Challenges and Limitations:

- **Availability of dataset:**
There is a limited number of datasets available. As we are working with an image dataset, the model requires a maximum number of images to predict accurately. More than 1000 images are required for every dish. However, hardly 600-800 images are available online. We cannot define the exact number of images that are required for every class to achieve an acceptable accuracy rate (70% - 90%). The likelihood that more data may contain meaningful information rises, which is helpful.
- **Recognition of multiple foods:**
Current system is not capable of identifying various cuisines, in future research will be conducted to list down the further requirements and expand the scope of the project capable of identifying all cuisine areas.
- **Quality of the dataset:**
Only good-quality images can be used for model training. Even if we have more than 1000 images for every dish, it won't be able to predict or give a

good accuracy rate unless the images are of high quality. The dataset we are using is a mixture of good and medium quality. Additionally, it is crucial that the dataset is prepared in such a way that our model can fully comprehend the data. The model will then be able to successfully use that dataset for learning.

- **Training the model:**
The number of recipes and images is directly proportional to the training time of the model. In general, it is understood that a poor approximation is caused by a lack of training data. Poor performance will be the result of both an under-constrained model that will likely overfit the training data and an over-constrained model that will likely underfit the limited training dataset. A model's performance estimation with insufficient test data will be too positive and subject to large variance.
- **Flutter:**
Flutter does not support python script. Hence there is a requirement to build API containing additional python code that predicts British items and then call API from the flutter application. API containing additional python code should be deployed or published on the server to avoid running localhost again and again.
- **Converting Keras model into TensorFlow lite:**
This is a challenging task as Flutter supports only TensorFlow lite model, required for Android iOS.
- **System requirement and GPU:**
GPUs (Graphic Processing Unit) are ideal for developing deep learning and artificial intelligence models since they are capable of handling numerous computations at once. Various parallel processes may be computed more effectively as they have several cores. Moreover, deep learning calculations must process enormous volumes of data; hence, for this reason, a GPU's memory bandwidth is ideal. GPU or an electric circuit is specialized and created to manipulate and change memory in order to speed up the production of images in a frame buffer that is intended to be output to a display device. GPUs are utilized in game consoles, personal computers, mobile devices, and so on.
- **Edamam API:**
Edamam API can be hit only 10 times a minute. Thus, the user has to wait once it has reached its limit after ten attempts. Extracting the recipe by scanning the image is very challenging. We have to create an account to access the Edamam API. We will be provided with an Id and a key for data extraction.
- **Accuracy rate:**
As we are currently facing many challenges with regard to dataset quality, availability, and GPU requirement, it is very challenging to achieve a good accuracy rate. An acceptable accuracy rate is anything between 70% - 90%.

The current system has issues identifying the images accurately due to the limitations with dataset availability and quality.

- **Scope of the project:**
Due to the above factors, we have restricted our scope to eleven dishes. In order to create a complete application, the scope of the project should be wider covering at least a complete cuisine. However, in our case, this is one of the biggest challenges that we are facing.
- **Google's Teachable machine:**
A web-based application called teachable machine allows building machine learning models quick, simple, and available to everyone. It can offer a good accuracy rate ranging from 75% - 92% with good batch size, epoch, and learning rate. As GPU and system configuration are not a constraint, we are free to use as many images as we can.

15. Evaluation, Results and Discussions:

15.1. Evaluation:

It is very helpful to find the model that best depicts our data and predicts how well the model will perform in the future. There are various approaches to evaluating a machine learning model's performance. We are using the accuracy score and confusion matrix.

Accuracy Score: The accuracy score is used to gauge the model's effectiveness by calculating the ratio of true positives to total true negatives across all made predictions. The accuracy score which goes from 0 to 1, represents the percentage of samples out of all samples that were properly identified.

The formula to calculate the accuracy score is,

$$\frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}}$$

```
from sklearn.metrics import accuracy_score, confusion_matrix, f1_score

print("Accuracy Score for Support Vector Machine algorithm:- ",accuracy_score(y_pred, y_test)*100)
print()
print("Accuracy Score for K-Nearest-Neighbor Lagorithm:- ",accuracy_score(pred_nbr, y_test)*100)
print()
print("Accuracy Score for Random Forest Lagorithm:- ",accuracy_score(pred_random, y_test)*100)
print()
print("Accuracy Score for Decision Tree Lagorithm:- ",accuracy_score(pred_dTree, y_test)*100)

Accuracy Score for Support Vector Machine algorithm:- 85.45454545454545

Accuracy Score for K-Nearest-Neighbor Lagorithm:- 76.36363636363637

Accuracy Score for Random Forest Lagorithm:- 83.63636363636363

Accuracy Score for Decision Tree Lagorithm:- 69.0909090909091
```

Figure 26: Accuracy score of SVM, KNN, Random Forest, and Decision Tree algorithm.

Confusion Matrix: A table called a confusion matrix is used to describe how well a classification system performs. The output of a classification algorithm is shown and summarized in a confusion matrix. In other words, the confusion matrix uses the key values. It provides information on the type of errors being produced, which is more essential than just the faults the classifier is making. This breakdown gets over the drawback of relying solely on categorization accuracy.

```

: print("##### Models comparison with Confusion Matrix #####")
print()
print("In Support Vector Machine algorithm, out of 55 model predict 47 correct predictions and 8 incorrect ")
print(confusion_matrix(y_pred, y_test))
print("#####")
print("In K-Nearest-Neighbor algorithm, out of 55 model predict 42 correct predictions and 13 incorrect ")
print(confusion_matrix(pred_nbr, y_test))
print("#####")
print("In Random Forest algorithm, out of 55 model predict 46 correct predictions and 9 incorrect ")
print(confusion_matrix(pred_random, y_test))
print("#####")
print("In Decision Tree algorithm, out of 55 model predict 38 correct predictions and 17 incorrect ")
print(confusion_matrix(pred_dTree, y_test))

#####

In Support Vector Machine algorithm, out of 55 model predict 47 correct predictions and 8 incorrect
[[26  6]
 [ 2 21]]
#####
In K-Nearest-Neighbor algorithm, out of 55 model predict 42 correct predictions and 13 incorrect
[[21  6]
 [ 7 21]]
#####
In Random Forest algorithm, out of 55 model predict 46 correct predictions and 9 incorrect
[[25  6]
 [ 3 21]]
#####
In Decision Tree algorithm, out of 55 model predict 38 correct predictions and 17 incorrect
[[19  8]
 [ 9 19]]

```

Figure 27: Model Comparison with confusion matrix

For model evaluation, only a few classes are considered, and images are downloaded online using the 'bing-image-downloader' library which is a python library for mass image download from Bing.com and because async URL is used, this package downloads incredibly quickly. However, we won't be using this package for dataset collection and rather will download the images from various sources with focus on getting the best quality dataset which can be used for model training. We are working with image dataset, and we have already defined the scope. SVM, KNN, Random Forest and Decision tree will give same result throughout. CNN has a scope of improvement in image classification and CNN performs well with large classes and datasets specifically dealing with image classification.

15.2. Results:

It is possible, that our model produces different predictions every time, even though it is trained on the same set of data repeatedly. This is to be anticipated and may possibly be a benefit of the algorithm rather than a flaw. Running the same algorithm on different data will also produce different results. This is called as the variance of the machine learning algorithm. The variance serves as a gauge of how responsive an algorithm is to the training data. With the current approach, we managed to get an accuracy rate of 81%.

Dataset	Classes	Number of Images per class	Epochs	Accuracy	Quality of the dataset	Source of the dataset and other comments
10 Dessert Items	10	150-200	20	Less than 10%	Good quality image	Images downloaded from ImageSearchDownload All apk manually. Limited images available
10 Dessert Items	10	700-800	20	11%-13%	60-80 images were with good quality, quality of other images was not good	Images downloaded from Google by adding an extension "Download All Images to the chrome. 700-800 images were available per class
10 Dessert Items	10	700-800	35	60%	75% images are of good quality	Good quality images collected from Unsplash website and also used Google's good quality images.
5 Dessert Items	5	700-800	20	60%	More images are required for good accuracy rate. However, this is not achievable with the existing system	Retraining the model gives varying accuracy rate. Reduced the scope to 5 classes for trial purpose and the accuracy rate was 60%.
5 Dessert Items & 6 other cuisines	11	700-800	10	78%	75% images are of good quality	Increased the scope and used 11 classes with an addition of 6 other cuisines and the accuracy was 78%
5 Dessert Items & 6 other cuisines	11	700-800	20	More than 81%	75% images are of good quality	Retrained the model again with a change in the epoch and achieved 83% on first attempt, the second attempt gave 81%

Table 1: Results with the accuracy rate

Note: The accuracy rate is fluctuating. Hence exact numbers are not defined in the table. The acceptable accuracy rate is 70% to 90%.

We have compared the accuracy score of Support Vector Machine, K Nearest Neighbour, Random Forest, Decision Tree, and CNN. Considering the model training time and system configuration, we have evaluated the performance of all algorithms using three classes. It is evident from the below comparison that Neural Networks are the best image classification algorithm.

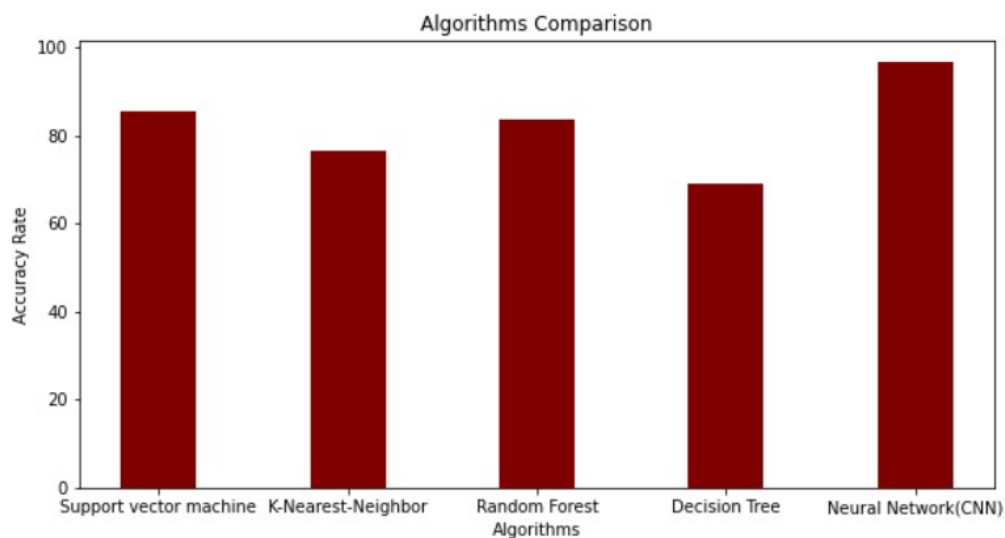


Figure 28: Model comparison

15.3. Discussion:

The results are changing every time the model is being trained. Result also depends upon the number of factors specifically the parameters, the number of images used to train the model, and the quality of the images used. Considering the existing limitations with regard to data quality and quantity along with the system specifications, it is very challenging to develop a complete system with the ability to identify all the dishes. Initially, we have restricted the number of images supplied to a model to 150-200 per dish. The accuracy score, in this case, was below the acceptance level i.e., less than 10%. The accepted accuracy rate is between 70% to 90%. Thus, there is a requirement to increase the dataset and the model was re-trained by supplying 700-800 images which resulted in 11% to 13% accuracy (accuracy rate changes as and when we train the model). Increased the epoch to 35 and the model has achieved 60% accuracy this time. However, the model was not predicting anything in evaluating the system. Thus re-trained the model and the accuracy rate suddenly came down to 11.67%. There are only two possibilities to get the model trained with a good accuracy score: one is to get a maximum number of datasets and the second is dataset should be of good quality. We couldn't get more than 800 images for the dessert items and the system configuration was not up to the mark as expected in computer vision. Computer vision requires millions of images in order to achieve a perfect or best accuracy

rate. Thus, the scope of the project was reduced, and implemented the same model in 5 dessert items. Again, the accuracy rate was not acceptable which came to around 60% this time. Just a random thought and used a combination of British dessert items and other cuisines. This time we took 5 dessert items and 5 other cuisines with an epoch set to 10 and the accuracy rate was 78% which was pretty good compared to previous work. Again, the epoch was increased to 20 and the model's accuracy rate came up to 81% which is good and acceptable.

16.Conclusion:

In this research work, we mainly focused on computer vision and speech recognition to identify cuisines along with detecting any allergens present in the recipe and calculating the calorie in the dish. We have discussed various algorithms that can be used to train the model in order to identify the dish. Support Vector Machine, K Nearest Neighbour, Random Forest, Decision tree, and CNN are the algorithms that we talked about. On comparing the model performance of all these algorithms, we have decided to use CNN in our application. Having discussed the major challenges and concerns of the dataset, we have built an application that can currently recognize 5 dessert items and 6 other cuisines with an 81% accuracy rate. Our future work includes expanding the scope and the system could then recognize a major area of dishes precisely and evaluating the system. System evaluation is another important factor post building a complete system. From the result, it is evident that there is a high possibility of an increase in the accuracy rate by increasing the dataset and its quality. Having a system with GPU can quickly train the model.

17.References:

AltexSoft. 2022. Image Recognition with Deep Neural Networks and its Use Cases. [online] Available at: <<https://www.altexsoft.com/blog/image-recognition-neural-networks-use-cases/>> [Accessed 29 August 2022].

Anthimopoulos, M., Gianola, L., Scarnato, L., Diem, P. and Mougiakakou, S., 2014. A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model. *IEEE Journal of Biomedical and Health Informatics*, 18(4), pp.1261-1271.

British Food: A History. 2022. Shortbread. [online] Available at: <<https://britishfoodhistory.com/2012/06/10/shortbread/>> [Accessed 19 August 2022].

Brownlee, J., 2022 [online] Available at: <<https://machinelearningmastery.com/keras-functional-api-deep-learning/>> [Accessed 26 August 2022].

Carrotmuseum.co.uk. 2022. The History of Carrot Cake. [online] Available at: <<http://www.carrotmuseum.co.uk/carrotcake.html>> [Accessed 19 August 2022].

Cooksinfo.com. 2022. Swiss Roll - CooksInfo. [online] Available at: <<https://www.cooksinfo.com/dishes/desserts/cakes/swiss-roll/>> [Accessed 19 August 2022].

Culinary Ginger. 2022. British Brandy Snaps. [online] Available at: <<https://culinaryginger.com/brandy-snaps/>> [Accessed 19 August 2022].

Day, I., Day, I. and profile, V., 2022. Battenburg Cake - the Truth. [online] Foodhistorjottings.blogspot.com. Available at: <<http://foodhistorjottings.blogspot.com/2011/08/battenburg-cake-truth.html>> [Accessed 19 August 2022].

Delia Online. 2022. Summer Pudding. [online] Available at: <<https://www.deliaonline.com/recipes/international/european/british/summer-pudding>> [Accessed 19 August 2022].

Encyclopedia Britannica. 2022. Chelsea bun | Definition, Ingredients, & History. [online] Available at: <<https://www.britannica.com/topic/Chelsea-bun>> [Accessed 19 August 2022].

Fang, A., 2022. How to Use the Edamam Food and Grocery Database API with Python, PHP, Ruby & Javascript Examples - RapidAPI. [online] The Last Call - RapidAPI Blog. Available at: <<https://rapidapi.com/blog/edamam-food-and-grocery-database-api-with-python-php-ruby-javascript-examples/>> [Accessed 18 August 2022].

GitHub. 2022. deep-learning-keras-tf-tutorial/cnn_cifar10_dataset.ipynb at master · codebasics/deep-learning-keras-tf-tutorial. [online] Available at: <https://github.com/codebasics/deep-learning-keras-tf-tutorial/blob/master/16_cnn_cifar10_small_image_classification/cnn_cifar10_data_set.ipynb> [Accessed 1 September 2022].

Johan, E. and Rizal, A., 2021. Allergen Recognition in Food Ingredients with Computer Vision. *Ultima Computing : Jurnal Sistem Komputer*, 13(2), pp.44-49.

Johnson, D., 2022. CNN Image Classification in TensorFlow with Steps & Examples. [online] Guru99. Available at: <<https://www.guru99.com/convnet-tensorflow-image-classification.html>> [Accessed 30 August 2022].

Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V. and Ma, Y., 2022. DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment. [online] arXiv.org. Available at: <<https://arxiv.org/abs/1606.05675>> [Accessed 22 August 2022].

Lutkevich, B., 2022. What is Speech Recognition?. [online] SearchCustomerExperience. Available at:

<<https://www.techtarget.com/searchcustomerexperience/definition/speech-recognition>> [Accessed 30 August 2022].

Mimi, A., 2022. GitHub - afsanamimii/Multiclass_image_classification. [online] GitHub. Available at: <https://github.com/afsanamimii/Multiclass_image_classification> [Accessed 1 September 2022].

Ozimek, S., 2022. Authentic British Scones. [online] Curious Cuisiniere. Available at: <<https://www.curiouscuisiniere.com/british-scones/>> [Accessed 19 August 2022].

Mishra, M., Sarkar, T., Choudhury, T., Bansal, N., Smaoui, S., Rebezov, M., Shariati, M. and Lorenzo, J., 2022. Allergen30: Detecting Food Items with Possible Allergens Using Deep Learning-Based Computer Vision. Food Analytical Methods,.

Renotte, N., 2022. GitHub - nicknochnack/ImageClassification: Jupyter notebook showing how to build an image classifier with Python and Tensorflow. [online] GitHub. Available at: <<https://github.com/nicknochnack/ImageClassification>> [Accessed 1 September 2022].

Ross, G., Bremer, M. and Nielen, M., 2018. Consumer-friendly food allergen detection: moving towards smartphone-based immunoassays. Analytical and Bioanalytical Chemistry, 410(22), pp.5353-5371.

Salim, N., Zeebaree, S., Sadeeq, M., Radie, A., Shukur, H. and Rashid, Z., 2021. Study for Food Recognition System Using Deep Learning. Journal of Physics: Conference Series, 1963(1), p.012014.

Selvan, A., 2022. Blogs/app-v1.py at main · AgnelSelvan/Blogs. [online] GitHub. Available at: <https://github.com/AgnelSelvan/Blogs/blob/main/upload_image_to_server/flask-server/app-v1.py> [Accessed 2 September 2022]. (Selvan, 2022)

Shen, Z., Shehzad, A., Chen, S., Sun, H. and Liu, J., 2020. Machine Learning Based Approach on Food Recognition and Nutrition Estimation. Procedia Computer Science, 174, pp.448-453.

What A Girl Eats. 2022. Strawberry Parfait (Strawberry Fool). [online] Available at: <<https://whatagirleats.com/strawberry-fool/#:~:text=A%20Fool%20is%20a%20classic,blended%20with%20the%20whipped%20cream.>> [Accessed 19 August 2022].

2022. [online] Available at: <<https://healthitanalytics.com/news/deep-learning-detects-allergic-reactions-in-patient-safety-reports>> [Accessed 24 August 2022]. Srivastava, A., 2022. Developing Cross-Platform (Android/iOS) Apps Using a Single Codebase With Flutter. [online] Medium. Available at: <<https://medium.com/walmartglobaltech/developing-cross-platform-android-ios->

apps-using-a-single-codebase-with-flutter-f380ff15303a#:~:text=Flutter%20is%20Google's%20mobile%20UI,how%20exactly%20does%20it%20work%3F&text=Development%20in%20Flutter%20is%20done,to%20any%20other%20OOP%20language).> [Accessed 25 August 2022].

18. Bibliography:

Youtube.com. 2022. [online] Available at: <https://www.youtube.com/watch?v=dw96S_iFFbl> [Accessed 7 September 2022].

Youtube.com. 2022. [online] Available at: <https://www.youtube.com/watch?v=Su_qbc98xsE> [Accessed 7 September 2022].

19. Appendix C:

Ethics Approval: Approved ethics form is attached below.



Project status

Status

● ● ● Approved

Actions

Date	Who	Action	Comments
20:27:00 07 July 2022	Drishty Sobnath	Supervisor approved	
19:43:00 07 July 2022	Surya Gopu Pillai	Principal investigator submitted	

Figure 29: Ethics Application approval

**SOLENT UNIVERSITY SOUTHAMPTON
FACULTY OF BUSINESS, LAW AND DIGITAL
TECHNOLOGIES**

MSc Applied AI and Data Science

2022

Surya Gopu Pillai - Q15784037

***“Computer Vision and Speech
Recognition - Auto Recipe detection
and alert system using AI”***