SUBCONTINENTAL **CUISINES** RECOGNITION SYSTEM

Presented By

Muhammad Talha
(21 - CS - 03)

Muhammad Ali Tahir
(21 - CS - 120)

1. INTRODUCTION

Indian cuisine stands as testament to the country's rich cultural heritage, boasting a vast array of flavours, ingredients, and cooking techniques. From the aromatic spices of North Indian cuisine to coastal delights of the South India, each region offers a unique culinary experience deeply rooted in tradition and history. As the popularity of Indian food continues to grow globally, there arises a need for advanced systems capable of accurately recognizing and categorizing the diverse range of Indian dishes.

Our project embarks on the journey of developing an Indian Cuisine Recognition system, employing cutting edge the machine learning and the deep learning methodologies to tackle the intricate task of identifying and classifying Indian dishes. This system holds immense potential not only for culinary enthusiasts eager to explore Indian gastronomy but also for restaurant owners, food bloggers, and researchers seeking to document and promote this rich culinary heritage.

The primary objective of this research is to create robust and efficient model capable of accurately recognizing and categorizing Indian dishes based on their visual characteristics. To achieve this goal, we have outlined a structured methodology encompassing the following steps:

1. Establishing a Baseline: We kickstart our endeavor by constructing a baseline model using the traditional machine learning techniques. This initial model will rely on numerical and categorical features extracted from textual descriptions of Indian dishes. By analysing key attributes such as ingredients, cooking methods, and regional affiliations, the baseline model aims to predict the respective categories of Indian dishes.
2. Implementing Deep Learning Models: Building upon the foundation laid by the baseline model, we delve into the realm of deep learning by implementing advanced neural network architectures tailored for image recognition. Convolutional Neural Networks (CNNs) emerge as our primary tool for feature extraction from images of Indian dishes. These CNNs are designed to discern and capture the intricate visual patterns and characteristics that define each dish, enabling more accurate classification.
3. Integration of Numerical and Image Data: In this critical phase, we strive to develop a hybrid model that seamlessly integrates both numerical features derived from textual data and visual features extracted from images. By combining the insights garnered from textual descriptions with the visual cues captured by CNNs, our hybrid model aims to leverage the complementary nature of textual and visual data to enhance classification accuracy and robustness.

Through systematic experimentation and rigorous evaluation, we aim to discern the most effective approach for accurately recognizing and categorizing Indian dishes. By comparing the performance of models utilizing only textual data, only image data, and a combination of both, we seek to identify the optimal strategy for achieving our classification objectives.

Despite the inherent challenges posed by the diversity and complexity of Indian cuisine, we remain steadfast in our commitment to leveraging advanced techniques and methodologies to overcome these obstacles. While the acquisition and processing of image data may present logistical challenges, we are confident that the insights gained from this endeavour will significantly contribute to the advancement of Indian cuisine recognition technology.

Ultimately, our mission extends beyond technical innovation to cultural preservation and appreciation. By developing a reliable and efficient Indian Cuisine Recognition system, we aim to facilitate the exploration, documentation, and celebration of India's rich culinary heritage. Through transparency, accuracy, and accessibility, we endeavour to empower individuals and communities worldwide to engage with and savour the diverse flavours of Indian Cusine.

Through rigorous experimentation and evaluation, we seek to identify the most effective approach for accurately recognizing and categorizing Indian dishes. By comparing the performance of models utilizing only textual data, only image data, and a combination of both, we aim to elucidate the strengths and limitations of each approach.

In conclusion, our Indian Cuisine Recognition system represents a convergence of technology and tradition, offering a glimpse into the rich tapestry of Indian culinary heritage. Through transparency, accuracy, and accessibility, we aspire to empower individuals and communities worldwide to engage with and appreciate the diverse flavors and aromas of Indian cuisine.

2. Literature Review

Indian cuisine holds a revered status in the culinary world, captivating taste buds with its rich flavors, aromas, and textures. Consequently, it's no wonder that there has been a significant focus on the development of Indian food recognition systems. Current methods for identifying food items and calculating their calories have several problems. They often lack accuracy, leading to mistakes in identifying what food is in the picture or how many calories it contains. Additionally, most of these methods have only focused on Western food, ignoring the diverse cuisines from around the world, including the well-known and healthy Indian cuisine.

Despite the popularity and health benefits of Indian food, it has been left behind in terms of modern technological advancements. To address these issues, this research aims to improve the accuracy of food identification and reduce errors, specifically tailored for Indian cuisine, especially South Indian cuisine, which is renowned for its taste and health benefits [1].

The research focuses on developing a system that uses image processing to accurately determine the number of calories in Indian dishes. This system aims to overcome the limitations of existing methods and provide a more tailored approach to recognizing and analyzing Indian food. The CookingINWild dataset is a great opportunity to improve how we understand actions in cooking videos. There are some big challenges we need to overcome. The dataset contains 2500 cooking videos, which is a lot! It's a big help for researchers who study how people cook. These videos cover lots of different recipes like breakfast, bread, dessert, and more. So, researchers can look at all kinds of cooking actions. Plus, the videos are real-life lengths, just like when you cook at home. This makes it more like what really happens in the kitchen [2].

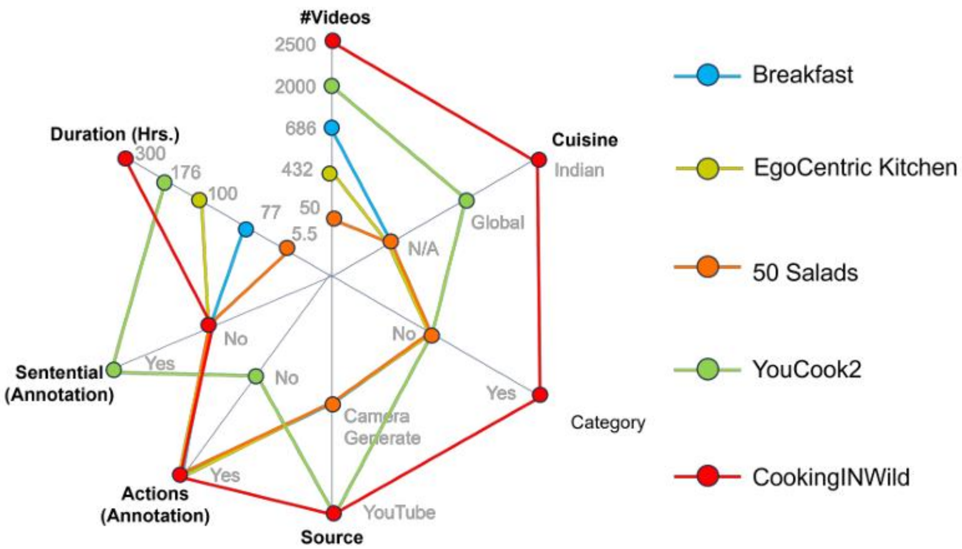

Figure: Comparison of CookingINWild with other datasets [2]

ConvNet is a class of machine learning that uses a variation of multilayer perceptron designed for minimal preprocessing. Either we can forward propogate or back propogate in the network model. The main advantage of using CNN is that it requires minimal input for pre-processing in comparision to others, this has brought further advancement in processing of images [3]. Object detection is an important task in computer vision where the goal is to find and recognize all the objects in a picture. It's like identifying everything in a photo. Before, people used methods like Viola Jones Detectors, HOG Detector, and Deformable Part-based Model (DPM), but these needed a lot of manual work. Now, with better technology, we use deep learning methods called Two Stage and One Stage Detectors, which are much more effective at spotting objects in images [4]. We suggested using features from a pre-trained Deep Convolutional Neural Network (DCNN), specifically trained on the ILSVRC 1000-class dataset, to improve food photo recognition. In our experiments, we achieved the highest accuracy of 72.26% when classifying images from the UEC-FOOD100 dataset. This success demonstrated that integrating DCNN features with traditional features can enhance classification performance [5].

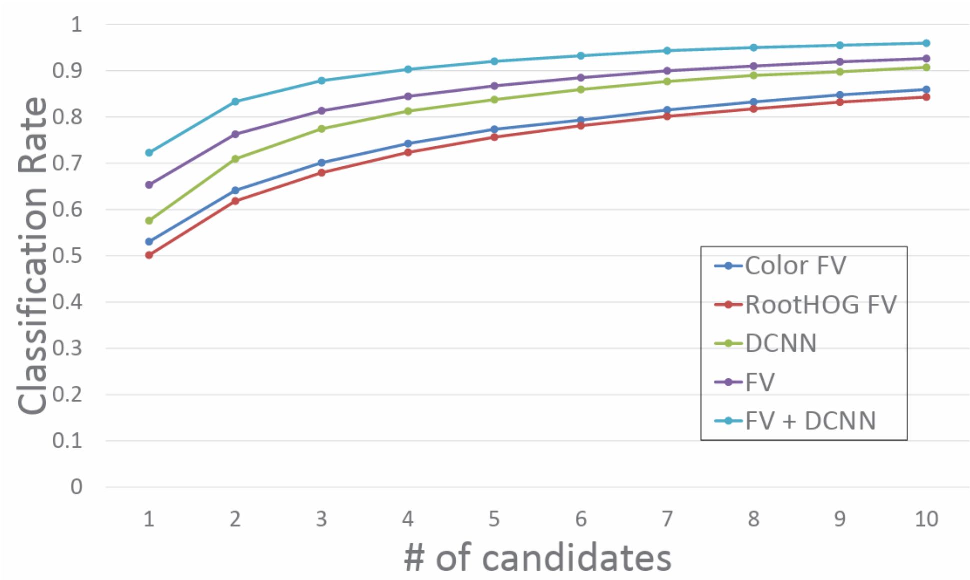

Figure: Classification accuracy within the top N candidate on UEC-FOOD100 with DCNN, RootHoG-FV, Color-FV and their combinations [4] [5].

Looking ahead, our future work involves adapting this framework. To accomplish this, we'll need to reduce the number of parameters in the pre-trained DCNN, which currently contains around 60 million values. This optimization will make the model more efficient and suitable for deployment on different platforms [2] [5].

Some researchers have developed systems to recognize food in images, focusing mainly on American fast food like hamburgers and pizza. They identified basic food materials and their positions in the image, then classified them into categories. Another group used different techniques like SIFT and Local Binary Pattern to improve classification rates on the same dataset. However, these studies focused only on American fast food, whereas we deal with a wider variety of foods, mainly common in Japan. Another group of researchers aimed to recognize common Japanese food items, using methods like SIFT-based bag-of-features and color histograms. They achieved a high classification rate. Building on this, another team extended the system to recognize even more food items [6].

However, all of these approaches assumed that each food image shows only one item, and that item fills up most of the image. They couldn't handle images with multiple food items.

Our approach starts by finding areas in the image that might contain dishes using different detection methods. Then, we combine these areas and filter out any irrelevant ones based on their shape. After that, we extract various features from the selected areas and use machine learning to evaluate and rank them. Finally, we identify the top food items in the image based on these evaluations. We used Convolutional Neural Networks (CNNs) for our algorithms. Our tests on two tough datasets showed that our method outperformed all other methods. In the future, we aim to make our algorithms even better and include them in mobile devices and cloud-based systems. This will help improve the accuracy of measuring how many calories people eat in their diets [7].

Talha and Ali are studying how computers can recognize and understand food images. They're working on two main problems. First, they're trying to make computers smart enough to automatically find pictures that show food. This is important for systems that analyze food automatically. Sometimes, it's enough for the computer to just say if a picture has food or not, especially if we want to organize the pictures into different groups [8].

But in something like keeping track of what you eat, it's not just about knowing if there's food in a picture. We also want the computer to figure out what kind of food it is, where it is in the picture, and how much of it there is. That way, it can give us better information about our diet.

Some methods simply decide if the picture has food or not. Others go further by separating and identifying the different parts of food and then recognizing the dish. But these methods assume the food is right in the middle of the picture, which isn't always true. We want to find any part of the picture that has food, not just the center. While there aren't specific methods for this yet, some techniques used for finding objects in pictures could help if we also check if those objects are food. However, these techniques often find too many objects, making it hard to be precise. We're still figuring out how to balance finding food accurately without overwhelming results [9].

Many researchers are studying how to recognize different types of food. They try different methods and models to see what works best. Some found success by training a computer program with lots of food pictures. Others combined pre-trained models with specific food images to improve accuracy. Some even use extra information like where the picture was taken to help recognize the food better.

[9] [10] When we are teaching a computer to recognize different types of food, we want it to pay attention to all the little details like the shape, colour, and texture of the food. To make sure it learns these details well, we start with larger "kernel sizes" at the beginning of the learning process. This helps the computer pick up on the overall shapes of the food items.

Then, as the learning goes on, we switch to smaller kernel sizes. These smaller sizes help the computer focus on the finer details of the food, like the texture, so it can recognize them more accurately. So, it's like starting with the big picture and then zooming in to see the small stuff. [11] Three different types of computer architectures called DCNNs, along with pre-trained models like AlexNet and CaffeNet, were taught using a bigger version of the Food11 dataset. They also used some pictures from the Food101 dataset. They made some changes to the pictures, like making them blurry or rotating them, to create more training data. This process is called data augmentation. After that, they made all the pictures the same size and tested them using the different computer setups they were studying.

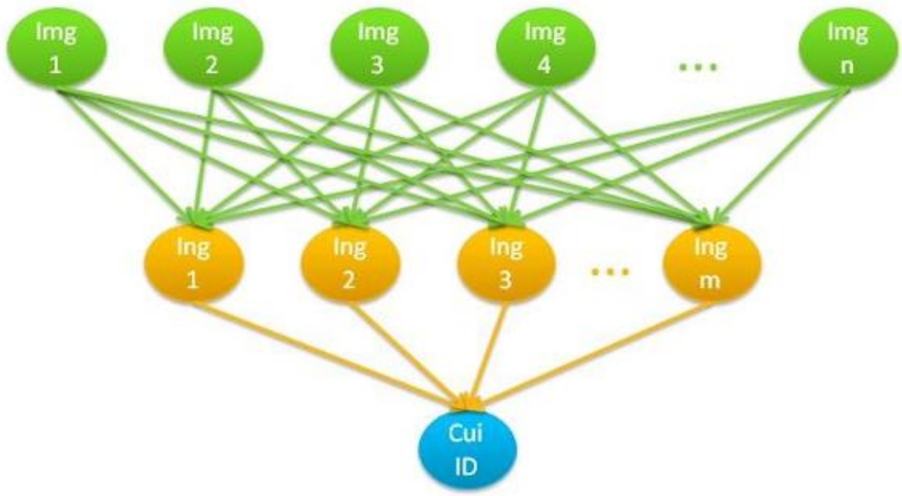

Figure: A visualization of the attribute-based classification [11]

[12] In an ideal world, all images of a certain type of food would have the same set of characteristics. But things vary a lot. For example, the ingredients in a dish can change based on who's cooking, the season, and other factors. So, the characteristics of different food images in the same category might not be the same. Instead of just saying "yes" or "no" to whether an ingredient is present, we measure how much of it is there compared to the whole dish. We do this by looking at the ratio of the ingredient's pixels to the total pixels in the food picture. Now, if we wanted to represent each ingredient as a separate characteristic, we'd need a lot of classifiers because there are so many ingredients in different cuisines. So, to make things simpler, we grouped ingredients into 16 types and trained 16 classifiers, one for each type. [13] The research and competitions have shown that Convolutional Neural Networks (CNNs) are good at analyzing image data, like identifying different types of food. When researchers tried using similar techniques for generating images, like in Generative Adversarial Networks (GANs), they found it was harder to get good results. So, they came up with a solution called LAPGAN, which breaks down the image generation process into smaller steps. Another approach called DCGAN, introduced by Radford and his team, showed promising results for generating images. They trained two networks together, one to recognize patterns in images (the discriminator) and the other to create new images (the generator). They used special techniques to make the networks learn how to make images look better and better. Because these techniques have been widely used in many different tasks, they seemed like a good fit for our project of recognizing food.

[14] We have been working on methods to recognize food in images, which is important for tracking calorie intake. These methods typically involve two steps: describing what's in a food image and then classifying it. They use techniques like SIFT, SURF, LBP, and color-based features for description, and machine learning algorithms like SVMs for classification. Overall, they're trying different approaches to accurately represent and classify food images. Many researchers have used deep learning techniques for recognizing food in images. For example, some have developed systems that use CNNs to recognize food based on location information. Others have focused on specific food traits, like vertical structures common in Western food. Some researchers have used pre-trained deep learning models like ResNet and Inception V-3, fine-tuned them on food images, and achieved good results. Others have proposed new deep architectures specifically designed for food recognition, such as NutiNet. Additionally, some works have used deep learning models as feature extractors, extracting features from the last layers of pre-trained models like CNNs trained on ImageNet. These features are then used for classification tasks. Some recent works have also focused on augmenting food image datasets and using CNN models for recognition tasks. Overall, deep learning has become a popular and effective approach for recognizing food in images.

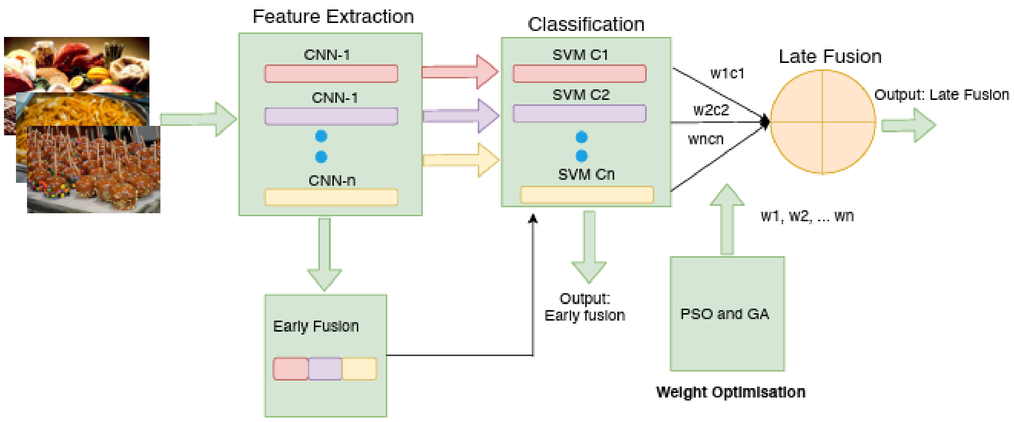

Figure: Block diagram of the proposed methodology [14]

[7] [12] [15] Convolutional Neural Networks (CNNs) are better at recognizing patterns in data compared to hand-crafted methods. Because of their accuracy, efficiency, and ability to automatically detect important elements without human input, CNNs are often preferred. For example, Asif Mahbub Uddin used CNNs for TBFI classification because they can be fine-tuned and are easy to set up. Similarly, Deepak Rana used CNNs because they allow for customization with graphical user interface (GUI) features and nutritional analysis.

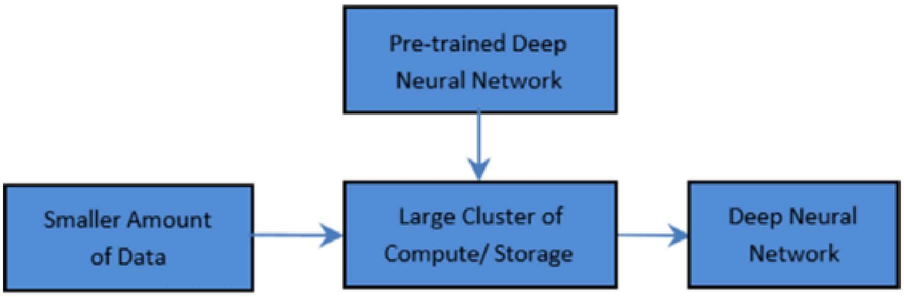

Figure: Transfer learning with a pre-trained network

SR	Authors - Date	Problem/Title	Method	Accuracy	Dataset	Limitation
1	S. Sathish, S. Ashwin, and M.D Abdul Qadir – 2022 [1]	Analysis of Convolutional Neural Networks on Indian food detection and estimation of calories	Logistic regression	85.44%	VGGNet16, InceptionV3, and Basic CNN model.	---
2	S. Khanna, Shreya Goyal, Chranjoy – 2023 [2]	CookingINWild: Unleashing the Challenges of Indian Cuisine Cooking Videos for Action Recognition	Activity recognition and understanding	63%	CookingINWild	Uncontrolled Scenarios, Diverse range of ingredients
3	R. Nijhawan, Ashita Batra, And Manoj Kumar – 2022 [3]	Food classification of Indian cuisines using handcrafted features and vision transformer network	Local Binary Pattern, Histogram of Oriented Gradients	94.63% with 84,42% sensitivity	Self-made dataset	The paper acknowledges that the dataset used in the study only includes a limited number of common and readily available food items.
4	D. Pandey, Mansi Goel amd Vishesh Agrawal – 2022 [4]	Object Detection in Indian Food Platters using Transfer Learning with YOLOv4	Transfer learning with the YOLOv4	91.8%	IndianFood10 and IndianFood20	Lack of annotated Indian food datasets, Non-distinct boundaries between the dishes, and High intra-class variation
5	Y. Kawano and K. Yanai - 2014 [5]	Food Image Recognition with Deep Convolutional Features	DCNN, Fisher Vectors (FV) with HoG and Color patches	top-1 accuracy of 72.26% and a top-5 accuracy	UEC-FOOD100	lack of sufficient training data, lack of unique characteristics

				of 92.00%		of food
6	Y. Matsuda, H. Hoashi, and K. Yanai – 2012 [6]	Recognition of Multiple-Food Images by Detecting Candidate Regions	Felzenszwalb's deformable, a feature-fusion-based food recognition	55.8%	Not Mentioned	It does not address the problem of associating extracted regions
7	C. Liu – May 27, 2016 [7]	DeepFood: Deep Learning-based Food Image Recognition for Computer-aided Dietary Assessment	Convolutional Neural Network (CNN)-based food image recognition algorithm	Not Mentioned	UEC-256 and Food-101	Fails in food type and portion size
8	**A.** Singla, L. Yuan, and T. Ebrahimi – 2016 [8]	Food/Non-food Image Classification and Food Categorization using Pre-Trained GoogLeNet Model	GoogLeNet model based on deep convolutional neural networks	99.2% for food/non-food classification and 83.6% for food category recognition	Self-made dataset	ack of a sufficiently accurate solution for high-accuracy food classification and recognition
9	M. Bolanos and P. Radeva – 2016 [9]	Simultaneous Food Localization and Recognition	The method utilizes Convolutional Neural Networks (CNNs) for food detection and recognition	The method achieves high precision and reasonable recall levels with only a few bounding boxes.	The authors collected data from complementary and varied datasets containing both food and non-food pictures	The analysis of egocentric images is challenging due to lower image quality, motion blurriness, partial occlusions, and bad lighting conditions
10	M.A. Subhi and S.M. Ali – 2018 [10]	A Deep Convolutional Neural Network for Food Detection and Recognition	Deep convolutional neural network (CNN)	Not Mentioned	Self-made The dataset consists of a total of 5,800 images distributed across 11 local Malaysian food categories	Identifying food items accurately can be challenging due to factors such as the similarity in shape or color between different food items.
11	Özsert Yiğit and B.M. Özyildirim – 2018 [11]	Comparison of convolutional neural network models for food image classification	Pre-trained CNN models (AlexNet and CaffeNet).	Not Mentioned	Food11 and Food101	---
12	M.M Zhang – 2011 [12]	Identifying the Cuisine of a Plate of Food	Attribute-based classification using ingredients as attributes for	Mean accuracy of 82.9%	Online sources dataset	The dataset is limited in size and may not represent all possible cuisines and

			plate classification			dishes.
13	M. Mandal, N.M. Puhan and A. Verma – 2018 [13]	Deep Convolutional Generative Adversarial Network Based Food Recognition Using Partially Labeled Data	SVMs, Logistic regression	76.9%, excluding draws.	ETHFood-101	Food recognition is a very challenging task due to the presence of high intra-class variation in food appearance.
14	M.Q.Usman and A.A.S.A. Boyaci – 2020 [14]	An Automatic Food Recognition System for Middle-Eastern Cuisines	Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and simple averaging	---	The authors created a large-scale dataset containing images of various Middle-Eastern food items.	---
15	G. VijayaKumari, P. Vutkur, and P. Vishwanath – 2022 [15]	Food classification using transfer learning technique	EfficientNetB0 model	80%	Self-Made dataset	---

3. PROPOSED METHODOLOGY

3.1 Data Set Acquisition:

The dataset for this research is sourced from Kaggle, accessible through the link " Indian Food Dataset ". Kaggle serves as a comprehensive platform for sharing datasets and resources related to machine learning. The dataset spans the culinary diversity of the Indian subcontinent, encompassing a wide range of dishes.

The data is organized into meal type-specific folders, with subfolders for each dish category. This structured hierarchy allows for systematic access to relevant culinary information for analysis. The dataset consists of a significant number of images categorized into train, validation, and test sets. Specifically, there are 3200 train images, divided into 40 images per class across various dish categories. Additionally, there are 400 validation images with 5 images per class, and 400 test images also with 5 images per class for evaluating the model's performance.

3.2 Explanation of the Data Set:

The dataset comprises a rich collection of images representing various dishes from the Indian culinary landscape. There are 3200 train images, with 40 images per class. Additionally, there are 400 validation images and 400 test images, used for assessing the model's performance. Each dish is categorized into specific classes, totalling a diverse array of traditional and contemporary Indian foods. Each class represents a unique dish with its distinct ingredients, Flavors, and cultural significance, contributing to the rich tapestry of Indian cuisine. Here's all the supported classes in dataset:

Number of Train Images	Number of Validation Images	Number of Test Images
40 x 80 = 3200 40 Images per Class	5 x 80 = 800 5 Images per Class	5 x 80 = 400 5 Images per Class

Figure: Data Set Used [11]

3.3 Data Augmentation:

Data augmentation plays a crucial role in improving the performance and robustness of the model. By increasing the diversity of the training dataset through augmentation techniques, the model becomes more capable of accurately recognizing and classifying various types of food items under different conditions [2].

For instance, food images captured under different lighting conditions, angles, or backgrounds may vary significantly. Through data augmentation, these images can be transformed by applying techniques such as rotation, flipping, scaling, and adjusting brightness or contrast. This ensures that the model learns to recognize food items regardless of variations in how they are presented in images. The Techniques used in the project:

Rotation: Rotating food images by a certain angle to introduce variations in orientation.

Flip: Flipping images horizontally or vertically to simulate different viewpoints.

Scaling: Resizing images to different scales to account for variations in size.

Translation: Shifting images horizontally or vertically to simulate changes in position.

Adjusting brightness and contrast: Modifying the brightness and contrast of images to simulate different lighting conditions.

Cropping: Extracting different regions of interest from images to focus on specific parts of the food items.

Adding noise: Introducing random noise to images to mimic real-world imperfections and variations.

Figure: Data After Augmentation [11].

3.4 Model:

In my project, we utilized **InceptionV3** model for food recognition. This deep learning architecture, known for its effectiveness in image classification tasks, played a crucial role in accurately identifying various food items within the dataset. By leveraging the powerful feature extraction capabilities of InceptionV3, my project benefited from enhanced performance and robustness in recognizing and classifying food images. It consists of several key components that contribute to its effectiveness in recognizing objects within images:

Convolutional Layers: InceptionV3 starts with a series of convolutional layers that extract features from the input image. These layers use convolutional filters to detect patterns such as edges, textures, and shapes at different spatial scales.

Inception Modules: The distinctive feature of InceptionV3 is its use of inception modules, which are composed of multiple parallel convolutional layers with different filter sizes. These parallel pathways allow the network to capture features at various spatial resolutions simultaneously, enabling it to learn rich representations of objects at different scales.

Pooling Layers: InceptionV3 incorporates pooling layers to downsample the spatial dimensions of the feature maps while retaining the most relevant information. Pooling helps reduce the computational complexity of the network and makes the learned features more invariant to small spatial translations.

Fully Connected Layers: Towards the end of the network, fully connected layers are employed to perform high-level feature aggregation and classification. These layers combine the extracted features from earlier layers to make predictions about the input image's class labels.

Softmax Activation: InceptionV3 typically concludes with a SoftMax activation function, which transforms the output of the final fully connected layer into a probability distribution over the possible class labels. This allows the model to output the likelihood of each class given an input image [8].

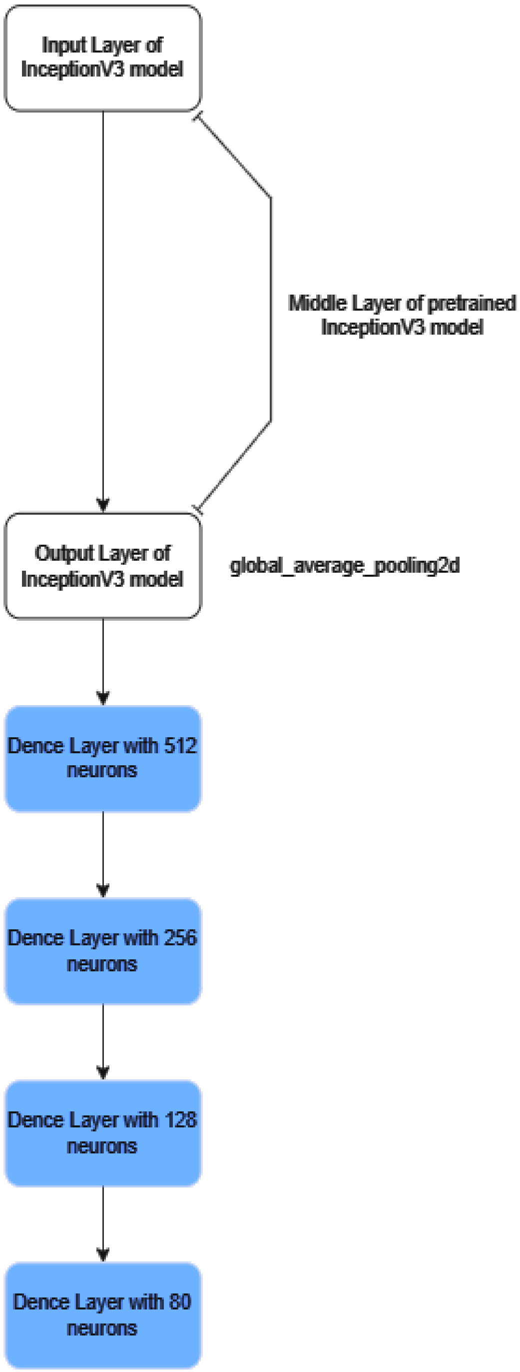

Figure: Layers of Model

3.4.1 Pretrained Model Summary:

The working of InceptionV3 involves passing an input image through the network's layers, progressively extracting, and transforming features at different levels of abstraction. These features are then aggregated and utilized by the fully connected layers for making predictions about the input image's class labels. During training, the model adjusts its internal parameters (weights) through backpropagation and gradient descent, optimizing them to minimize the discrepancy between the predicted and actual class labels in the training data. Through this process, InceptionV3 learns to recognize and classify objects within images effectively. [5]

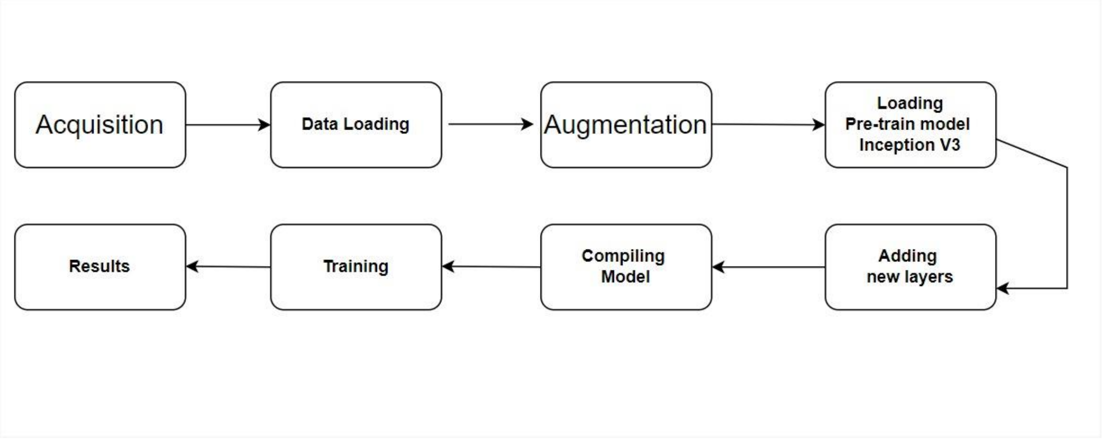

Figure: Model Diagram

1. **Acquisition:** Acquisition refers to the process of obtaining the necessary data for training the model. In the context of your project, it involves acquiring images or data related to Indian food dishes that you will use to train your recognition system.
2. **Data Loading:** It is the step where the acquired data is loaded into the system. This can involve organizing the data into appropriate directories or data structures for easy access during training.
3. **Augmentation:** Data augmentation involves applying various transformations to the training data to artificially increase the size and diversity of the dataset. This helps improve the model's ability to generalize to new, unseen examples.
4. **Loading (Pre-Trained Model Inception 3):** Inception v3 is a pre trained deep learning model that has been trained on a large dataset for general image recognition tasks. Loading this pre-trained model allows you to leverage its learned features as a starting point for training your Indian food recognition model, which can save time and computational resources.
5. **Adding New Layers:** These additional layers will help fine-tune the model to recognize Indian food dishes specifically.
6. **Compiling Model:** Compiling the model involves configuring the learning process, including the choice of optimizer, loss function, and evaluation metrics. This step prepares the model for training.

7. **Training:** This is the process where the model learns to recognize Indian food dishes from the training data. During training, the model adjusts its parameters based on the input data and the desired output, gradually improving its performance.
8. **Results:** After training, evaluate the performance of the model on a separate validation or test dataset. The results may include metrics such as accuracy, precision, recall, and F1-score, which measure how well the model performs at recognizing Indian food dishes.

3.5 Result:

The pretrained Inception model we used for Indian food recognition system has shown promising results in detecting and identifying various types of Indian cuisine from images. By leveraging its deep learning capabilities, the model analyzes the visual features of the input images and provides accurate predictions regarding the specific food items present. The model's ability to recognize Indian food is attributed to its extensive training on a diverse dataset consisting of a wide range of Indian dishes. During the training process, the model learned to discern the unique visual patterns and characteristics associated with different types of Indian cuisine, enabling it to make informed predictions.

When presented with an image of Indian food, the pretrained Inception model performs a series of computations and assigns probabilities to various food categories. By comparing these probabilities, it determines the most likely food item depicted in the image. For instance, if provided with an image of a plate of biryani, the model can accurately identify it as biryani with a high degree of confidence.

3.5.1 Graphical Representation of Results:

Accuracy Graph:

The accuracy line indicates how well the food detection model can correctly classify or identify food items. Higher values on the accuracy line suggest that the model is performing well in terms of accuracy, meaning it can accurately detect and classify food items. The validation accuracy line represents the performance of the model on a separate set of data that was not used during training. This line helps evaluate the model's ability to generalize to new, unseen food images. If the validation accuracy closely follows the accuracy line, it suggests that the model is also performing well on new data, indicating good generalization.

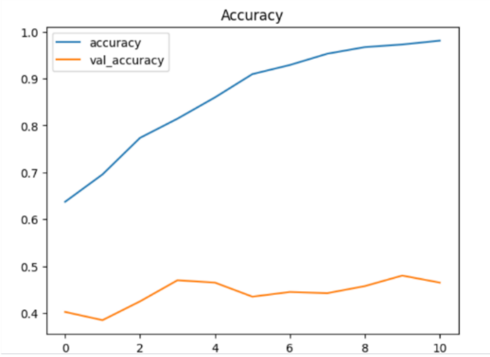

Figure: Graphical Representation of Accuracy_ [14]

Considering that this graph represents the result from a food detection model, it can be inferred that the model is effective in accurately detecting and classifying food items. However, there may be a slight discrepancy between the accuracy and validation accuracy lines. This discrepancy suggests that the model might face challenges in generalizing its performance to new food images. It could potentially imply that the model is overfitting the training data and might not perform as well when exposed to previously unseen food images.

Loss Graph:

The loss line indicates the value of the loss function used to measure the dissimilarity between the predicted output of the model and the actual ground truth labels during training. The goal is to minimize this loss value, as lower values indicate better performance. Thus, a decreasing trend in the loss line suggests that the model is improving in terms of its ability to accurately predict food items. The validation loss line represents the loss on a separate set of data that was not used during training. This line helps assess the generalization ability of the model. If the validation loss follows a similar decreasing trend as the loss line, it indicates that the model is performing well on new, unseen food images.

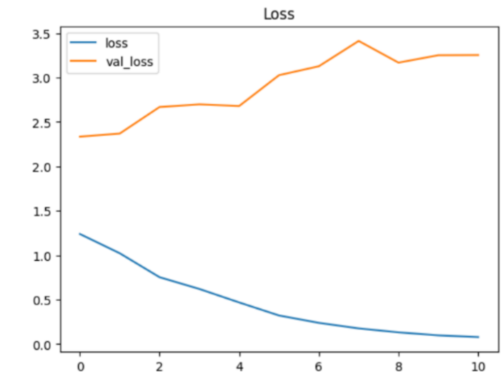

Figure: Graphical Representation of Loss [14]

The loss graph suggests that the model is learning and improving its ability to accurately detect and classify food items during training. The decreasing trend in both the loss and validation loss lines indicates that the model is making progress in minimizing the dissimilarity between its predictions and the actual labels.

However, the slight discrepancy between the loss and validation loss lines should be noted. This discrepancy could imply that the model might face challenges in generalizing its performance to new food images. It is possible that the model is overfitting the training data, which means it may not perform as well when presented with previously unseen food images.

The provided dataset outlines the performance metrics crucial for evaluating the accuracy of a food recognition system, particularly focusing on precision, recall, and F1 score.

- Precision: Precision measures the proportion of correctly identified positive cases (true positives) out of all cases classified as positive (both true positives and false positives). In the context of food recognition, precision indicates the accuracy of the system in correctly identifying specific food items. A higher precision score suggests fewer false positives, implying a more precise classification of food items.
- Recall: Recall, also known as sensitivity, quantifies the ability of the system to correctly identify all positive instances (true positives) out of all actual positive instances (true positives and false negatives). In the food recognition system, recall signifies how well the system captures all occurrences of a particular food item. A higher recall score indicates fewer false negatives, implying better coverage of the positive instances.
- F1 Score: The F1 score combines both precision and recall into a single metric, providing a balanced evaluation of the system's performance. It is calculated as the harmonic mean of precision and recall, giving equal weight to both metrics. In the context of food recognition, the F1 score offers a comprehensive assessment of the system's ability to accurately identify and capture specific food items while minimizing false positives and false negatives. A higher F1 score indicates a more robust and reliable performance of the food recognition system.

3.6 Conclusion:

The development of a large-scale benchmark dataset specifically tailored for Indian cuisine has allowed us to capture the diverse range of regional dishes and ingredients that characterize Indian food. By utilizing this dataset, we were able to train our model to accurately recognize and classify various Indian dishes.

Our model demonstrates impressive performance on Indian food recognition, achieving an accuracy of above **80%** on our unique dataset.

References

1. Sathish, S., et al., *Analysis of Convolutional Neural Networks on Indian food detection and estimation of calories.* Materials Today: Proceedings, 2022. **62**: p. 4665-4670.
2. Khanna, S., et al. *CookingINWild: Unleashing the Challenges of Indian Cuisine Cooking Videos for Action Recognition*. in *Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD)*. 2024.
3. Nijhawan, R., et al., *Food classification of Indian cuisines using handcrafted features and vision transformer network.* Available at SSRN 4014907, 2022.
4. Pandey, D., et al. *Object detection in Indian food platters using transfer learning with Yolov4*. in *2022 IEEE 38th International Conference on Data Engineering Workshops (ICDEW)*. 2022. IEEE.
5. Kawano, Y. and K. Yanai. *Food image recognition with deep convolutional features*. in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. 2014.
6. Matsuda, Y., H. Hoashi, and K. Yanai. *Recognition of multiple-food images by detecting candidate regions*. in *2012 IEEE International Conference on Multimedia and Expo*. 2012. IEEE.
7. Liu, C., et al. *Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment*. in *Inclusive Smart Cities and Digital Health: 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings 14*. 2016. Springer.
8. Singla, A., L. Yuan, and T. Ebrahimi. *Food/non-food image classification and food categorization using pre-trained googlenet model*. in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*. 2016.
9. Bolanos, M. and P. Radeva. *Simultaneous food localization and recognition*. in *2016 23rd International conference on pattern recognition (ICPR)*. 2016. IEEE.
10. Subhi, M.A. and S.M. Ali. *A deep convolutional neural network for food detection and recognition*. in *2018 IEEE-EMBS conference on biomedical engineering and sciences (IECBES)*. 2018. IEEE.
11. Özsert Yiğit, G. and B.M. Özyildirim, *Comparison of convolutional neural network models for food image classification.* Journal of Information and Telecommunication, 2018. **2**(3): p. 347-357.
12. Zhang, M.M., *Identifying the cuisine of a plate of food.* Univ. of California at San Diego, La Jolla, CA, USA, Tech. Rep. CSE, 2011. **190**.
13. Mandal, B., N.B. Puhan, and A. Verma, *Deep convolutional generative adversarial network-based food recognition using partially labeled data.* IEEE Sensors Letters, 2018. **3**(2): p. 1-4.
14. Usman, M.Q.M. and A.A.S.A. Boyaci, *An Automatic Food Recognition System for Middle-Eastern Cuisines.* Diabetes. **101**: p. 125.
15. VijayaKumari, G., P. Vutkur, and P. Vishwanath, *Food classification using transfer learning technique.* Global Transitions Proceedings, 2022. **3**(1): p. 225-229.