Food Ingredient Detection Using Deep Learning

Akanksha Mane[1], Gayatri Sharma[2], Aditya Soraganvi[3], Anita Devkar[4], Roshani Raut[5]

[1] Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India
amane07022003@gmail.com
gayatri7588sharmaa@gmail.com
adityasoraganvi32@gmail.com
anitadevkar@gmail.com
roshani.raut@pccoepune.org

Abstract. Ingredients play a very crucial role in any dish, as they determine the flavor, taste, and texture of the dish, and they can impact the nutritional value of the dish. as if a dish includes many high-quality ingredients and protein sources that can provide necessary nutrients, vitamins, and minerals for a healthy diet. It's very crucial to know which ingredients to use in a dish to get the desired outcome. The strategy we've taken in this paper can help resolve this issue. A model for identifying food ingredients was put into practise based on image datasets using convolutional neural networks (CNNs).This model involves CNNs to recognise the items included in training images of Indian sweet dishes. In order to expand the amount and diversity of the dataset, the model calls for gathering a dataset of 1000 images of Indian sweets dishes, preprocessing the images, and using data augmentation techniques. We then trained a Convolutional Neural Network (CNN) model using a modified version of the Inceptionv3 architecture, fine-tuned on the dataset using transfer learning. The results demonstrated a promising accuracy rate of 99.6% in ingredient detection, which is noteworthy given the dataset's size of 1000 images. Consumers with dietary choices , such vegetarians or those who have food allergies, can benefit from ingredient detection utilizing CNNs. It is possible to provide healthy eating decisions by properly identifying the contents in a food item.

Keywords: Inceptionv3, Object recognition, Food Ingredients detection, LSTM, SVM, CNN.

1 Introduction

The food industry is constantly evolving, and technological advancements are revolutionizing the way we approach food and cooking. Food recognition has drawn considerable research interest due to its importance for applications in

R. Murugan et al. (eds.), *Proceedings of the International Conference on Signal Processing and Computer Vision (SIPCOV 2023)*, Advances in Engineering Research 239,
https://doi.org/10.2991/978-94-6463-529-4_20

the field of health[1]. Machine learning has become a powerful resource for food analysis in latest days, with applications ranging from food quality inspection to recipe recommendation systems. One interesting area of research within this field is ingredient detection in food, which can be used to help people with food allergies, dietary restrictions, or who are simply curious about what's in their food.

We concentrate on the issue of ingredient detection in this research in various sweet dishes using Machine Learning and Convolutional Neural Network (CNN) models. Sweet dishes, such as cakes, pastries, and desserts, are popular around the world and often contain a complex mix of ingredients, making them an ideal subject for this type of analysis. Our goal is to develop a model that can accurately identify the ingredients present in sweet dishes using images as input.

To address these issues, technology had to be improved, and so object recognition technology was made available to everyone. The technology for object recognition has advanced significantly. A system known as object recognition recognises the things in an image based on their unique characteristics. Hence, object recognition is a method that extracts elements from a picture and locates them in the real world. For the purposes of this research, CNN is used in conjunction with the Deep Learning model, OpenCV, Tensorflow, Keras, Pandas, and Numpy[2].

The processing and representation of images can be done using OpenCV. Machine learning can be performed numerically because of deep learning APIs like NumPy.

To find the image and develop deep learning image recognition models, utilize the TensorFlow API. The features in the food photographs have been extracted using both conventional machine learning methods like Histogram of Gradient (HOG) and Scale-Invariant Feature Transform (SIFT) and cutting-edge deep learning models like VGG16, MobileNet, ANN, Resnet18, Resnet50, and Densenet121[3]. Using pre-trained architecture has the benefit of allowing for transfer learning-based modification. These packages, APIs, and architectural frameworks make it simple to identify objects in a picture[4].

To achieve this goal, we will first review the relevant literature on ingredient detection and image analysis using Machine Learning. We will then describe our methodology for collecting and preprocessing the data, building and training the CNN models, and evaluating their performance in comparison to conventional methods using manually derived features, CNN greatly improved recognition performance [5]. Finally, we will present and discuss our results, highlighting the strengths and limitations of our approach and identifying areas for future research.

Overall, this research has the potential to contribute to the development of more accurate and efficient food analysis techniques, with applications in areas such as food safety, quality control, and personalized nutrition.

The literature survey is offered in section 2, followed by descriptions of the suggested approach in sections 3 , results in sections 4, and a conclusion in section 5 followed by future scope in section 6.

2 Literature Survey

Indian sweets are a diverse and popular category of desserts that often contain a complex mix of ingredients.Recently, there has been an increase in interest in using CNN and machine learning models for ingredient detection in Indian sweets[5]. Several studies have explored this area of research, with a focus on different types of Indian sweets and different approaches to data collection and model development.

There is a growing interest in designing systems that can properly identify the ingredients present inside a food item and offer acceptable recipes based on the discovered elements as technology and the availability of food-related data increase.

One early study by Hongsheng He et al. (2015) [6] proposed a method by combining a texture verification model and a deformable part-based model that allows for the detection of food items. A multiview multi kernel SVM is used to classify food groups based on the observed ingredients. They used a database with 15,262 food photos representing 55 different food categories.

Marlies De Clercq et al. (2016) [7] proposed a method with Non-negative matrix factorization (NMF) and two-step regularized least squares (TSRLS) are both approaches for creating models of recipe completion. Real-world data is applied to evaluate the outcomes.

Another study by Suyesh Maheshwari et al. (2019) [8] proposed a method by using two machine learning models-vector space model and Word2Vec model to search top ingredient pairs from different cuisines and to suggest alternate ingredients. They also used web scraping to collect images from the web, which can then be labeled and used to train image recognition models.

One early study by Md. Kishor Morol et al. (2021) [9] proposed a method for food ingredient recognition and recipe recommendation algorithm using a

CNN model. They used a dataset of 9,856 images of food and had 32 different classes and achieved an accuracy of 94%. They find CNN is suitable for food ingredient image recognition.

Ziyi Zhu et al. (2021) [10] offered two methods for identifying ingredients. The first kind of method combines salient ingredient classifiers with salient ingredient identifiers. One element of the Type 2 technique is the segment-based classifier in which each classifier and identifier on Resnet50 are trained through transfer learning. They obtain 91.97% in the test set and 82.48% on the dataset of test dish photos.

These studies demonstrate the potential of Machine Learning and deep learning models for ingredient detection in Indian sweets such as by using a combination of CNN and LSTM models. SVMs can be used for classification tasks, such as identifying the cuisine type of a recipe or detecting specific ingredients. Hybrid approaches combine multiple techniques, such as collaborative filtering and content-based filtering, to improve the accuracy and effectiveness of food recommendation systems.

3 Methodology

3.1 Description of dataset

As real-world datasets for this topic did not contain images of various Indian sweet dishes like ras-malai, ghevar, etc. online, we had to generate our own. We combined the data we generated for this study with the Indian food 101 dataset from Kaggle. There are around 255 conventional and well-known Indian dishes in the Indian Food 101 collection. We gained insight regarding ingredients, particularly those used in sweet meals, from there. The data we generated consists of photos of 10 classes (distinct sweet dishes), with 100 photographs per class (256,256). We created a dataset with about 1000 photos in total.

Table 1. Table captions should be placed above the tables

Dishes	Dishes Name	Ingredients
	Balushahi	Maida flour, yogurt, oil, sugar
	Gajar ka Halwa	Carrots, milk, sugar, ghee, cashews, raisins

	Ghevar	Flour, ghee, kewra, milk, clarified butter, sugar, almonds, pistachio, saffron, green cardamom
	Gulab Jamun	Milk powder, plain flour, baking powder, ghee, milk, sugar, water, rose water
	Jalebi	Chhena, sugar, ghee
	Kalakand	Milk, cottage cheese, sugar
	Kheer	Milk, rice, sugar, dried fruits
	Laddu	Gram flour, ghee, sugar
	Mysore pak	Besan flour, semolina, mung bean, jaggery, coconut, skimmed sugar, ghee milk powder
	Ras Malai	Chhena, reduced milk pistachio

3.2 Proposed work

As shown in Fig. 1, the proposed work contains the following components -

Data collection: Images of various Indian sweet dishes are gathered in this stage from various sources and are then saved in a dataset.

Data Preprocessing: After the images have been resized to a fixed resolution and the pixel values have been normalized, To create training and test datasets, the dataset was split.

Data augmentation: The training images are improved through using various transformations, such as rotation, shifting, flipping, zooming, and shearing, to expand the training set and enhance the model's generalization performance[11].

Model Architecture: To learn the features and categorize the pictures into various sweets dishes, a convolutional neural network (CNN) model is created with numerous convolutional, pooling, dropout, and fully connected layers.

Model training: To minimize the loss function and maximize precision on the validation set, the CNN model is trained on the augmented training data using backpropagation and stochastic gradient descent.

Model Evaluation: The trained model's accuracy, precision, memory, and F1-score for each class are assessed on the test set.

Model Fine-Tuning: To further enhance the model's performance, the hyperparameters learning rate, group size, optimizer, activation functions, and regularization methods are adjusted.

The flowchart shows the sequential order of the steps involved in the ingredient detection project and helps to visualize the overall process from data collection to model deployment.

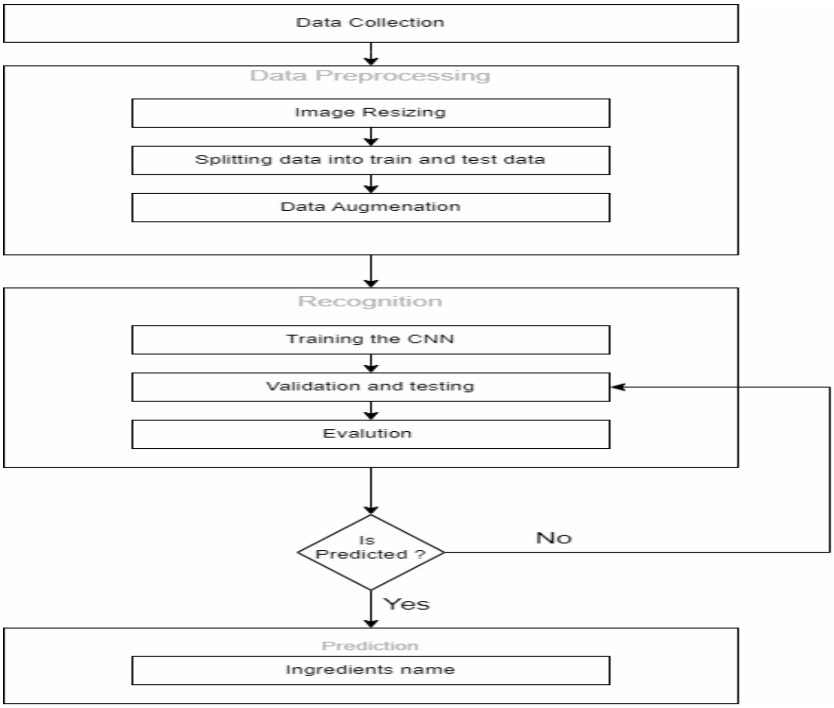

Fig. 1. Flow of the proposed strategy

3.3 Architecture

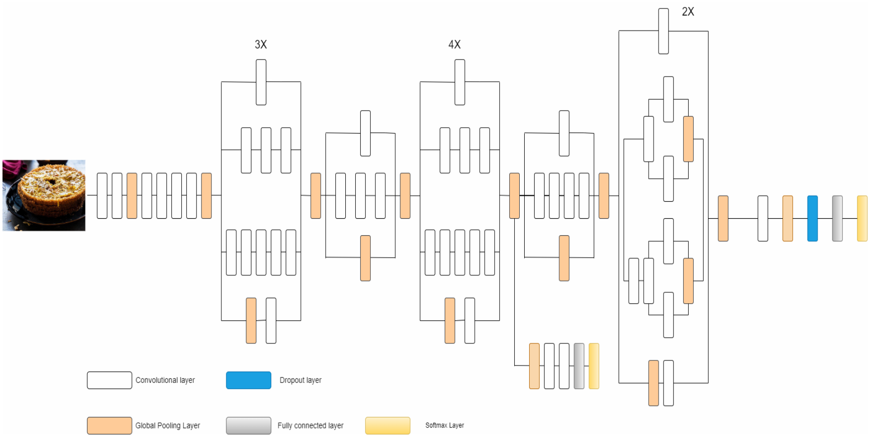

Fig. 2. Architecture of InceptionV3

In our research paper, as shown in Fig. 2 we employed a modified version of the Inceptionv3 architecture, which is a deep convolutional neural network initially introduced for the ImageNet Large Scale Visual Recognition Challenge. The modified Inceptionv3 architecture was specifically tailored for detecting ingredients in food using CNN.

The key components of our modified Inceptionv3 architecture are:
Input Layer: This layer takes a 3-dimensional array of the food ingredient image, which includes the height, width, and color channels.

Inceptionv3 Blocks: The model is composed of several Inceptionv3 blocks, with each block having multiple parallel convolutional branches that utilize distinct filter sizes and pooling operations. The outputs of these branches are combined and fed into the next block.

Global Average Pooling: The average value of each feature map is determined using this layer at the end of the last Inceptionv3 block.
Dropout: We used a dropout layer to prevent overfitting by randomly discarding some of the features during training.

Fully Connected Layers: The dropout layer output is flattened and passed through multiple fully connected layers, which employ a set of learnable weights to predict the presence of each ingredient in the input image.

Output Layer: A probability distribution across the potential ingredient classes is output by the model's softmax layer, which is the last layer to be considered.

Our model was fine-tuned on a dataset of 1000 Indian sweet dish images using transfer learning and achieved a high accuracy of 99.6% in detecting ingredients. These results showcase the effectiveness of our modified Inceptionv3 architecture for ingredient detection using CNN and have significant implications for applications in the food industry, including ingredient labeling and dietary analysis.

3.4 Data Augmentation

By creating additional training examples from existing ones, data augmentation is a technique used in machine learning and computer vision to enhance the size of a dataset. By introducing noise, biases, or other changes

to the raw data, the goal is to increment diversity in the data and decrease overfitting.
The goal of data augmentation is to add variation to the data so that the model can become more adaptable to various changes in the same data. For instance, frequent enhancements in picture categorization include rotating an image horizontally or vertically, rotating it, zooming in or out, adding noise, or altering the brightness and contrast. We have added roughly 2000 additional photos to our dataset by using these changes, which increases data diversity and decreases overfitting.

4 Results

4.1 Model performance

We have a dataset with a total of 1000 photos of dishes, as we previously said. Therefore, we used 80% of the photos for training and 20% for testing. Our CNN-based model had 99% accuracy for the training dataset after running 200 transfer learning epochs. In Fig. 3 there are many fluctuations in the accuracy of testing dataset and is much stable in training dataset.

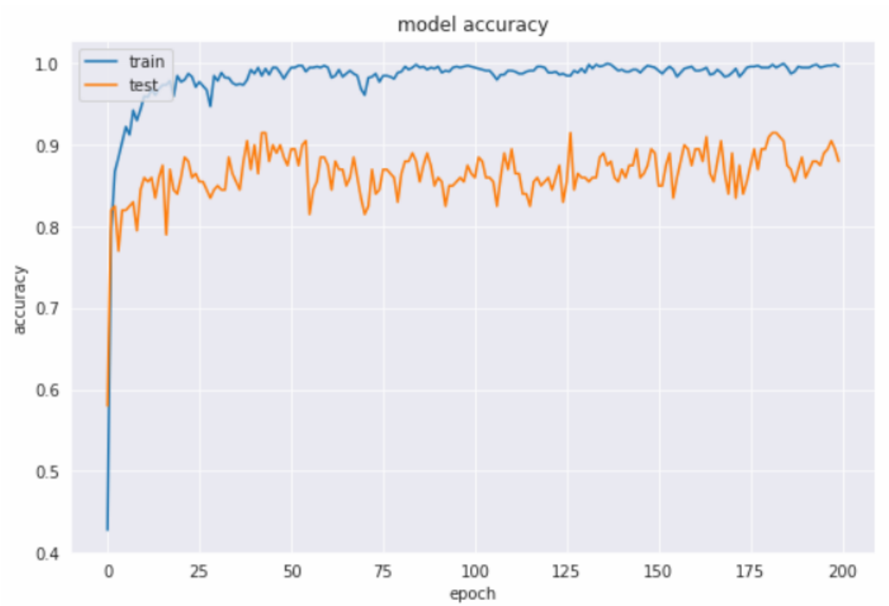

Fig. 3. Training v/s Validation accuracy

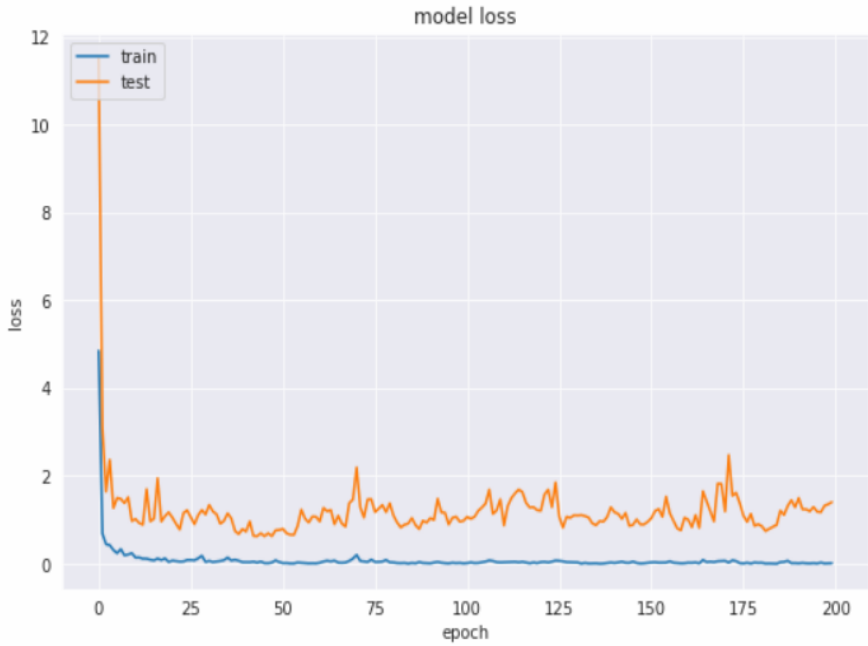

Fig. 4. Training v/s validation loss

In Fig .4 there are many fluctuations in the loss of testing dataset and is much stable in the loss of training dataset. After analyzing the results and metrics, we can state that the CNN model has admissible performance on identifying ingredients. Accuracy, Precision, F1 score, training accuracy and validation accuracy, loss matrix, and confusion matrix are a few of the metrics that may be used to assess the effectiveness of an ingredient identification model employing CNNs.

In Fig .5 Images from class 1 are predicted accurately i.e accuracy is 100% whereas images in class 5 have less accuracy(i.e 75%).

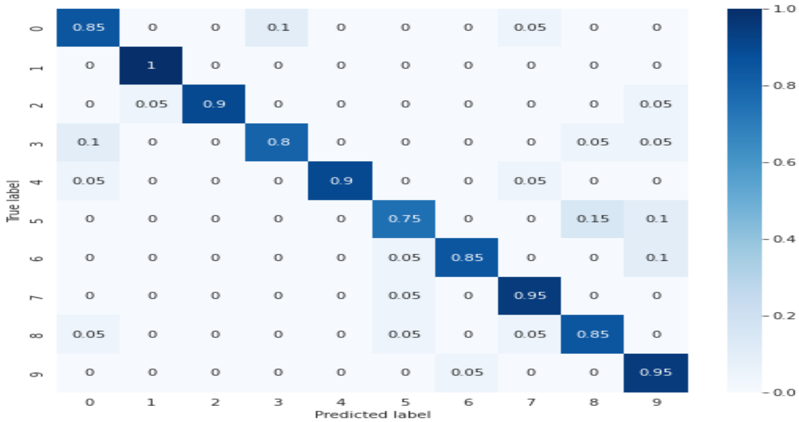

Fig. 5. Confusion Matrix

5 Conclusion

This project successfully developed a deep learning model for identifying ingredients in Indian sweet dishes using convolutional neural networks. The model achieved an impressive accuracy of 99.6%, indicating its potential applications in the food industry and dietary analysis. Transfer learning and data augmentation techniques were crucial in enhancing the model's performance, and the use of a modified Inceptionv3 architecture proved to be effective in balancing accuracy and model efficiency.

Future research could expand the model to recognize ingredients in other types of dishes and cuisines, as well as explore its utility in recipe recommendation systems and dietary analysis tools. This project highlights the potential of deep learning techniques in food ingredient recognition and related applications, providing a promising avenue for further investigation.

6 Future scope

The model implemented has a number of potential future directions for further research and development. Firstly, the model could be extended to detect ingredients in other types of cuisine, which would require building a new dataset of images and annotations. Secondly, the model could be optimized for real-time ingredient detection, which would require further architectural optimization and parallelization techniques to improve processing time. Thirdly, the model could be integrated into a recipe recommendation system, which would require the development of a database of recipes and their corresponding ingredient lists. Finally, the model could be extended to detect allergens in food, which would be useful for people with food allergies or intolerances. Overall, this project has the potential to be highly impactful in a range of fields, and further research is needed to explore these potential directions.

References

1. J. Chen, B. Zhu, C. -W. Ngo, T. -S. Chua and Y. -G. Jiang, "A Study of Multi-Task and Region-Wise Deep Learning for Food Ingredient Recognition," in IEEE Transactions on Image Processing, vol. 30, pp. 1514-1526, 2021, doi: 10.1109/TIP.2020.3045639.
2. Keshri, A., Singh, A., Kumar, B., Pratap, D. and Chauhan, A., 2022. Automatic detection and classification of human emotion in real-time scenario. Journal of IoT in Social, Mobile, Analytics, and Cloud, 4(1), pp.41-53.
3. Q. -L. Tran, G. -H. Lam, Q. -N. Le, T. -H. Tran and T. -H. Do, "A Comparison of Several Approaches for Image Recognition used in Food Recommendation System," 2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), Purwokerto, Indonesia, 2021, pp. 284-289, doi: 10.1109/COMNETSAT53002.2021.9530793.
4. G. Zheng, G. Han and N. Q. Soomro, "An inception module CNN classifiers fusion method on pulmonary nodule diagnosis by signs," in Tsinghua Science and Technology, vol. 25, no. 3, pp. 368-383, June 2020, doi: 10.26599/TST.2019.9010010.
5. M. A. Subhi and S. Md. Ali, "A Deep Convolutional Neural Network for Food Detection and Recognition," 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia, 2018, pp. 284-287, doi: 10.1109/IECBES.2018.8626720.
6. He, H., Kong, F. and Tan, J., 2015. DietCam: multiview food recognition using a multi kernel SVM. IEEE journal of biomedical and health informatics, 20(3), pp.848-855.
7. De Clercq, M., Stock, M., De Baets, B. and Waegeman, W., 2016. Data-driven recipe completion using machine learning methods. Trends in Food Science & Technology, 49, pp.1-13.
8. Maheshwari, S. and Chourey, M., 2019. Recipe recommendation system using machine learning models. International Research Journal of Engineering and Technology (IRJET), 6(9), pp.366-369.
9. Morol, M.K., Rokon, M.S.J., Hasan, I.B., Saif, A.M., Khan, R.H. and Das, S.S., 2022, March. Food recipe recommendation based on ingredients detection

using deep learning. In Proceedings of the 2nd International Conference on Computing Advancements (pp. 191-198).
10. Zhu, Z. and Dai, Y., 2021. Food Ingredients Identification from Dish Images by Deep Learning. Journal of Computer and Communications, 9(4), pp.85-101.
11. Shorten, Connor, and Taghi M. Khoshgoftaar. "A survey on image data augmentation for deep learning." Journal of big data 6.1 (2019): 1-48.