

**Table of Contents**

[**1. Introduction 2**](#_bpufx6vp18tv)

[Problem Statement 2](#_32ilqzr9r3un)

[Key Objectives 2](#_hm3zd6v20yoa)

[Stakeholders 3](#_illbee3e3wcx)

[**2. Data Understanding 4**](#_hoi6q0shklhw)

[**3. Model Architecture 5**](#_x1u0s8101ppo)

[Googlenet 5](#_aah52ajrx9wh)

[Mobilenetv3 7](#_1knlp66b36qd)

[Nasnet 7](#_yuqyz28ijcoa)

[**4. Evaluation 8**](#_52qj2lov8v5a)

[Result and Analysis 8](#_8w8pm0gj2uz7)

[Limitation 11](#_enl24d2jbdbd)

[**5. Conclusion 13**](#_vw3gl58lmyvi)

[**REFERENCE 14**](#_s5qmwrofunoh)

# 

# Introduction

Deep learning significantly impacts various industries by enhancing the analysis and interpretation of large volumes of unstructured data, such as images and videos. This branch of machine learning is notably effective in identifying and extracting valuable patterns and insights. Applying deep learning through food image classification in the food industry introduces transformative improvements by boosting food item identification accuracy and efficiency.

This project focuses on the reliable classification of food images from digital photographs using cutting-edge computer vision and machine learning techniques. This technology is increasingly crucial as industries aim to automate and optimise tasks traditionally performed by humans. In the food service sector, automated food recognition can aid in inventory management, menu customisation, and adherence to health regulations by precisely identifying and recording served food items. Nutritionists and dietitians are leveraging this technology to monitor better and analyse patients’ dietary habits, thus providing more personalised and accurate nutritional advice, which leads to improved management of health conditions such as diabetes, obesity, and heart disease, where diet plays a crucial role in patient care.

## Problem Statement

The challenge lies in developing a robust model to classify food items under different conditions and presentations accurately. Food items can appear vastly different depending on the cuisine, preparation style, and serving method, which introduces complexity in the classification task. Additionally, the model must perform well across diverse lighting conditions and image qualities.

## Key Objectives

1. **Accuracy Improvement**: Enhance the accuracy of food classification models to handle various food items and presentation styles.
2. **Efficiency**: Ensure the model can classify images efficiently to support applications.

## Stakeholders

1. **Healthcare Professionals**: Dieticians and nutritionists who could use this technology to track and analyse patients' dietary habits more efficiently.
2. **Restaurant Owners and Chefs** Can use this technology for inventory management, portion control, and menu planning based on visual analysis of dishes served.
3. **Fitness Tracking Companies**: Integrating food image classification to enhance the accuracy of dietary tracking within fitness apps.
4. **Consumers**: Individuals who use apps to monitor their eating habits or to find nutritional information based on meal photos.

# Data Understanding

This dataset features 101 distinct food categories, encompassing 101,000 images. Each category has 250 test images and 750 training images, all of which have undergone manual review for quality assurance.

# Model Architecture

## Googlenet

GoogLeNet, also known as Inception v1, is a breakthrough in the design of neural networks for image processing, focusing on being deep yet efficient. It introduces a unique feature called the Inception module, which contains several different-sized filters that work simultaneously. This setup allows the network to collect detailed information from images at multiple scales. By combining the results from these filters, GoogLeNet can handle complex images without needing more computing power. The architecture has 22 layers, which helps it perform well by using resources efficiently and preventing it from learning irrelevant details. The diagram shows how GoogLeNet organises its convolutional, pooling, and fully connected layers to process images effectively.

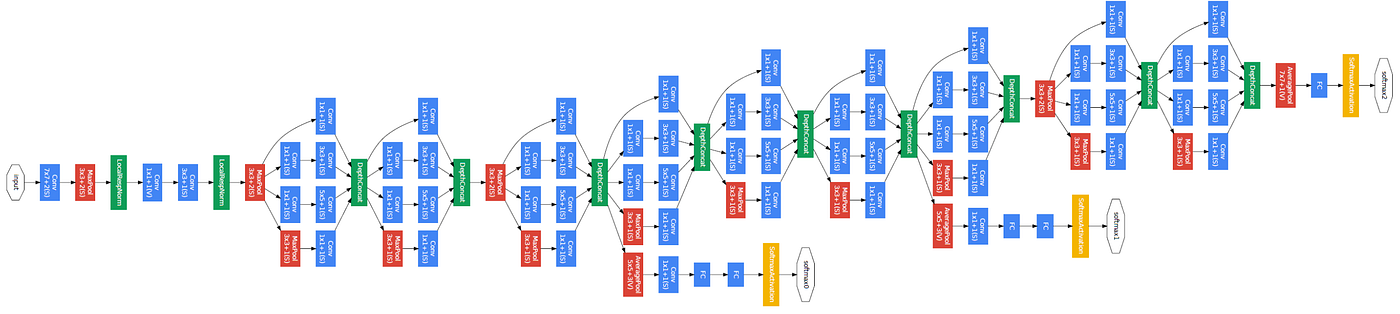
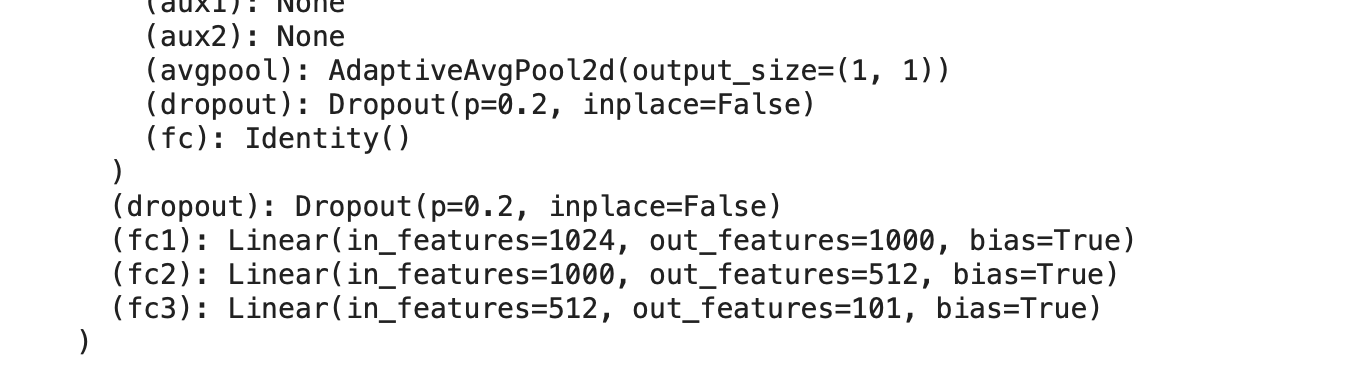


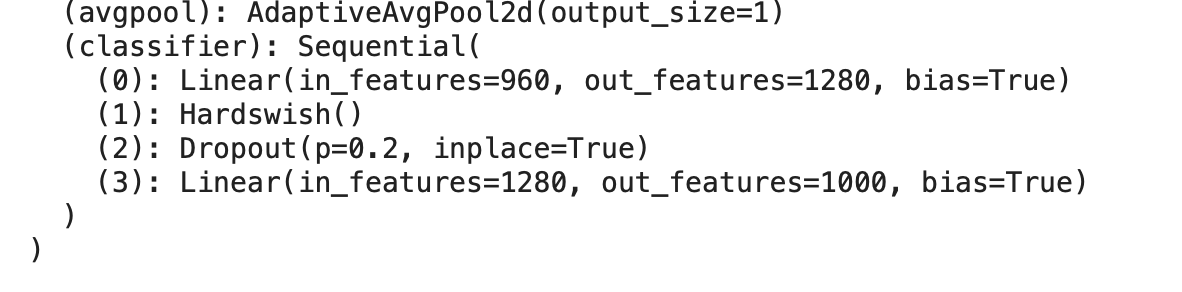
Figure: *GoogleNet Architechture (Source: (PyTorch, n.d.))*

Added Layers for the experiment:



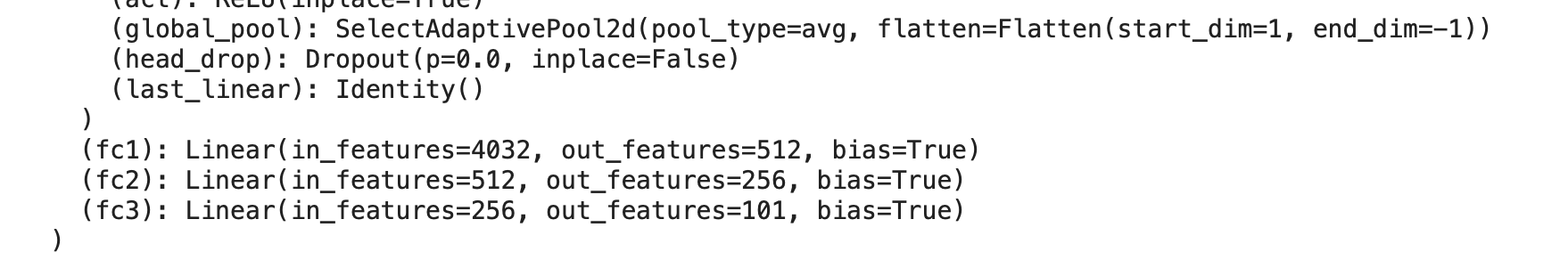
## Mobilenetv3

MobileNetV3 is a streamlined neural network designed for mobile devices. It uses specialised techniques to reduce power and computational needs while maintaining accuracy. The network, available in small and large versions, optimises performance for mobile apps by intelligently arranging layers to process images efficiently and save energy.



## Nasnet

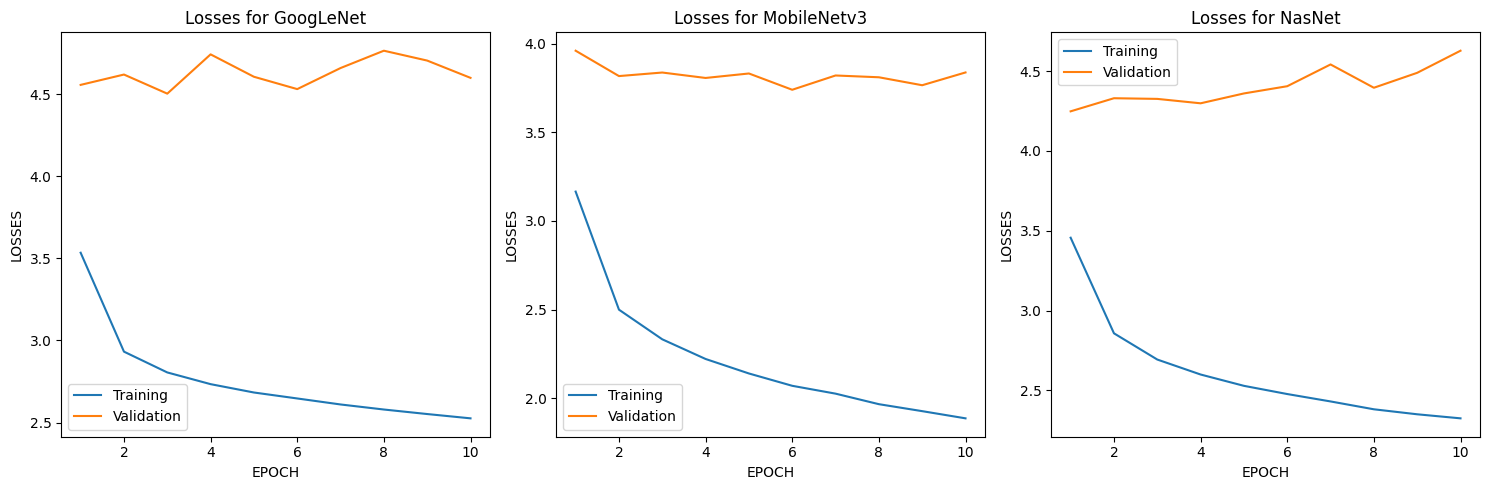
NASNet is a versatile neural network architecture developed using an automated process that tests many network structures to find the most efficient one. It's designed to be scalable, which can be adjusted to perform well across different devices and tasks by optimising its performance and efficiency.

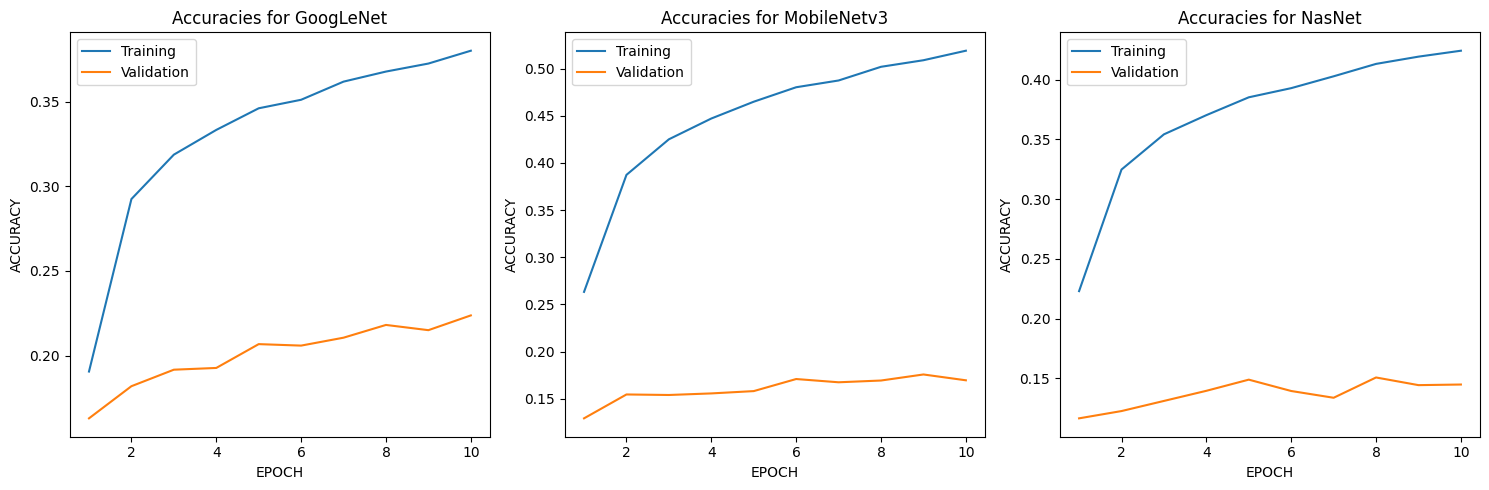


# Evaluation

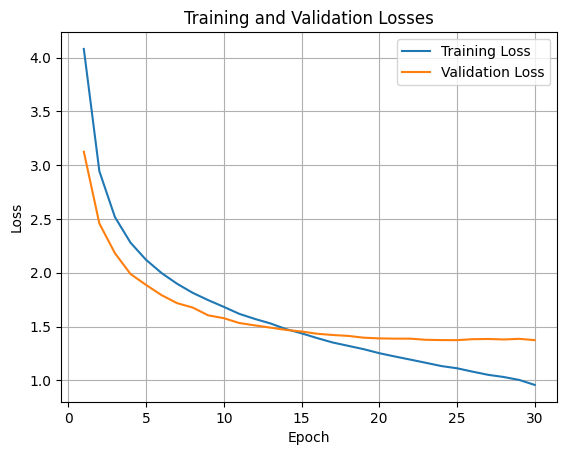
## Result and Analysis

This project evaluated three pre-trained models, GoogleNet, MobileNetv3, and NasNet, by freezing all their layers and altering the model head to incorporate an average pooling layer and three fully connected layers. After training each model for 10 epochs, GoogleNet demonstrated the best performance, achieving higher accuracy.

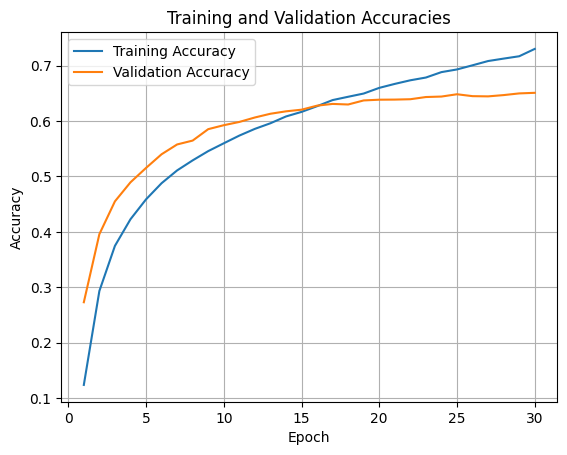




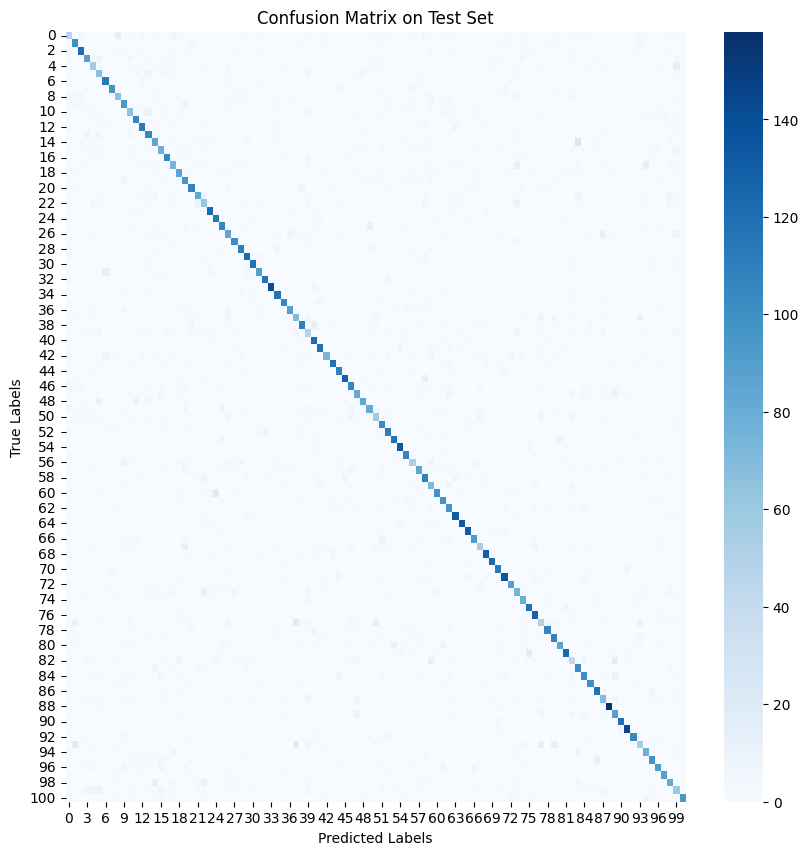
To further enhance GoogleNet's performance, the model underwent fine-tuning by unfreezing two layers. During this phase, strategies like data augmentation and early stopping were implemented. Additionally, a specific optimiser and loss criterion were selected to refine the training process and improve the model's ability to generalise to unseen data.



This figure shows the training and validation loss of a model over 30 epochs as early stopping triggered. Both losses decrease as the number of epochs increases, indicating that the model is learning and improving its prediction accuracy over time.



This plot displays the training and validation accuracy of a model across 30 epochs. Both accuracies increase over time, with training accuracy consistently higher, suggesting the model is learning effectively, though possibly beginning to overfit.



The confusion matrix shown illustrates the classification results for 101 categories, with a prominent diagonal indicating high accuracy for most classes. Off-diagonal marks represent misclassifications, which are minimal, suggesting the model performs well across diverse classes.

## Limitation

1. Similar and complex dishes make it hard for the model to tell them apart, especially when ingredients overlap.
2. The dataset does not include important details like cooking style or region, which are key for correct classification.
3. The dataset must be regularly updated with new dishes and cuisines, requiring significant resources.

# Conclusion

Food image classification technology offers significant advantages by providing automated food recognition tools, which can benefit numerous sectors, including healthcare and food service. Continued model accuracy and efficiency improvements can lead to more innovative applications that enhance user engagement and operational effectiveness.

# REFERENCE

data.vision.ee.ethz.ch. (n.d.). *Food-101 -- Mining Discriminative Components with Random Forests*. [online] Available at: https://data.vision.ee.ethz.ch/cvl/datasets\_extra/food-101/.

huggingface.co. (n.d.). *NASNet*. [online] Available at: https://huggingface.co/docs/timm/en/models/nasnet [Accessed 1 May 2024].

PyTorch. (n.d.). *GoogLeNet*. [online] Available at: <https://pytorch.org/hub/pytorch_vision_googlenet/>.

Pytorch (n.d.). *MobileNet V3 — Torchvision main documentation*. [online] Available at: https://pytorch.org/vision/main/models/mobilenetv3.html [Accessed 1 May 2024].