Improving the Food classification rate using Transfer Learning methods

MSc Research Project
Data Analytics

Dhanashree Subhash Rane
Student ID: x20142498

School of Computing
National College of Ireland

Supervisor: Dr.Christian Horn

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Dhanashree Subhash Rane
Student ID:	x20142498
Programme:	Data Analytics
Year:	2021
Module:	Msc Research Project
Supervisor:	Dr.Christian Horn
Submission Due Date:	16/12/2021
Project Title:	Improving the Food classification rate using Transfer Learning methods
Word Count:	6308
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Dhanashree Subhash Rane
Date:	31st January 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Improving the Food classification rate using Transfer Learning methods

Dhanashree Subhash Rane
x20142498

Abstract

Nowadays due to growing food consumption and variance in dietary behaviour, the diet may get imbalanced. The mainstream part of the diet is the calorie intake, which shall be taken care of. To appropriately assess and monitor the diet, various advanced learning frameworks have been incorporated. Therefore, in our study, we have introduced a food classification model which will recognize the food product and state the calories of each. For this, we have utilized the DNN algorithms such as Mobile Net V2, Inception V3, and Custom Architecture. Each of these algorithms is evaluated and the optimal algorithms based on performance are considered for further implementation. The application for this model can be widened with the growing technological advances.

1 Introduction

The growing population of humans has exponentially increased food intake. Food is the crucial part of the life cycle on which the whole living ecosystem relies upon. The increasing food demand has also stressed the food web chain. Additionally, people are nowadays inclined towards unnatural food intake from the natural food intake. These unnatural foods may be processed, adulterated, or related which is harmful to the human diet at a longer-term view. Although, such type of foods is the best way to meet the day-to-day demands of the consumers. But this raises various concerns for the food intake such as quality, safety, requisite nutrients intake, etc. The consumption of food must provide the requisite intake of nutrients and cope with the optimal diet level of the consumer. Intake both below and above the required level may cause harm to the consumer. Therefore, neither one should have an under-provisioned diet nor over provisioned diet but to only have a balanced diet. Under various detailed assessments, it is seen that various peoples around the world are prone to health risks due to the imbalance in their daily diet. To ensure a balanced diet, there are various essential factors to be considered such as calories, carbohydrates, proteins, fats, fibres, vitamins, minerals, and water. The foremost factor of all these is calories which is the energy from the consumed food. This energy is sourced for food and utilized for physical human activities. Although, calories shall be availed until a certain threshold upon which it may have an adverse effect on the health of a person. Therefore, an effective and autonomous approach for dietary assessment and monitoring systems must be suggested.

In recent times, technology has been advancing rapidly in the field of computer vision and artificial intelligence. The domain of computer vision has allowed the processing and identification of objects in a real-time environment. This system can be implemented for various types of applications such as facial recognition, object detection, food classification, etc. For food classification, these computer vision algorithms can detect various kinds of complex features in the food which may not be possible through human visions. Computer vision can interpret features such as size, weight, dimension, texture, and food colour. Through these systems, the nutritional and diet intake can be assessed and monitored. Apart from this, the application of this system can also be utilized in inventory management, restaurant billing management, food quality check-up, etc. This framework relies on two processes which are detection and recognition. Initially, the food product is detected and then the dish is recognized. In this various methods and deep learning algorithms such as neural networks are utilized for the pre-processing, feature extraction, training of the model during the implementations.

In our study, the model is implemented for the food classification which will then interpret the calories of each food recognized. For the model, we have sourced the dataset from the Kaggle repository which is the Food-11 image dataset. This dataset contains 11 food categories sample images with a total image of 16643. Furthermore, the dataset images are split into three folders which are evaluation, training, and validation. These datasets are then pre-processed using appropriate methods are to overcome any kinds of imbalance or noise in the images. In the given dataset, each food category sample images have variance in the number of samples. As our model utilizes the weighted approach, the image samples must be balanced using the generative adversarial network (GAN) model. This model generates similar images samples using the existing image samples and thus balances the whole dataset. Furthermore, the dominant features from the sample images are extracted utilizing various feature extraction techniques. For the implementation of the model, our study utilizes the weights of Image Net through pre-trained algorithms such as Mobile Net V2, Inception V3 and furthermore the custom architecture. Here, the transfer learning approach is utilized because it can increase the efficiency of the performance and save computational time by just fine-tuning the various parameters of the model. After the implementation of the algorithms, the models are assessed through different evaluation metrics such as PRF score, computational time, etc. Once the model classifies the food product, the intake calories for the classified food are shown to the consumer.

1.1 Research Question

- Does data augmentation techniques enhances the accuracy of the model for food classification ?
- Which model efficiently recognise the category of food products? Does reducing the number of classes enhances the model performance ?

2 Literature Review

2.1 Food Classification employing Transfer Learning Approach

Hafiz et al. (2020) in the study utilized the deep transfer learning approach for the classification of soft drink type in the model for diet assessment system. With the rapid increase of soft drink consumption around the world, it is necessary to track the consumption limit for a better health approach. Therefore, through this study, the author aimed to provide an effective approach to detect the nutritional output of the soft drink and make aware to the consumer. This model utilizes the convolutional neural network (CNN) algorithm combined with the transfer learning approach. For Dataset, it was acquired for the Image Net repositories and custom obtained. To increase the efficiency of the process, the author implements noise reduction and contrast enhancement during the pre-processing phase. With the completion of this process, the deep transfer learning approach is implemented. The nutritional output is interpreted through the value table provided whereas the capacity of the bottles or cans is interpreted through the implementation of the Bag of Features (BoF) model. In the final take, during the evaluation of the model, it achieved an accuracy score of 98.57% which outperformed some of the conventional approaches. Similarly, Islam, Wijewickrema, Pervez and O'Leary (2018) surveyed the approach of deep transfer learning for food classification. The paper in the discussion stated various types of food classification approaches such as nutrition management, diet intake calculators, and food menu demands calculators in restaurants. For the study, both Food-5K and Food-11 database is utilized in the approached model. The learning set and validation set were split into the ratio of 70% and 30%, respectively. Then various dominant characteristics from the sourced dataset were obtained by implementing the deep convolutional neural network framework during the phase of pre-processing of the dataset. Furthermore, various pre-trained algorithms such as Alex Net, Google Net, Res Net 50, etc were implemented to train the model. Nonetheless, these models also experimented with the feature extraction phase previously. In addition to this, different conventional algorithms such as support vector machines and convolutional neural networks were also utilized to interpret the detailed comparison of the pre-trained and convolutional algorithms. Also, the author introduced a customized dataset named Food-22 to evaluate the efficacy of the model. Although the proposed model evaluated an efficiency of 88.09% which was just moderate in comparison to other research. On the other hand, Burkapalli and Patil (2020) also incorporated the transfer learning mechanism for food classification. In the study, the customs classification algorithm is built based on the Inception v3 framework. This approach will effectively optimize the model and increase the performance efficiently. The sample dataset utilized in this model contained 16 different types of the food product. These samples were split in the ratio of 80:20 for the training. The model was implemented under the 100 epochs. The model here utilized the RMSprop optimizer which effectively tuned the output. In the final take, the model achieved an accuracy of 98.77%. Additionally, the author suggested the future scope of work using the framework of TFOD, YOLO and Dectoron2.

Rajayogi et al. (2019) studied the mechanism of transfer learning for the classification of Indian foodstuffs. To increase the computational rate and the efficiency of the model, the conventional algorithm could be replaced by the transfer learning algorithms which work on the previously learned framework. In the model, the Indian food dataset was utilized which contained 20 different food dishes which had 500 image repositories. Although the dataset had multiple distortions which could be removed during the pre-processing phase. The imbalance in the dataset was also phased out which enhanced the overall efficiency of the dataset. For the learning set and validation set, the samples were split into the ratio 80:20 due to which it had 8000 learning samples and 2000 validation samples. Here the CNN model was classified using the pre-trained model as it helped overcome the challenge of overfitting. The batch size considered in this model was 32 with an epoch of 30. Further, the stochastic gradient descent optimizer was utilized to reduce the loss function by enhancing the various features of the model. SoftMax function and the global pooling layer were utilized in the process to enhance the model. In the final stage, the model could extract the calorific value of the classified food. With the overall output, in the study, Inception V3 achieved a higher accuracy of 0.879 with a loss function of 0.5893 whereas Res Net achieved a lower accuracy of 0.6991 with a loss function of 1.0804. Although the study needed further enhancement for the real-time application. Thiodorus et al. (2021) proposed a novel approach to classify various food products. In the study, the tray box image mechanism was formed utilizing the convolutional neural network with the transfer learning method. Using this novel, the weight of the trained model is combined, and the loss function and computational duration are reduced. For the research, the model utilized in this study is Google Net and Res Net 18. During the computational process, each food sample is set in the tray boxes. The author in the study experiments with the different epoch and suggests the best suiting for the efficient model. Furthermore, the author set 242, 29, and 33 as the learning, validation, and test set, respectively from the 304 food image samples. During the evaluation of the final approach, the Res Net 18 model achieved higher performance than the Google Net model. The accuracy, precision, recall and f-1score for the Res Net 18 model stood at 1 whereas for the Google Net model was a bit low. This study showed that the combination of tray box image mechanism with the deep transfer learning algorithm could enhance the model in the real-time application.

2.2 Food Classification employing Local Binary Pattern Approach

Nguyen et al. (2014) proposed a mechanism for food image classification using local appearance and global structural information. This model will help to obtain the exterior features of the food product to recognize the object. In this model, the author utilized two types of descriptors to classify the food product. The two models are local binary pattern (LBP) and non-redundant local binary pattern (NRLBP) which would extract the various characteristics of the food product. Furthermore, the histogram dimension for the LBP and NRLBP are 59 and 30, respectively. Additionally, the author in the study utilized the Pittsburgh Fast Food Image (PFI) dataset for the model. This dataset had 6 food category samples in it. This proposed model also compared the model with other conventional models such as SVM, SIFT and histogram. Although, this model performed poorly achieving an average accuracy of 51%. Therefore, the author also provided different views and challenges in the paper to overcome in the future scope of the study. Similarly, Inunganbi et al. (2018) proposed the approach of food classification using a novel mechanism. The author here utilized interactive image segmentation. This proposed model utilizes the random forest algorithm with the combination of interactive image segmentation. To balance the noise and distortions in the image samples, various pre-processing approaches are implemented such as Gappy Principal Component Analysis and Boundary detection and filling algorithms were utilized. During the phase of feature extraction, the framework of LBP and NRLBP is utilized. These extracted components are compiled and sent for further implementation. Here, a support vector machine algorithm is also utilized for classification purposes. In the study, the author here utilized the dataset of Food 101. These models were also compared with different types of conventional algorithms. With the evaluation of the different models, the proposed framework showed an effective performance and outperformed all the other conventional mechanisms.

Madgi et al. (2015) proposed a different approach to classify and identify the vegetable images with the combination of RGB pattern and the local binary pattern framework. Through these approaches, the model initially extracts the fundamental statistic and texture characteristics. Then a feature vectorization is developed which consists of features with the colour and texture of the vegetable images. The vegetable image sample consists of both leafy and non-leafy vegetables. The sample dataset has 18 vegetable categories which are equally divided among leafy and non-leafy vegetables. Furthermore, the author here utilized an advanced learning algorithm called a multiple-layer neural network for the classification of the vegetables. In the final take, the model was evaluated using various types of evaluation metrics. After evaluation, the model achieved an accuracy of 93.3% among the classification of different vegetables. This proposed model was widely being adopted by various domains in real-time applications.

2.3 Food Classification employing Histogram of Oriented Gradient Approach

Rahmadani (2020) proposed a model for food classification employing the method of the histogram of oriented gradient (HOG) and k nearest neighbour (KNN) for feature extraction and classifier, respectively. The study aimed to integrate this model with a payment system to identify the food dishes and charge the amount for the customers autonomously. This proposed model aligned with the motto of the concept called Industri 4.0. Through this, the restaurant can limit the manpower with enhanced efficacy and a lower process period. The dataset utilized in this study had six different food image samples. Each combination of food samples had a certain mentioned price which would be shown when detected. Then the dominant features of each image sample are extracted using the method of HOG. This method consists of four different sequential processes which are image conversation, gradient computes, spatial orientation binning and normalization block. Once done, these extracted features are fed into the implementation of the KNN classifier to classify the food dishes in the model. It was seen that with the highest k value the model could effectively classify the outcome. These outcomes were then calculated using the appropriate evaluation metrics such as precision, recall and f-1 score. In the final experiment, the model achieved a higher accuracy of 82%. Although the model can be enhanced through alternate tuning for better efficiency. Islam et al. (2019) in a study investigated a model to automatically detect plants using the method of HOG and LBP with the combination of SVM classifier. Different types of plants have different kinds of applications and for that purpose, it is necessary to interpret each variety. Although recognition of the different varieties is a crucial task and therefore an effective autonomous is necessary for the domain. The author proposed an effective model to detect various kinds of plant species. The study here utilized the Flavia dataset for the model implementation. Primarily, the features of the image samples were extracted from the dataset through the methods such as histogram of oriented gradient and local binary pattern. Once the features are extracted, the SVM classifier is implemented to the model classify the different varieties of the plant. Various types of image dimensions were considered such as 2x2, 4x4 and 8x8 for the study. During the evaluation of the different combinations of the models, there was a notable variance in the accuracy rate.

The proposed combined model achieved an accuracy of 91.25%. This model could be utilized in a real-time environment.

2.4 Food Classification employing Scale Invariant Feature Transform Approach

Giovany et al. (2017) utilized the SIFT mechanism in association with the machine learning algorithms to recognize the Indonesian food image. The author here utilized the image samples sourced from the various online repositories. Before sourcing the images, various types of certain characteristics were considered. Then the images are pre-processed if there is any kind of imbalance in the samples. For the feature extraction phase, the author utilizes the method of SIFT. This method follows four steps to extract features which are constructing a scale space, Laplacian of gaussian application, finding key point candidates, determining the key points, orientation assignment, and image descriptor. Once the features are extracted from the image samples, the machine learning algorithms are implemented. The machine learning algorithm utilized in this study is the k-decision tree and backpropagation neural network (BPNN). The training and test set was divided into the ratio of 90% and 10%. In the final phase, these models were evaluated and compared. Although the model achieved a lower accuracy as the BPNN achieved around 51% whereas the k-DT achieved an accuracy of 44%. This proposed model could not be utilized in the real-time environment and should be tuned for better efficiency. On the other hand, Fakhrou et al. (2021) proposed a system of the smartphone-based food recognition program. In this system, the author utilized the SIFT descriptor for feature extraction and the CNN algorithm for food recognition. This model aimed to recognize the food dishes through a mobile application that would help persons with visual impairments. The recognized food dishes will speak by the application to the user. In the model, the FOOD-101 dataset was utilized. This dataset consists of 29 food dishes categories. Here, various frameworks are implemented such as deep convolutional neural network, transfer learning, and ensemble learning. Also, custom deep CNN models were implemented which was a combination of both transfer learning and ensemble learning. To assess the models, various evaluation metrics were implemented to evaluate the efficacy of the model. The study found that the ensemble learning framework performed effectively than the other considered algorithms. Furthermore, with the utilization of a customized dataset, the model achieved an accuracy of 95.55%. This proposed model could also be utilized in various other open datasets for better adoption.

2.5 Food Classification employing various other Deep Learning Approach

Al-Sarayreh et al. (2020) researched the species in meat categorization using deep learning in the study. Convolutional Neural Network was employed for autonomous fruit categorization utilizing deep learning methods and computer vision in another study by Pande Pande et al. (2019). Similarly, Nasiri employed the Deep convolutional neural network approach to select excellent grade date fruit by integrating categorization phases and extraction of features in a distinctive framework Nasiri et al. (2019). Momeny et al. (2020) suggested a model that used an upgraded Convolutional Neural Network (CNN) approach to construct cherry fruit packaging methods to decrease wastage and boost principal elements and commercial viability. Hu's work used RCNN to develop and evaluate a low-cost, fast approach for categorizing calamari in produced items Hu et al. (2019). Ponce Real et al. (2019) recognized olive fruits using computer vision techniques and a convolutional neural network. In another study, Fahira et al. (2019) proposed a model using the classical machine learning and deep learning algorithm to classify Sumatra traditional food images. In related research Islam, Karim Siddique, Rahman and Jabid (2018), the authors used the Convolutional neural network approach for image processing of food goods, and they identified the packaged foods using a food sample. The comparative analysis for the various studies by different researchers is shown in Table 2.5

Paper Title	Publish Year	Method	Advantages	Future Scope / Disadvantages
Image-based soft drink type classification and dietary assessment system using deep convolutional neural network with transfer learning	Nasiri et al. (2019)	This model utilizes the convolutional neural network (CNN) algorithm combined with the transfer learning approach	In the final take, during the evaluation of the model, it achieved an accuracy score of 98.57% which outperformed some of the conventional approaches	Could be further enhanced
TRANSFER LEARNING: INCEPTION-V3 BASED CUSTOM CLASSIFICATION APPROACH FOR FOOD IMAGES	Burkapalli and Patil (2020)	The model here utilized the RMSprop optimizer which effectively tuned the output	In the final take, the model achieved an accuracy of 98.77%	Additionally, the author suggested the future scope of work using the framework of TFOD, YOLO and Dectoron2
Food Detection Using Histogram of Oriented Gradient (HOG) as Feature Extraction and K-Nearest Neighbors (K-NN) as Classifier	Rahmadani (2020)	This proposed model aligned with the motto of the concept called Industri 4.0	In the final experiment, the model achieved a higher accuracy of 82%	Although the model can be enhanced through alternate tuning for better efficiency
Convolutional Neural Network with Transfer Learning for Classification of Food Types in Tray Box Images	Thiodorus et al. (2021)	For the research, the model utilized in this study is Google Net and Res Net 18	The accuracy, precision, recall and f-1score for the Res Net 18 model stood at 1 whereas for the Google Net model was a bit low	This study showed that the combination of tray box image mechanism with the deep transfer learning algorithm could enhance the model in the real-time application

Table 1 : Comparison of Different Works for this Study

3 Methodology

The main objective of this research is to classify the food products based on the Food image, with their calories intake. To achieve this objective, in this research we have proposed a framework for food classification, which utilizes the multiple deep learning architectures. The complete framework can be splitted into series of steps for achieving the better results. The steps of proposed framework involves Food image collection, image pre-processing, model training and evaluation of models. The proposed framework architecture is shown in Figure 1.

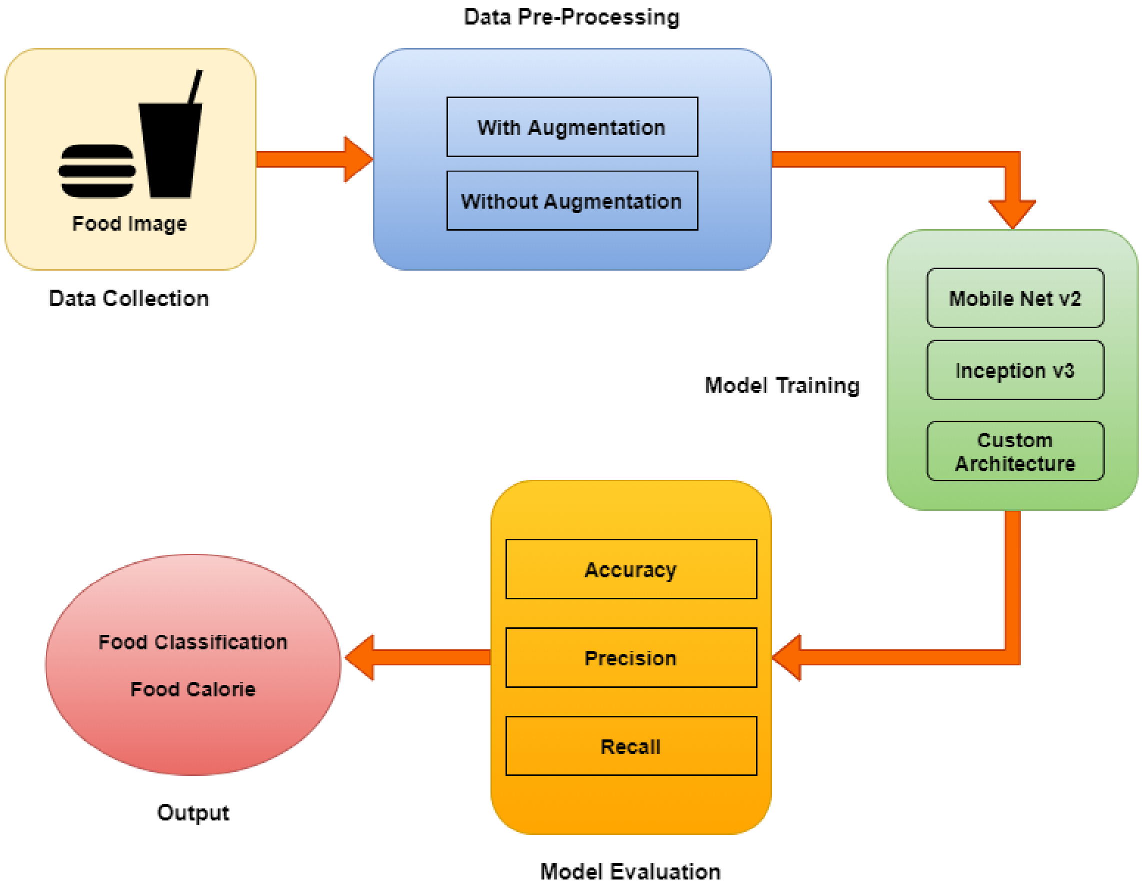

Figure 1: Method Flow Chart

3.1 Data Acquisition

To implement the model, the dataset for the model was acquired. We have obtained the Food-11 dataset from the Kaggle repositories *Food-11 image dataset* (n.d.) . This dataset consists of the 11 categories of food products with the total number of image samples standing at 16643. It had various kinds of food products such as bread, dairy product, dessert, egg, fried food, meat, etc. Although initially, we have utilized all the categories for the model, we have also evaluated considered only certain categories in further evaluation for assessing the efficacy of the model. In addition to this, the dataset was splitted into evaluation, training, and validation afore.

3.2 Data Pre-Processing

Once the dataset is acquired, the next phase comes is the data pre-processing. In our model, we have considered implementing the model both with data augmentation and without data augmentation. Through this approach, we can effectively evaluate the optimal output from the dataset. Each algorithm implemented in our model will have both data augmentation and without data augmentation. In data augmentation, various parameters have been transformed. Here, during the augmentation, the sheer range and zoom range are set at 0.2. In addition to this, the rotation range stood at 40 with the resize scale at 1:255. Although, we will not perform any data augmentation with validation data as we want to make the model perform well in the real-world environment. Considering the image samples without the data augmentation, the parameters have remained unaltered. The originally obtained images are utilized here for the model. The images with the augmentation will be well-refined whereas the images without the augmentation will be less effective.

3.3 Model Training

In this phase, the DNN algorithms are implemented on the acquired dataset. For our model to effectively recognize the food product, we have considered three algorithms. In addition, to recognize the food product, the calories of the food product are also evaluated. Then the models with the implemented algorithms are evaluated and an optimal is selected for further implementation. The three algorithms implemented in our model are Mobile Net v2, Inception v3, and Custom architecture. Here, the Mobile Net framework is a pre-trained convolutional neural network that generally has low latency for implementation in the embedding vision. Whereas the Inception framework is also a convolutional neural network that reduces the dimensionality of the samples for an effective computation. This framework could easily counter the challenges such as overfitting, etc. On the other hand, in the custom architecture, the convolutional layers are constructed here. Both the pre-trained algorithms which are Mobile Net and Inception are weighted using the ImageNet database. Furthermore, the batch size for training the model is set at 4 during the training of the model.

3.4 Evaluation of the Model

After the implementation of the algorithms, the models are evaluated using the appropriate metrics. Here, we have considered the two crucial metrics for the evaluation which are Accuracy and Loss. The number of the correct prediction of the total number of predictions is known as the accuracy whereas the loss is defined as the number of incorrect predictions over the total number of predictions. Generally, the accuracy is anti-proportionate of the loss value. The lower is the loss value of the model, the higher is the accuracy of the model. The algorithm with the highest accuracy and the lowest loss value will be considered as an optimal algorithm. This algorithm will be then implemented for further implementations. Other than the accuracy and loss, precision, recall and F1-Score for models also has been calculated.

4 Design Specification

In this section, we will discuss the detailed architecture of each algorithm implemented. The algorithms implemented in our models are Mobile Net v2, Inception v3, and Custom Architecture.

4.1 Mobile Net V2

The first algorithm that is Mobile Net V2 is a pre-trained convolutional neural network framework with a low latency output. It is generally focused on the application of the light-weighted computation which is for mobile and embedding vision. It also depends on the TensorFlow framework for the model implementation. Here in the layer framework, the standard convolutional layer is replaced by the depth-wise separable convolution layer. The computation standards for the model are less complex and the computation expenses are limited. The figure depicts the graphical representation of the architecture of the Mobile Net V2.

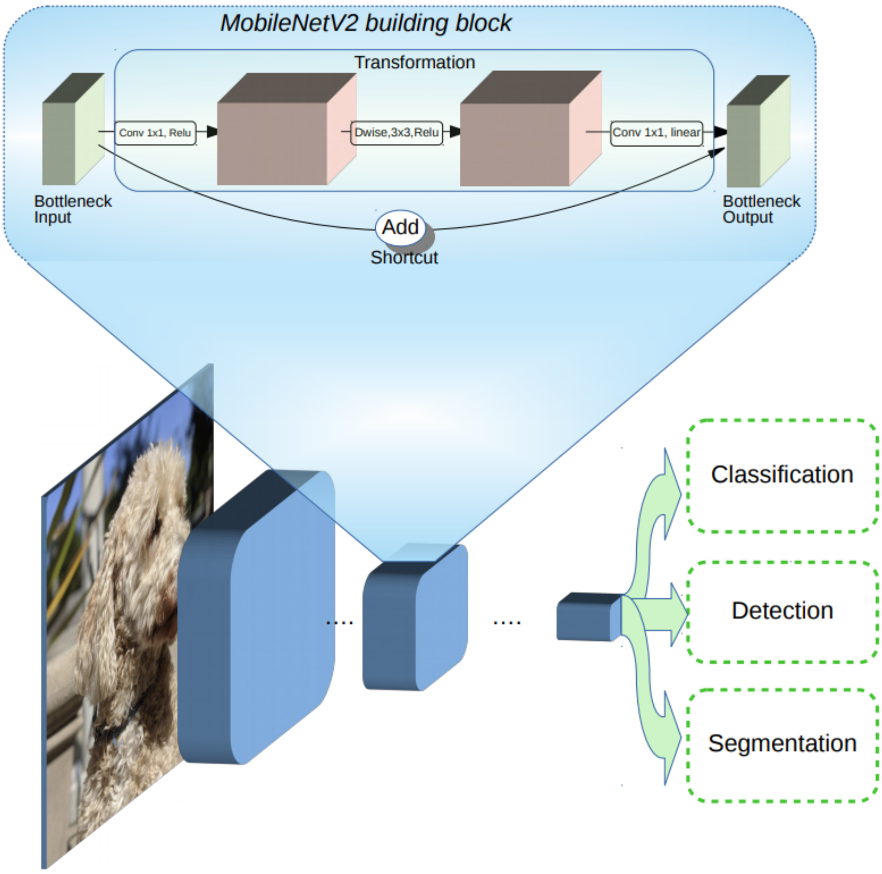

Figure 2: MobileNet V2 Architecture *MobileNetV2: The Next Generation of On-Device Computer Vision Networks* (n.d.)

4.2 Inception V3

The second algorithm here is the Inception V3 which relies on the framework of convolutional neural network. It is derived from the framework of the Google Net and the weights are trained from the Image Net repositories. This architecture consists of a 48-layer deep pre-trained convolutional neural network. Furthermore, this framework reduces the dimensionality of the input with stacked 1x1 convolutional. Inception V3 has achieved an accuracy of more than 78.1% when fed with the Image Net samples. In the figure, the graphical representation of the Inception architecture can be observed. Convolutions, average pooling, max pooling, concats, dropouts, and fully linked layers are among the symmetric and asymmetric architectural elements used in the model. Batch norm is utilized significantly in the model and is implemented to activation parameters.

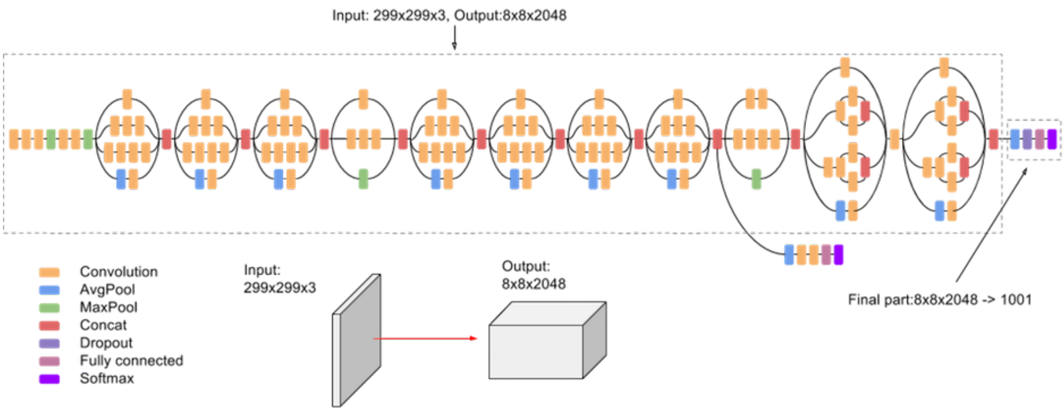

Figure 3: Inception V3 *Advanced Guide to Inception v3 | Cloud TPU* (n.d.)

4.3 Custom Architecture

In the third algorithm, the custom convolutional layers are constructed. Here, during the construction of the layers, the convolutional layer of 2D and Max Pooling 2D is placed. In this, the total number and the dimension of the pixel filters are addressed. The feature map is also extracted for the model. On the other hand, the max-pooling layer reduces the dimensionality to the described dimension pixels. Multiple activation functions are utilized such as ReLu and Softmax. Also, the optimizer such as Adam is utilized in the convolutional layer. Here, the dropout value is set, to counter the challenge of overfitting.

5 Implementations

Our model primarily aims to develop a model to recognize the category of food and specify the calories. Here, for the effective output, deep neural network algorithms are utilized which are Mobile Net v2, Inception v3, and Custom Architecture. Each of these algorithms consists of convolution layers. The Mobile Net framework with a low latency output which is generally utilized for mobile and embedding vision whereas the Inception framework is built under the framework of Google Net. On the other hand, in the custom architecture, the convolutional layers are set manually. In the model, initially, the food images are captured which interprets the category of the food. The image samples are pre-processed and trained to utilize the algorithms where these pre-trained algorithms handle the extraction of the features. Once the model is trained, using the image samples the output is obtained. After the food product is recognized, the calories of the food are also stated. The figure depicts the detection of the food. The model is implemented in the python programming language with a vast set of libraries. Various modules utilized during the implementations are NumPy, Pandas, Matplotlib, os, cv2, etc. Here, the cv2 module shows the computer vision library. During the implementations, the samples obtained are considered in two approaches which are with data augmentation and without augmentation. Furthermore, the model is evaluated in two ways wherein the dataset with 11 categories and 3 categories are utilized. In addition to these, the optimal algorithms are evaluated using the evaluation metrics such as Accuracy, Loss, Precision, Recall, and F-1 score.

6 Evaluation

The algorithms utilized in our model are Mobile Net v2, Inception v3, and Custom Architecture. Each algorithm implemented is evaluated using the metrics such as Accuracy, Loss, Precision, Recall, and F-1 score. The model is evaluated in two approaches wherein the 11 categories and 3 categories are considered.

6.1 Experiment 1 / Evaluation Based on Accuracy

In this section, the accuracy of each algorithm is considered. The algorithm with the higher accuracy is considered the optimal algorithm. Accuracy is defined as the number of correct predictions over the number of total predictions. Here, the accuracy for both 11 categories and 3 categories is discussed.

6.1.1 11 Food Categories

We have evaluated their pre-trained DNN algorithms here which are Mobile Net v2, Inception v3, and Custom Architecture. Initially, for each algorithm the dataset with augmentation and without augmentation. In the dataset with augmentation, various parameters are considered and transformed. Here, parameters such as shear range, zoom range, rotation range, etc were set. Whereas in the samples without the augmentation, these parameters are remained untouched. During the implementation, the batch size was set at 4 with a total of 15 epochs. In each algorithm, in the convolution layer, the ReLu activation function and Adam optimizer function were utilized. In the dense layer, the SoftMax function was utilized. Although, only in Mobile Net v2 and Inception v3, the weights of Image Net repositories were implemented. Generally, the accuracy of the training and test set is obtained, but in our case, we will only consider testing accuracy for the optimal algorithm. Considering the algorithms without the data augmentation, the Mobile Net v2 achieved a training accuracy of 27.41% and the testing accuracy achieved stood at 35.01% whereas for the Inception v3, the training accuracy achieved was 23.64% and the testing accuracy stood at 29.18%. On the other hand, for the custom architecture, the training accuracy stood at 28.69% and the testing accuracy stood at 30%. Here, it was observed that the custom architecture achieved higher training accuracy whereas the Mobile Net v2 achieved higher testing accuracy. Similarly, the lowest training and testing accuracy was scored by Inception v3. The graphical representation of accuracy for each algorithm is depicted in the Figure 4.

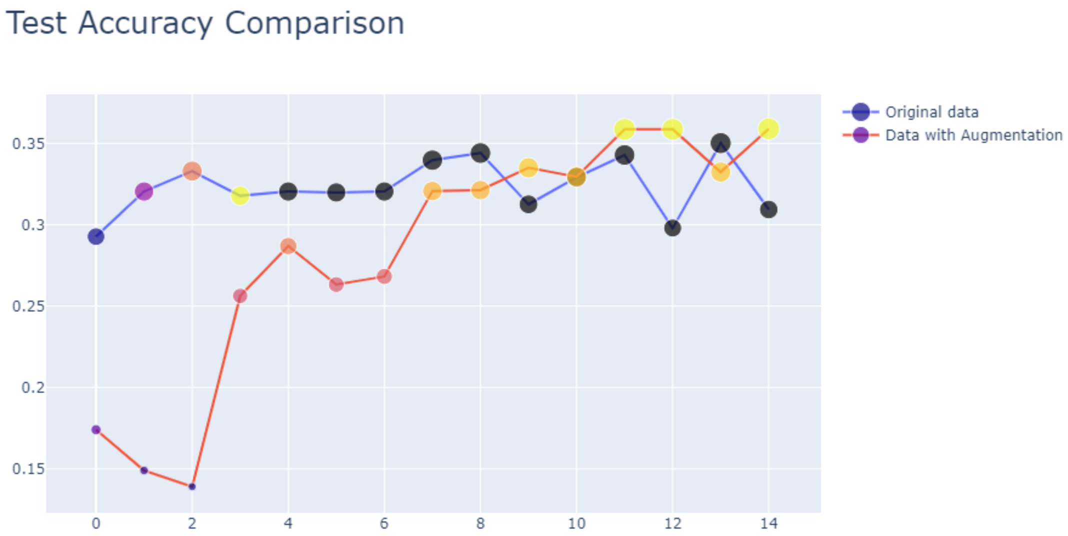

Figure 4: Mobile Net V2 Accuracy with Original data and Data With augmentation

For the algorithms with the data augmentation, the Mobile Net v2 achieved a training accuracy of 25.04% and the testing accuracy achieved stood at 35.89% whereas for the Inception v3, the training accuracy achieved was 51.8% and the testing accuracy stood at 68.57%.

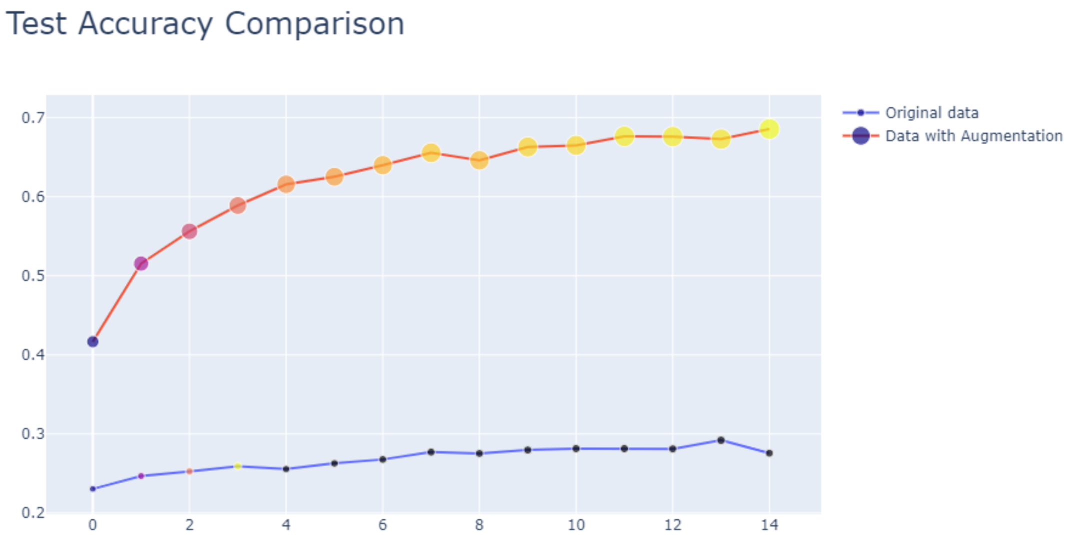

Figure 5: Inception V3 accuracy with Original data and Data With augmentation

On the other hand, for the custom architecture, the training accuracy stood at 33.79% and the testing accuracy stood at 38.22%. Here, it was observed that Inception v3 achieved higher training accuracy and testing accuracy. Similarly, the lowest training and testing accuracy was scored by Mobile Net v2. The graphical representation of each algorithm is depicted in the figure. Although with the overall comparison, the model with data augmentation achieved better than the model without the data augmentation. Moreover, to this, the Inception v3 outperformed the other algorithms.

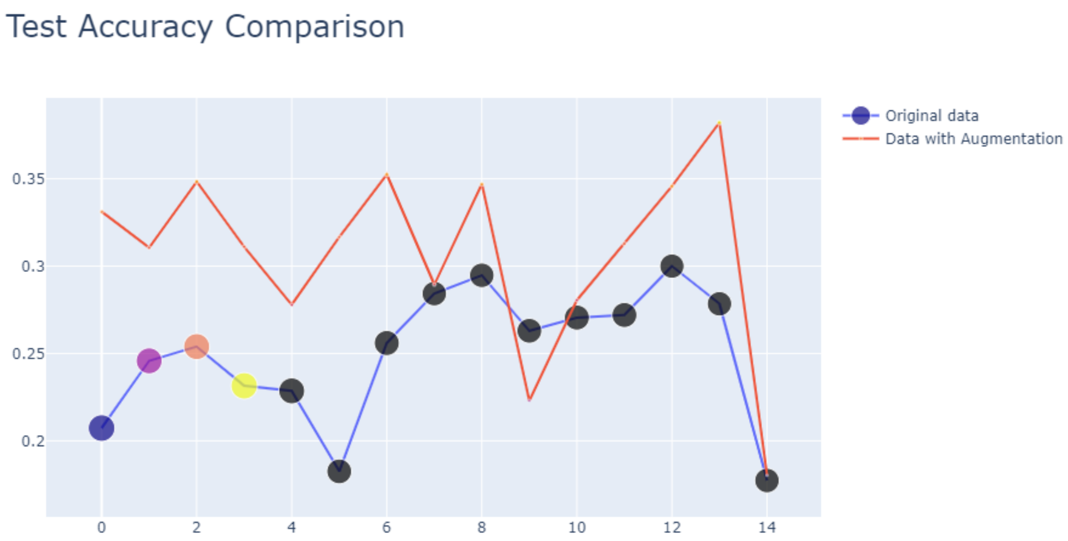

Figure 6: Custom Architecture Accuracy

6.1.2 3 Food Categories

In this case, the food categories have been reduced to 3 from 11. The categories with more number image samples were only considered. By this approach, the efficacy of the algorithm can be enhanced. The Inception v3 has achieved a training accuracy of 81.99% and the testing accuracy achieved stood at 91.65% whereas the Mobile Net v2 achieved a training and test accuracy of 86.87% and 94.2%, respectively. In the figure, here it was observed that the Mobile Net v2 outperformed the Inception v3, and it showed that the Mobile Net framework could effectively process with lesser image samples.

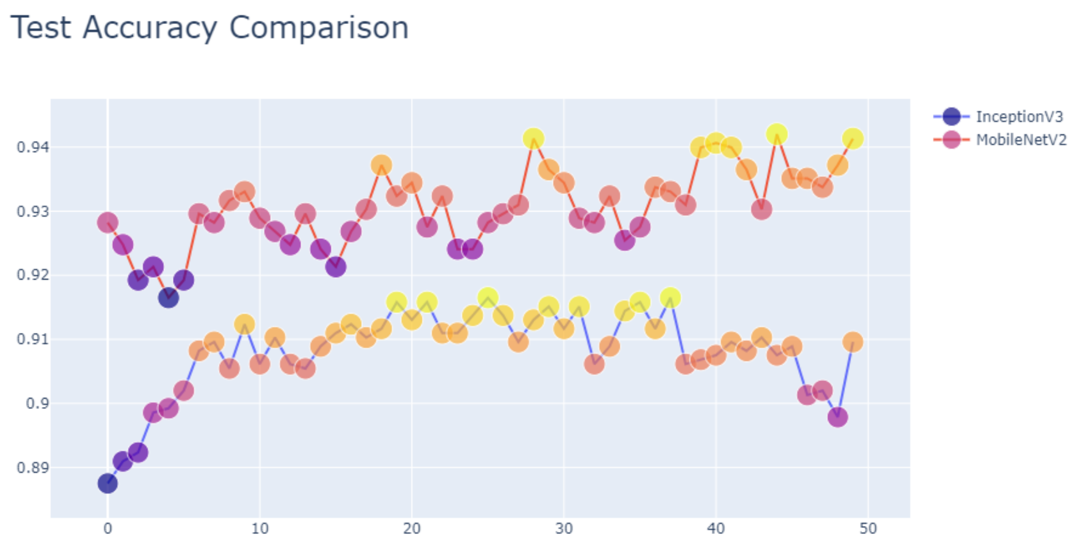

Figure 7: Accuracy in 3 Food Category

6.2 Experiment 2 / Evaluation Based on Precision and Recall

In this section, the metrics such as precision, and recall are discussed. Each metrics depends on the values of the confusion matrix which are truly positive, true negative, false positive, and false negative. In addition to these, the following subsections discuss the PRF score of each algorithm with and without augmentation considering the dataset with 11 and 3 food categories.

6.2.1 11 Food Categories

Here, for the implemented algorithms, the data considered was with augmentation and without augmentation. In the algorithms without data augmentation, the Mobile Net v2 achieved a training precision, and recall of 0.2145 and 0.4319, respectively whereas the testing precision and recall achieved was 0.2573 and 0.5216, respectively. For the Inception v3, the training precision, and recall stood at 0.1992 and 0.3983, respectively whereas the testing precision and recall achieved stood at 0.2296 and 0.4592, respectively. In the Custom Architecture, the training precision, and recall stood at 0.2305 and 0.4611, respectively whereas the testing precision and recall achieved stood at 0.242 and 0.484, respectively. Here it is observed that the highest testing precision and recall is achieved by Mobile Net v2 whereas the lowest was achieved by Inception v3.

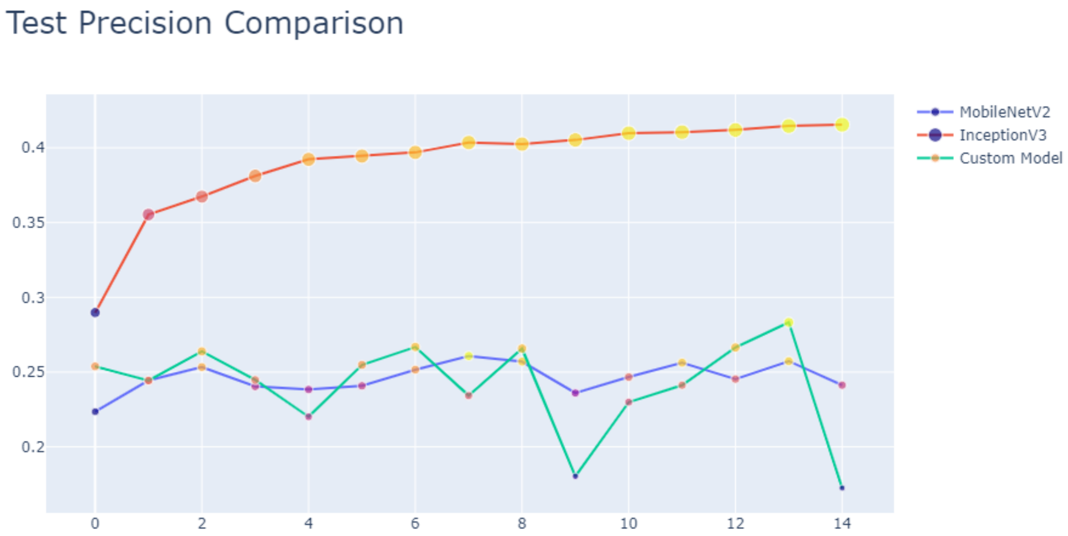

Figure 8: Precision Comparison

Whereas for the algorithms with the data augmentation, the Mobile Net v2 achieved a training precision, and recall of 0.2587 and 0.5175, respectively whereas the testing precision and recall achieved was 0.2573 and 0.5216, respectively. For the Inception v3, the training precision, and recall stood at 0.3467 and 0.6934, respectively whereas the testing precision and recall achieved stood at 0.4155 and 0.8309, respectively. In the Custom Architecture, the training precision, and recall stood at 0.2563 and 0.5127, respectively whereas the testing precision and recall achieved stood at 0.2832 and 0.5665, respectively. Here it is observed that the highest testing precision and recall is achieved by Inception v3 whereas the lowest was achieved by Mobile Net v2. This experiment showed that the evaluation of precision and recall showed that with data augmentation, the Inception v3 outperformed the other algorithms. The figure here depicts the graphical representation of each algorithm.

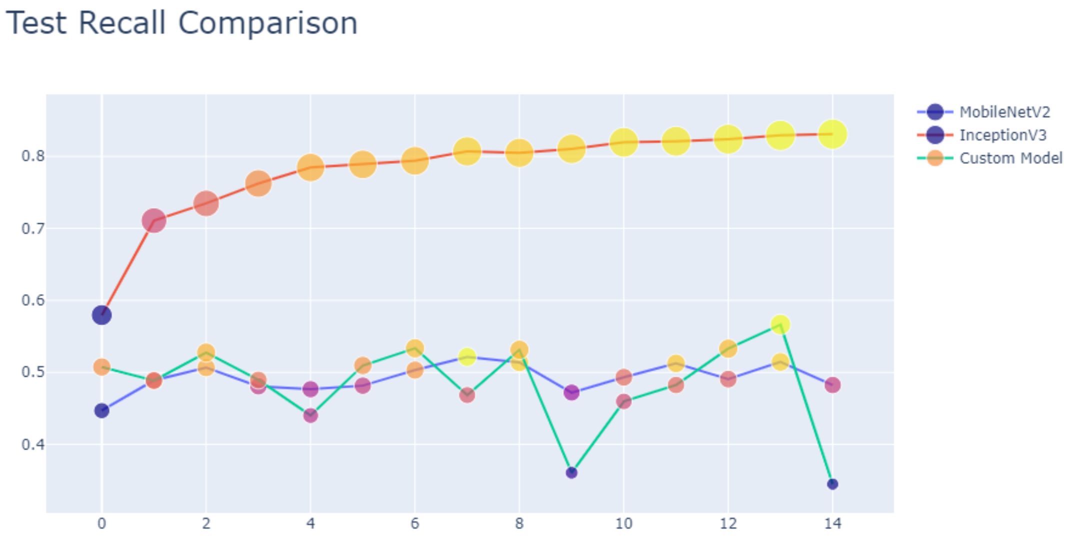

Figure 9: Recall Comparison

6.2.2 3 Food Categories

In this phase, the food categories have been reduced to 3 from 11 and the categories with more number image samples were only considered as discussed. Implementing this approach will enhance the efficacy of the algorithm. The Inception v3 has achieved a training precision and recall of 0.479 and 0.9579, respectively and the testing precision and recall achieved stood at 0.4928 and 0.9903, respectively whereas the Mobile Net v2 achieved a training precision and recall stood at 0.4855 and 0.9711, respectively and test precision and recall of 0.4972 and 0.9938, respectively.

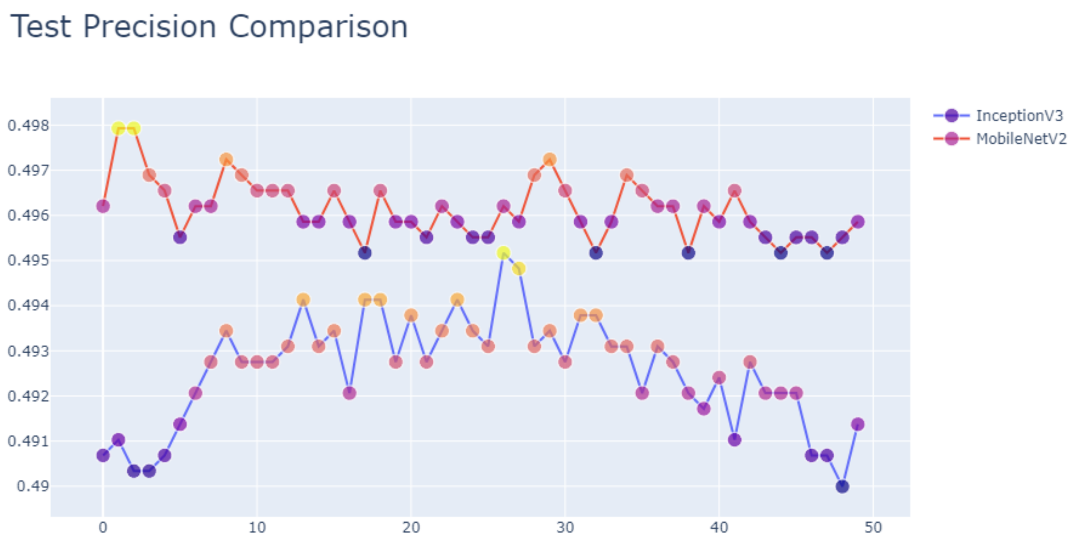

Figure 10: Precision Comparison

In the figure, here also it was observed that the Mobile Net v2 outperformed the Inception v3, and it also showed that the Mobile Net framework could effectively process with lesser image samples.

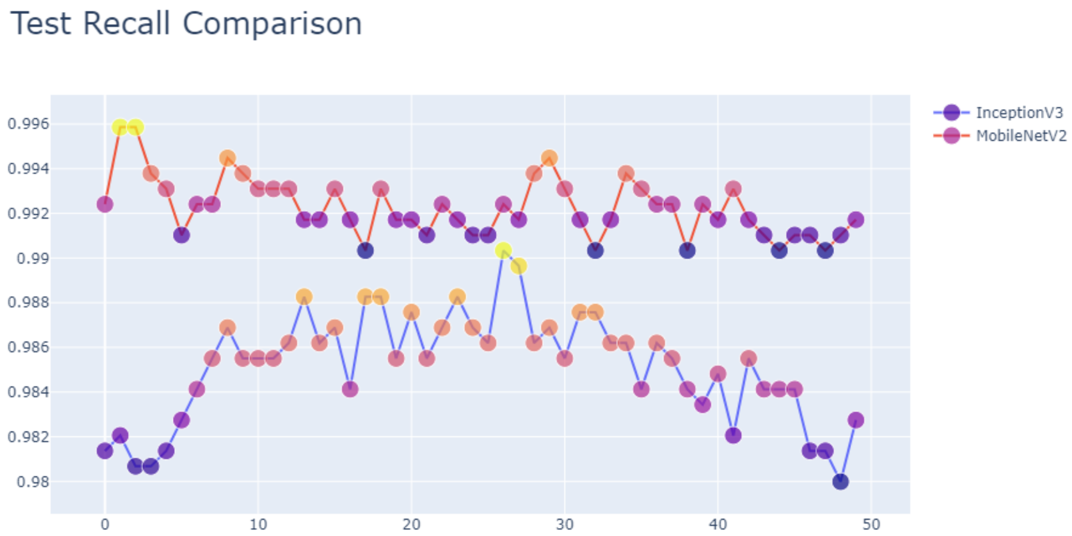

Figure 11: Recall Comparison

6.3 Discussion

Through our study, our fundamental motto was to introduce a robust model to classify the food products and state the calories of the recognized food. For the implementation, we have utilized their deep neural network algorithms out of which two are pre-trained named Mobile Net V2 and Inception V3 whereas the other one is the Custom Architecture. These algorithms rely on the framework of a convolutional neural network where the convolutional layers are constructed. Initially, we obtained the dataset for our model from the Kaggle repository which was named Food-11 *Food-11 image dataset* (n.d.). In the obtained dataset, image samples were split into 11 categories with each category consisting of training, validation, testing division. To evaluate the performance of each algorithm in-depth, 3 categories were also considered in the later phase. In addition to this, during the phase of pre-processing, two approaches are considered which are with data augmentation and without data augmentation. In data augmentation, several parameters of the samples are tuned for an effective result. Once the completion of pre-processing, the features are extracted internally by the training algorithms. The model is trained using the DNN algorithms. We processed each model and evaluated the performance of these to consider the optimal algorithm for further implementation. Here, we considered metrics such as accuracy, precision, and recall for the evaluation of the performance. First, in the figure, it was observed that the model with the data augmentation performed well in comparison with the model without the data augmentation. Then the model was evaluated based on dataset categorization where 11 and 3 food categories were implemented. During the evaluation of metrics in the 11 food categories, it was noted that the Inception V3 outperformed other models, but when the number of samples depleted such as in 3 food categories, the Mobile Net V3 outperformed the previous and the other algorithms. On the other hand, the custom architecture was poorly performed due to which it was not considered for further implementations.

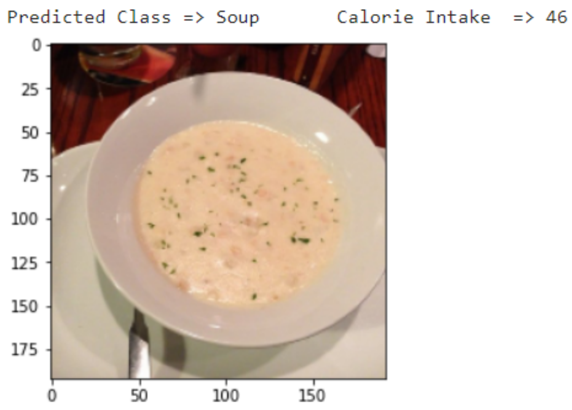

Figure 12: Class Soup

In the implementation of our model, we have fed the food images which will recognize the food and state the calories of each. Considering the figures, we have observed both the class soup and desert have been predicted. Furthermore, the calories have been stated.

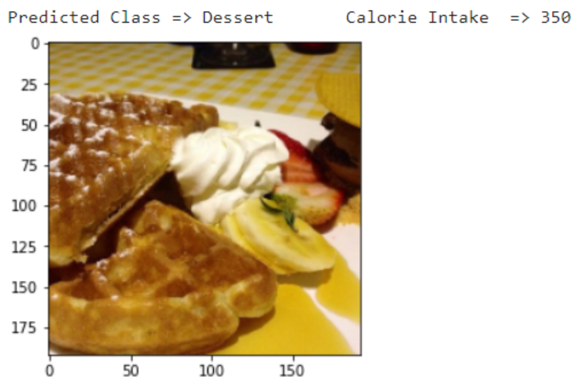

Figure 13: Class Desert

7 Conclusion

The growing population and the demand for food with the surging technological advances have brought these in corroboration. In addition to these, the imbalance in the consumption of food may imbalance the nutrient intake. Food is a crucial part of living being survivability, although a balanced diet also matters. To ensure a balanced diet, there are various essential factors to be considered such as calories, carbohydrates, proteins, fats, fibres, vitamins, minerals, and water. The foremost factor of all these is calories which is the energy from the consumed food. This energy is sourced for food and utilized for physical human activities. Although, calories shall be availed until a certain threshold upon which it may have an adverse effect on the health of a person. Therefore, an effective and autonomous approach for dietary assessment and monitoring systems must be suggested. Therefore, in our study, we introduced a model for food classification which will denote the calories of the food recognized. This model is developed utilizing computer vision technology. In the model, the deep neural network algorithms were considered which are Mobile Net v2, Inception v3, and custom architecture. Out of these, the first two algorithms were pre-trained on the weights of the Image Net database. In the custom architecture, the custom convolutional layers were developed. These were then evaluated using the metrics such as Accuracy, Precision, and Recall. In the experiment for the detailed evaluation, it was observed that the Inception V3 outperformed other models when implemented with the 11 food categories but in the case of 3 categories, the MobileNet V2 outperformed. Whereas the custom architecture performed poorly hence eliminated. This model could be enhanced in the future scope of work through appropriate and optimal tuning of the parameters. In addition to this, the application of this model can be extended in various domains such as in the food industry, medical, etc.

References

Advanced Guide to Inception v3 | Cloud TPU (n.d.).
URL: *https://cloud.google.com/tpu/docs/inception-v3-advanced*

Al-Sarayreh, M., Reis, M., Yan, W. and Klette, R. (2020). Potential of deep learning and snapshot hyperspectral imaging for classification of species in meat, *Food Control* **117**: 107332.

Burkapalli, V. C. and Patil, P. C. (2020). Transfer learning: Inceptionv3 based custom classification approach for food images', *ICTACT Journal on Image and Video Processing* **11**(01).

Fahira, P., Wibisono, A., Wisesa, H., Rahmadhani, Z., Mursanto, P. and Nurhadiyatna, A. (2019). Sumatra traditional food image classification using classical machine learning, pp. 1–5.

Fakhrou, A., Kunhoth, J. and Al-ma'adeed, S. (2021). Smartphone-based food recognition system using multiple deep cnn models, *Multimedia Tools and Applications* **80**: 1–22.

Food-11 image dataset (n.d.).
URL: *https://kaggle.com/trolukovich/food11-image-dataset*

Giovany, S., Putra, A., Hariawan, A. and Wulandhari, L. (2017). Machine learning and sift approach for indonesian food image recognition, *Procedia Computer Science* **116**: 612–620.

Hafiz, R., Haque, M., Rakshit, A. and Uddin, M. (2020). Image-based soft drink type classification and dietary assessment system using deep convolutional neural network with transfer learning, *Journal of King Saud University - Computer and Information Sciences* .

Hu, J., Zhou, C., Zhao, D., Zhang, L., Guijun, Y. and Chen, W. (2019). A rapid, low-cost deep learning system to classify squid species and evaluate freshness based on digital images, *Fisheries Research* **221**.

Inunganbi, S., Seal, A. and Khanna, P. (2018). *Classification of Food Images through Interactive Image Segmentation*, pp. 519–528.

Islam, K. T., Wijewickrema, S., Pervez, M. and O'Leary, S. (2018). An exploration of deep transfer learning for food image classification, *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–5.

Islam, M., Billah, M. and Yousuf, M. (2019). Automatic plant detection using hog and lbp features with svm, *International Journal of Computer (IJC)* **33**: 26–38.

Islam, M. T., Karim Siddique, B. N., Rahman, S. and Jabid, T. (2018). Image recognition with deep learning, *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, Vol. 3, pp. 106–110.

Madgi, M., Danti, A. and Anami, B. (2015). Combined rgb colour and local binary pattern statistics features-based classification and identification of vegetable images, *International Journal of Applied Pattern Recognition* **2**: 340.

MobileNetV2: The Next Generation of On-Device Computer Vision Networks (n.d.).
URL: *http://ai.googleblog.com/2018/04/mobilenetv2-next-generation-of-on.html*

Momeny, M., Jahanbakhshi, A., Jafarnezhad, K. and Zhang, Y.-D. (2020). Accurate classification of cherry fruit using deep cnn based on hybrid pooling approach, *Postharvest Biology and Technology* **166**: 111204.

Nasiri, A., Taheri-Garavand, A. and Zhang, Y.-D. (2019). Image-based deep learning automated sorting of date fruit, *Postharvest Biology and Technology* **153**: 133–141.

Nguyen, D. T., Zong, Z., Ogunbona, P., Probst, Y. and Li, W. (2014). Food image classification using local appearance and global structural information, *Neurocomputing* **140**.

Pande, A., Munot, M., Sreeemathy, R. and Bakare, R. (2019). An efficient approach to fruit classification and grading using deep convolutional neural network, *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pp. 1–7.

Ponce Real, J., Aquino, A. and Andujar Marquez, J. (2019). Olive-fruit variety classification by means of image processing and convolutional neural networks, *IEEE Access* **7**: 147,629–147,641.

Rahmadani, D. (2020). Food detection using histogram of oriented gradient (hog) as feature extraction and k-nearest neighbors (k-nn) as classifier, *International Journal of Advanced Trends in Computer Science and Engineering* **9**: 219–225.

Rajayogi, J. R., Manjunath, G. and Shobha, G. (2019). Indian food image classification with transfer learning, *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, Vol. 4, pp. 1–4.

Thiodorus, G., Sari, Y. A. and Yudistira, N. (2021). *Convolutional Neural Network with Transfer Learning for Classification of Food Types in Tray Box Images*, Association for Computing Machinery, New York, NY, USA, p. 301–308.
URL: *https://doi.org/10.1145/3479645.3479687*