# Indian Food Classification


Simran Singh
*21D070073*
*IIT Bombay*
Mumbai,India
21D070073@iitb.ac.in

Mrunal Chourey
*21D070045*
*IIT Bombay*
Mumbai,India
21D070045@iitb.ac.in

Nikita Jangid
*21D070032*
*IIT Bombay*
Mumbai,India
21D070032@iitb.ac.in



***Abstract*—In recent years, machine learning has transformed various domains and extended its use in all the domains including culinary.In this study we will compare three models- Random Forest, Convolutional Neural Network and Transfer Learning usng VGG16 model to classify Indian Food dishes.The primary objective is to evaluate and compare the performance of these models in accurately identifying Indian dishes.Random forest model known for handling complex data, reducing overfitting serves as a baseline.Custom CNN architecture leverages deep learning capabilities to enhance accuracy by capturing intricate features in food images.Incorporating transfer learning using VGG16 further refines the accuracy in classifying the images.We have done a computational analysis across these models and insights from this study contribute valuable perspectives on the suitability of different ML models for culinary image classification tasks.**


## I. Introduction

In recent years the application of machine learning has revolutionized in various domains including the culinary domain.Our study majorly focuses on the development and comparison of three distinct models for the classification of Indian dishes based on their images. This models include the Random Forest Classification, Custom Convolutional Neural Network (CNN) architecture and transfer learning through VGG16 model.The primary objective of this study is to assess the performance of different ML models in accurately identifying and categorizing Indian dishes from their visual representations. Firstly Random forests model was used which is widely used for classification and regression functions known for its ability to handle complex data, reduce overfitting, and provide reliable forecasts in different environments.The second model we used was custom CNN architecture .Leveraging the power of deep learning, this model is designed to learn intricate features and patterns present in food images, enhancing classification accuracy compared to traditional ML approaches.To further improve the accuracy and decrease the computational time we incorporated transfer learning using the VGG16 model.Transfer learning allows our CNN model to leverage pre-trained weights and architectures from VGG16, a renowned deep learning model,which enhances the ability of the model to classify the images more accurately. Throughout the report, we have compared the performance of the three models: Random Forest Classification model, our custom CNN model, and the CNN model enhanced with transfer learning. By the end of this study, we aim to provide valuable insights into the suitability of different ML models for image-based classification tasks

## II. Image Data generator

The Image Data Generator method in TensorFlow's Keras API is designed for image preprocessing, data augmentation, and batch generation.It simplifies the process of loading and augmenting image data, making it suitable for training machine learning models.This method is used to load images from the directories and then perform various data augmentation techniques such as scaling, rotation, resizing, flipping and splitting the data.We have normalised the pixel values to the range [0,255] to ensure numerical stability and convergence.We have reshaped the image data array to 2D arrays which is suitable for training machine learning models.

## III. Models

### A. Random Forest Classifier

Random Forest is a machine learning algorithm that uses multiple decision trees to achieve precise results in classification and regression tasks. It resembles the process of choosing the best path amidst multiple options.When combined the Random Forest Classifier with the ImageDataGenerator method from TensorFlow's Keras API, it becomes a powerful tool for image classification tasks, offering robustness and efficiency in handling image data.Random Forest is a machine learning algorithm that belongs to the ensemble learning group. It works by constructing a multitude of decision trees during the training phase. The decision of the majority of trees is chosen by the random forest algorithm as the final decision. In the case of regression, it takes the average of the output of different trees, and in the case of classification, it takes the mode of different tree outputs. Utilizing the Random Forest Classifier with the ImageDataGenerator method offers a reliable and efficient approach to image classification tasks. It leverages the strengths of both algorithms, resulting in accurate and robust models capable of handling various image datasets. This methodology is widely used in applications such as object recognition, medical imaging, and satellite imagery analysis.We have trained the data using Random Forest classifier and we observed that the computational time for random forest classifier is pretty high.This high complexity time includes major factors involving combining multiple trees to make predictions(the hyperparameter nestimators indicate the number of trees in the ensemble and more the value of this hyperparamter more is the computational time involved).Additionally, Random Forest's feature selection process involves evaluating a random subset of features at each split in each tree, adding to the computational load.

### *B. Convolutional Neural Network*

Convolutional Neural Networks (CNNs) are a powerful tool for machine learning, especially in tasks related to computer vision.CNNs are a specialized class of neural networks designed to effectively process grid-like data, such as images.It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The architecture of CNNs is inspired by the visual processing in the human brain, and they are well-suited for capturing hierarchical patterns and spatial dependencies within images.CNNs can achieve state-of-the-art accuracy on a variety of image recognition tasks, such as image classification, object detection, and image segmentation.But on the other hand CNNs can be complex and difficult to train, especially for large datasets and it can require a lot of computational resources to train and deploy.Fig. 1 showing steps involved in CNN .Accuracy and loss over epochs using this model is shown in Fig. 2 and Fig. 3 respectively.

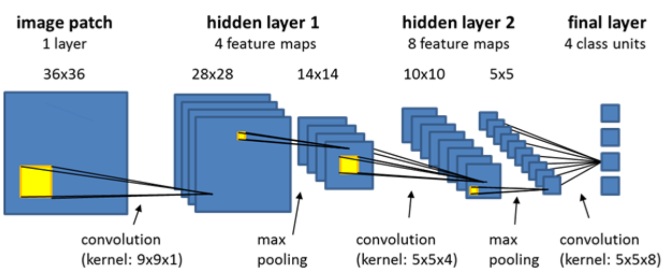


Fig. 1. Typical CNN model [1]

#### *1) Steps involved in CNN:*

- Multiplying two matrices and yielding a third, smaller matrix.
- The Network takes an input image and uses a filter (or kernel) to create a feature map describing the image.
- In the convolution operation, we take a filter (usually 2x2 or 3x3 or 5x5 matrix ) and slide it over the image matrix. The corresponding numbers in both matrices are multiplied and added to yield a single number describing that input space. This process is repeated all over the image.
- We use different filters to pass over our inputs and take all the feature maps, put them together as the final output of the convolutional layer.
- We then pass the output of this layer through a non-linear activation function. The most commonly used one is ReLU.
- The next step of our process involves further reducing the dimensionality of the data which will lower the computation power required for training this model. This is achieved by using a Pooling Layer. The most commonly used one is max pooling which takes the maximum value in the window created by a filter. This significantly reduces the training time and preserves significant information.

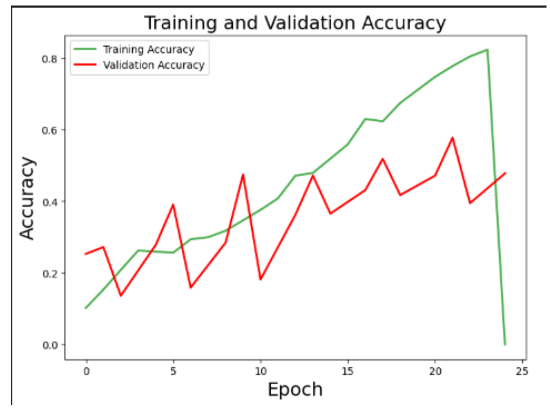


Fig. 2. Accuracy over epochs for CNN


Fig. 3. Loss over epochs for CNN

### *C. Transfer Learning*

Transfer learning is a popular deep learning method that follows the approach of using the knowledge that was learned in some task and applying it to solve the problem of the related target task. So, instead of creating a neural network from scratch we "transfer" the learned features which are basically the "weights" of the network. To implement the concept of transfer learning, we make use of "pre-trained models".Pre-trained models are the deep learning models which are trained on very large datasets which are developed and are made available by other developers who want to contribute to this machine learning community to solve similar types of problems. It contains the biases and weights of the neural network representing the features of the dataset it was trained on. The features learned are always transferrable.The basic necessity of transfer learning is Low-level features from model A (task A) should be helpful for learning model B (task B).We have used VGG16 model to train the images.Fig. 4

shows the architecture of the VGG16 model. We have used

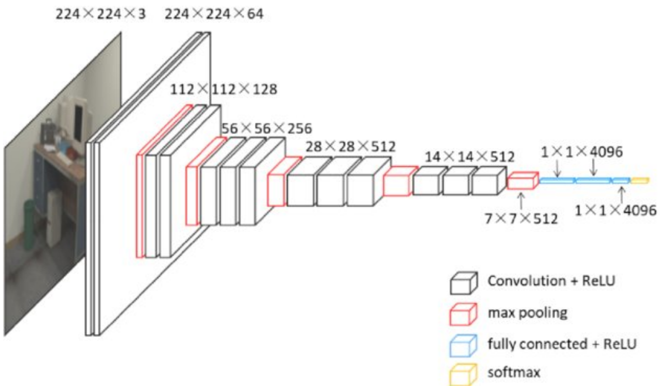


Fig. 4. Architecture of VGG16 [1]

two Dense layers with ReLU activationfor feature extraction and transformation along with a final dense layer with a softmax activation function for multi-class classification (15 classes in this case).Then combined these additional layers with the VGG16 base model.VGG16 comes pre-trained on large datasets like ImageNet which allows it to capture a wide range of features. This can significantly boost performance on tasks with limited data compared to training a custom CNN from scratch.Also it has a deeper architecture with 16 layers, capturing more intricate features in images compared to simpler custom CNN architectures.Although due to deeper architecture, VGG16 requires more computational resources for training compared to simpler custom CNN models. This can lead to longer training times and higher computational costs.Fig. 5 and Fig. 6 shows the accuracy and loss over the epochs respectively for the VGG16 model.

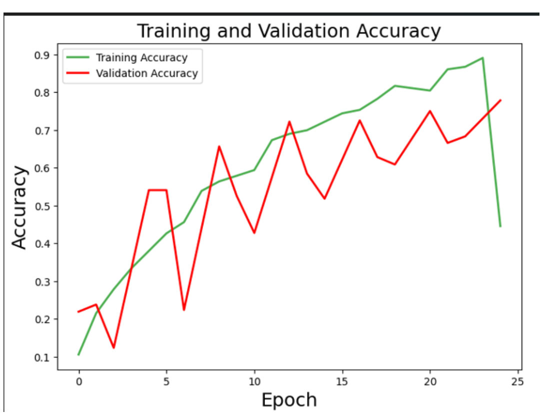


Fig. 5. Accuracy over epochs for VGG16

## IV. Conclusion

After a detailed comparison between the three models-Random Forest, Custom CNN and Transfer Learning for classifying images of Indian food dishes, VGG16 with transfer learning emerges as the most suitable model.Its ability to leverage pre-existing knowledge, capture intricate visual features, and demonstrate robust performance outweighs the limitations of traditional machine learning models like Random Forest and simpler custom CNN architectures.However this may require more extensive training computational compared to simpler models.


Fig. 6. Loss over epochs for VGG16

## References


[1] Chen Y, Chen R, Liu M, Xiao A, Wu D, Zhao S. Indoor Visual Positioning Aided by CNN-Based Image Retrieval: Training-Free, 3D Modeling-Free. Sensors (Basel). 2018 Aug 16;18(8):2692. doi: 10.3390/s18082692. PMID: 30115845; PMCID: PMC6111796.