Food Recipe Recommendation System

1st Martina D'souza, 2nd Mitesh Rege, 3rd Shivam Mishra, 4th Prajna Shetty
[1,2,3,4]*Department of Information TechnologyXavier Institute of Engineering* Mumbai, Maharashtra

***Abstract*—Everyone appreciates food in this world. The most important thing for humans to survive is food. Yet, not every person can consume food or consume the components used to prepare it. It's crucial to understand the food's ingredientsand cooking process. This food recipe recommendation aids in understanding the ingredients and preparation techniques forthe dish. This makes it easier for folks to understand a recipe's ingredients and preparation time. The three transfer learning models, Xception, InceptionResNetV2, and InceptionV3, were fine-tuned and implemented for image classification. Indian FoodImage is the dataset that was used in this. A training accuracyof 99.58% and testing accuracy of 88.9% was attained.**

***Index Terms*—Transfer learning, Image Classification, InceptionV3**

I.INTRODUCTION

Food image recognition can help humans in several ways:

- Dietary assessment: Food image recognition can be used to accurately assess a person's dietary intake, which can be helpful for individuals who are trying to monitor their food intake for health or weight management purposes. By taking pictures of their meals and using a food image recognition app, individuals can get a better understanding of the nutritional content of their meals.
- Food logging and tracking: Food image recognition can also be used to log and track the foods that a personeats. This can be particularly helpful for individuals who are following a specific diet or trying to monitor their food intake for other reasons. By using a food image recognition app to log their meals, individuals can get a more accurate picture of their food intake over time.
- Restaurant menu analysis: Food image recognition canbe used by restaurant owners to analyze their menus and optimize their offerings. By analyzing the popularity of different dishes and identifying areas where they can im- prove their offerings, restaurant owners can provide better service to their customers and increase their revenue.
- Cultural exploration: Food image recognition can also help individuals explore different cultures and cuisines. By taking pictures of unfamiliar foods and using a food image recognition app, individuals can learn more about the ingredients and flavors of different dishes and expand their culinary horizons.

II.DEEP LEARNING

Deep learning is a subset of machine learning that uses artificial neural networks to learn from data. In image recognition, deep learning algorithms, such as convolutional neural networks (CNNs), are used to analyze the features of images and identify patterns that can be used to classify the images into different categories.

Deep learning algorithms work by breaking down the images into smaller parts and analyzing each part separately. This process is similar to how the human visual system works, with the brain processing individual features of an image, such as edges, shapes, and colors, to form a complete perception ofthe image.

CNNs are particularly effective at image recognition because they can analyze images at multiple levels of abstraction. The first layers of the network analyze basic features, such as edgesand textures, while the later layers analyze more complex features, such as shapes and patterns. This hierarchical approach allows the network to learn complex representations of the images and make accurate classifications.

During training, the deep learning algorithm is presented with a large dataset of labeled images. The algorithm uses these images to learn the features and patterns that are associated with each category. Once the algorithm has been trained, it can be used to classify new images that it has not seen before. Overall, deep learning is a powerful tool for image recognitionbecause it can learn from large amounts of data and can identify complex

patterns that might be difficult for humans to detect.

Transfer learning is mainly used for natural language process- ing tasks such as computer vision and emotion analysis due to the large amount of computing power required. Transfer learning is actually a machine learning technology. No, but wecan think of it as. For example, design methodology in areas such as active learning. It is not the exclusive part or research area of machine learning. Nevertheless, it is very popular in combination with neural networks that require large amounts of data and processing power [6] [7]

III.LITERATURE REVIEW

1) Salvador, Amaia, et al." Learning cross-modal embeddings for cooking recipes and food images." Proceedings of the IEEE conference on computer visionand pattern recognition. 2017. [1]

In This Paper it introduced Recipe1M, a new, massive, structured corpus comprising over 800,000 food images and 1,000,000 recipes for cooking. Recipe1M, the biggest publicly accessible collection of recipe data, enables the training of high-capacity models on multi- modal, aligned data. These data are used to train a neural network that successfully completes the image- recipe retrieval task using a joint embedding of recipes and photos. Furthermore, we show how regularisation, when combined with a high-level classification target, enhances retrieval performance to match that of humans while also enabling semantic vector arithmetic. These embeddings, in our hypothesis, will serve as a starting point for further investigation of the Recipe1M dataset and of food and cooking in general. Models, data, and code are all available to the public.

2) Shen, Zhidong, et al. "Machine learning based approach on food recognition and nutrition estimation." Procedia Computer Science 174 (2020): 448-453. [2]

In this paper, they introduce a novel machine learning based system that accurately classifies food photos automatically and predicts food qualities. In the training stage of the prototype system, this study suggests a deep learning model made up of a convolutional neural network that categorises food into distinct groups.The major goal of the suggested strategy is to increase the pre-trainingmodel's accuracy. In the study, a client- server prototype system is designed.A photo detection request is sent by the client and handled by the server.A pre-trained CNN is one of the three primary software components of the prototype system.

3) Subhi, Mohammed Ahmed, Sawal Hamid Ali, and Mohammed Abulameer Mohammed. "Vision-based approaches for automatic food recognition and dietary assessment: A survey." IEEE Access 7 (2019): 35370- 35381. [3]

The focus of many dieticians and medical organisations is on eating the proper quantity and type of food. Maintaining a nutritious diet is essential to avoiding obesity and other health-related problems, such as diabetes, stroke, and many cardiovascular illnesses, in addition to physical activity and exercises. Automatic or semi-automatic nutritional assessment solutions have been made available by recent developmentsin machine learning applications and technologies, which may be a more practical method to monitor daily food intake and control eating patterns. These methods are meant to address the issues with the conventional dietary monitoring systems, which have a lack of accuracy, underreporting, time requirements, and low adherence. This study extensively explores contemporary vision-based approaches and ways to specify these approaches and methodologies used for automatic dietary assessment, their performances, feasibility, and unaddressed challenges and issue

4) Kawano, Yoshiyuki, and Keiji Yanai. "Foodcam: A real-time food recognition system on a smartphone."Multimedia Tools and Applications 74 (2015): 5263- 5287. [4]

This described a method for a smartphone-based real-time food recognition system. There have been two different real-time image recognition techniques employed. One is the fusion of X2 kernel feature maps with bag-of-features (BoF) and colour histograms.The second one is a colour patch descriptor with the most recent Fisher Vector representation, together with the HOG patch descriptor. A classifier has been implemented using linear SVM. They have attained a classification accuracy rate of 79.2%.

5) Bolaños, Marc, Marc Valdivia, and Petia Radeva. "Where and what am i eating? image-based food

menu recognition." Computer Vision–ECCV 2018 Workshops:Munich, Germany, September 8-14, 2018, Proceedings, Part VI 15. Springer International Publishing, 2019. [5] For food image recognition, use this CNN. Their solu- tion makes use of two different inputs. The first gives a basic explanation of the food image. The first solution use the InceptionResNetV2 CNN's penultimate layer, while the second makes use of LogMeals API to providea high-level description of the food image. This Log- Meals API, which forecasts meal groups, ingredients, and dishes, provided three different CNNs

IV.PROPOSED METHODOLOGY

The Proposed System Consists of various phases that are:

A. Dataset:

The "Indian Food Images (Top 20)" dataset on Kagglehas images of 20 popular Indian dishes from different states, pre-divided into training and validation sets, each food item category contains images of that particular dish. The dataset comprises 5828 images distributed across 20 categories.

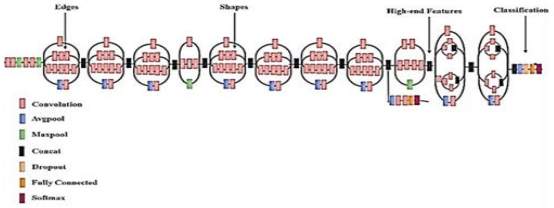

Fig. 1. Proposed Methodolgy

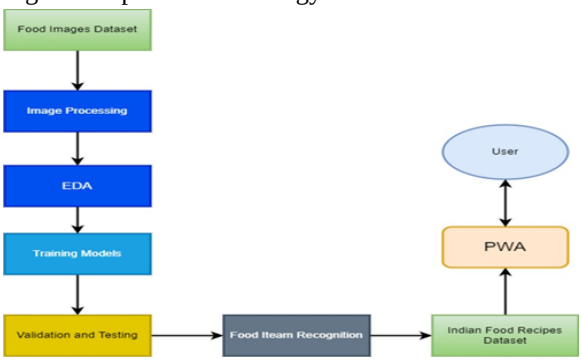

Fig. 3. Inception V3 Model

minimize variance and reduce complexity. Concatenation combines input tensors, dropout reduces over-fitting, and the softmax function produces the final output.

B. CNN Structure :

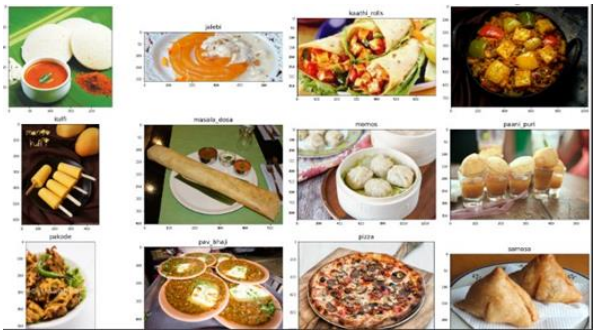

Fig. 2. Indian food dataset

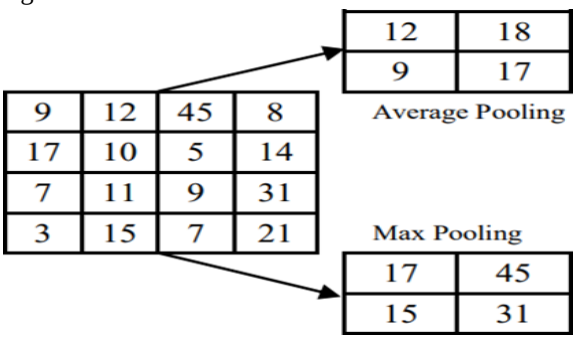

Fig. 4. Operation of Average pooling and Max pooling

A Convolutional Neural Network (CNN) is a type of artificial neural network that can perform feature extraction and classification on multiple raw images. The Inception-V3 model is a commonly used CNN architecture, which can be retrained with custom datasets. It consists of layers such as AvgPool, MaxPool, Convolution, Concat Layer, Fully Connected layer, Dropout, and Softmax Function. The convolution layer uses weight sharing and sparse connectivity to calculate neuron outcomes. The last layer reduces the entire image to a vector for class scores. A fully connected layer takes the Inception-V3 feature outputs and custom-generated segmented features for classification. Overall, Inception-V3is a powerful CNN architecture that can be customized for various applications. [8]

The food image classification model uses a CNN for feature extraction and softmax and fully connected layers for classification. Pooling function is used to decreasefeature resolution and counter noise in obtained features from convolution. Average pooling reduces data variance and computational complexity while max pooling extracts important features of edges. Max and average pooling

C. Image Preprocessing, EDA Data Augmentation:

Image data consists of variations due to resolution

differences between scenes, pixel intensities of an image and the environment around which the image was taken. This areaof image processing is critical in today's time with the rise of Artificial intelligence. Fthe research requires tremendous amount of work and can be seen as widely growing areas of computer vision.

Sobel filter is a basic way to get an edge magnitude/gradient image this works by calculating the gradient of image intensityat each pixel within the image. It finds the direction of the largest increase from light to dark and the rate of change in that direction. These are some common techniques used to preprocess images for classification. First, it is important to check the size, resolution, and color space of the images, normalize themto a common size, and convert them to grayscale or RGB basedon the requirements of the classifier. Augmentation techniqueslike rotation, flip, and zoom can also be applied to increase the dataset size and improve model performance, exploratory data

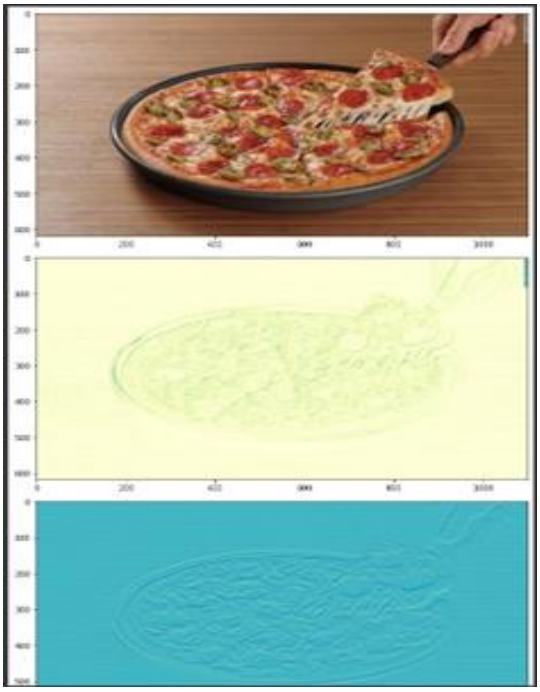

Fig. 5. Edge magnitude using Sobel filter

analysis (EDA) can be used to check the image distribution and quality, handle imbalanced classes, and visualize channel intensity and histograms between two color channels. [9] The Sobel filter can be used for edge detection, and brightness normalization can be applied to enhance image quality. [10] [11]Data augmentation techniques such as rotation, flipping, and zooming can be applied to images to increase datasetsize, improve classifier performance, reduce overfitting, and enhance model generalization capability. By utilizing these techniques, images can be properly preprocessed and preparedfor use in classification tasks. [12]

with the RMSprop optimizer and learning rate of 2e-5. The model is trained for 40 epochs with callbacks for progress logging and a scheduler learning rate. The training set is used for optimization, the validation set for hyperparameter tuning and performance monitoring, and the test set for evaluation the three transfer learning models, Xception, InceptionResNetV2, and InceptionV3, were fine-tuned and implemented for image classification. As these models have already learned powerful image representations, they are well-suited for the task. The checked pointer interface is used to save checkpoints in. hdf5 format, considering only the best score. [8]

E. Scraping Food Recipes Deploy

The Python tool Recipe-scrapers was used to extract recipe information from "Vegrecipesofindia.com", a website for veg- etarian Indian cuisine recipes. A frontend was developed usingFlask, a Python web framework, for building and deploying web applications.

V.RESULTS AND DISCUSSION

The Model InceptionV3 with pre-trained weights of image- net is included with input shape (150,150,3). In classification phase two more dense layers are added, where the first layer has 128 units, activation function Rectified linear unit (ReLu) with 0.5 dropouts, and Batch Normalization is considered. In the second layer 64 units, the activation function of ReLu with 0.5 dropouts is considered. The considered optimizer is RM- Sprop which is a momentum-based optimizer and has similar functionality of Ada-delta optimizer. three transfer learning models, Xception, InceptionResNetV2, and InceptionV3, were fine-tuned and implemented for image classification. As these models have already learned powerful image representations, they are well-suited for the task. From all three, our proposed model is by using inceptionV3 algorithm which gives us training and test accuracy as 99.59% and

88.9% resp. Below shown fig 6 , it gives us a clear explanation with graphical representation curve for model accuracy and model loss with two curves representing train accuracy and validation accuracy, as epoch increases the accuracy curve grows and loss curve lower down gradually.

D. Modal Training

	Xception	InceptionResNetV2	InceptionV3
Training Accuracy	97.91%	95.84%	99.58%
Testing Accuracy	32.14%	80.51%	88.9%

Fig. 6. Data Augmentation

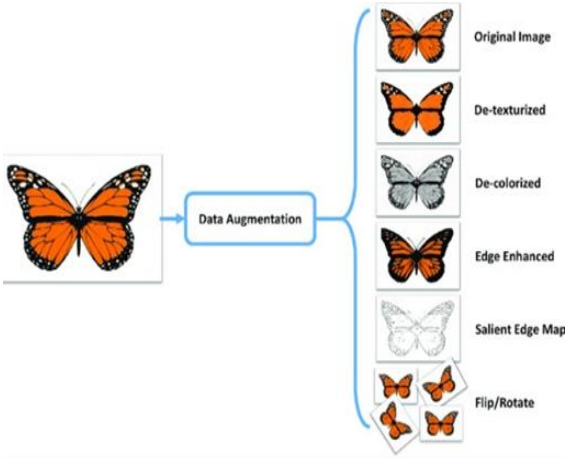

Fig. 7. Accuracy of Algorithms

At the end of training, we want the model to classify or detect objects based on features which are specific to the class.to validate how model attributes the features to class output, we can generate heat maps using gradients to findout which regions in the input images were instrumental in Here, the proposed model utilizes the custom Inception-V3 weights which are pre-trained using ImageNet [13] and it considers the reshaped size of 150×150×3 for all images. Average pooling is used on the food image dataset with a dropout rate to handle over-fitting. The softmax function is used to identify the output class, and a CNN classifier is used determining the class, below fig 9 shows the focus area where the model actually focuses in the food images and key arefrom where it classify which category this image belong to.

The input image passed to the model is shown on the left, followed by a heat map in the middle, and a class activation map on the right. The heat map shows what areas in the image

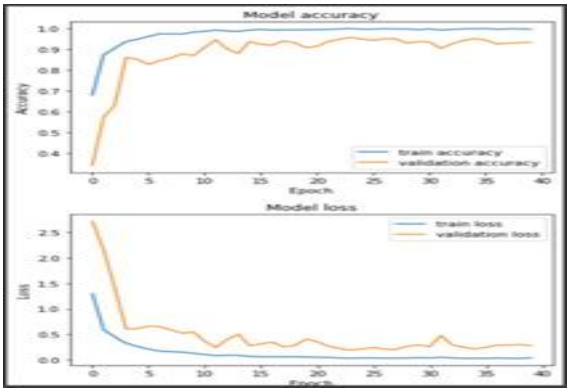

Fig. 8. Graphical representation

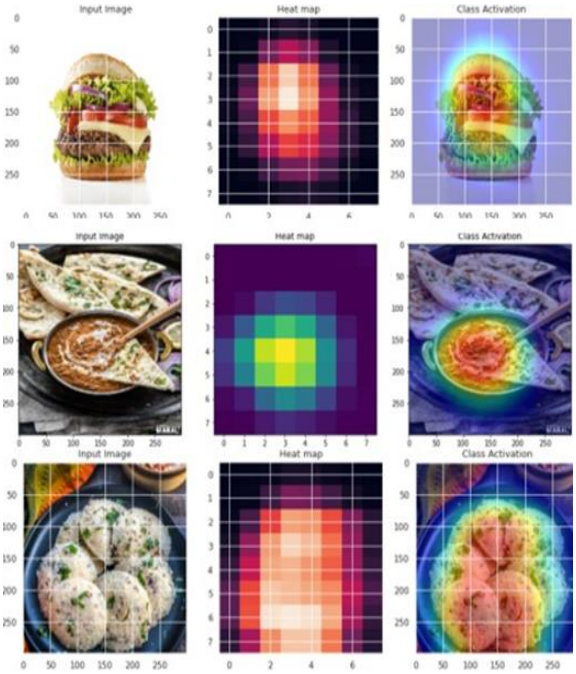

Fig. 9. Areas of Image to determine food

were used to determine the class of the image, making it clear what a model looks for in an image if it has to be classifiedas an idli or kadai panner or its burger shown in Fig 9.

In Frontend GUI there is a demo page where user will upload the food image and our proposed model will internally recognized it by classifying form all trained classes and result label of entered food as well as other crucial details like the whole recipe process to prepare it , ingredients to use , cooking time , preparation time along with other few information will be displayed to the user and all this information is being renderby web scraping using “Recipe-scrapers”, it is a tool developedin python to extract recipes from internet and we render all information extracting from “vegrecipesofindia.com” which is a good source for Indian cuisine recipes, frontend outputs shown in

below fig 10,11.

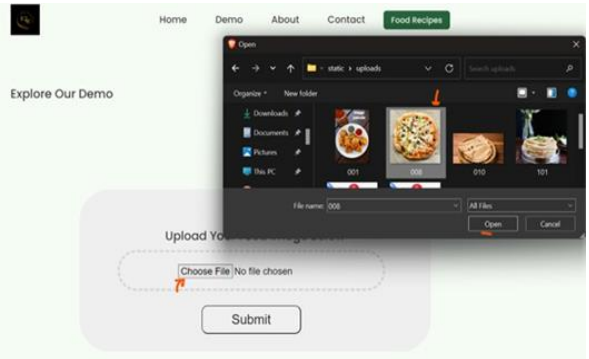

Fig. 10. Uploading Image

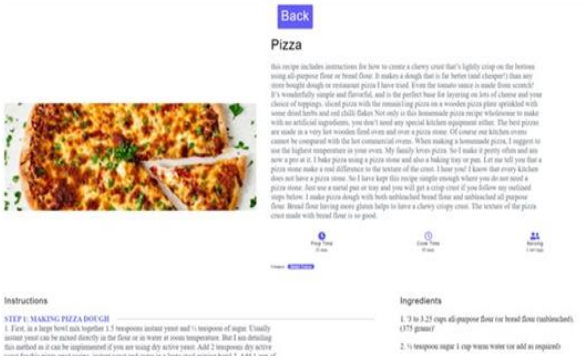

Fig. 11. Recipe Page (model recognized that its Pizza)

VI.CONCLUSION

This System introduces an image-to-recipe generation sys- tem, which takes a food image and produces a recipe consisting of a title, ingredients and sequence of cooking instructions. This model can easily be developed into applications for smartphone and tablets where potential users can customize their recipe library and easily query the recipes from their phones with any dish images they are interested at.

REFERENCES

1. Salvador, Amaia, et al. ”Learning cross-modal embeddings for cooking recipes and food images.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

2.Shen, Zhidong, et al. ”Machine learning based approach on food recog- nition and nutrition estimation.” Procedia Computer Science 174 (2020): 448-453.

3. Subhi, Mohammed Ahmed, Sawal Hamid Ali, and Mohammed Abu- lameer Mohammed. ”Vision-based approaches for automatic food recog- nition and dietary assessment: A survey.” IEEE Access 7 (2019): 35370- 35381.

4. Kawano, Yoshiyuki, and Keiji Yanai. ”Foodcam: A real-time food recog-nition system on a smartphone.” Multimedia Tools and Applications 74 (2015): 5263-5287.

5. Bolaños, Marc, Marc Valdivia, and Petia Radeva. ”Where and what am i eating? image-based food menu recognition.” Computer Vision–ECCV 2018 Workshops: Munich, Germany, September 8-14, 2018, Proceedings,Part VI 15. Springer International Publishing, 2019.

6. Wei Xi, Peng Li, X.B. Guo. “Application of Correlation Analysis Method based on multidimensional time series in prediction of equipment failure”. In the Journal Of Advances of Power System Hydroelectric Engineering, 000(012) (2014).

7. L. Breiman. “Random Forests”. In the Journal Of Machine Learning, 45(1) (2001).

8. Burkapalli, Vishwanath C., and Priyadarshini C. Patil. ”Transfer learning: inception-v3 based custom classification approach for food images.” ICTACT journal on image and video processing 11.01 (2020).

9.Morgenthaler, Stephan. ”Exploratory data analysis.” Wiley Interdisci- plinary Reviews: Computational Statistics 1.1 (2009): 33-44.

10.Aqrawi, Ahmed Adnan, and Trond Hellem Boe. ”Improved fault segmentation using a dip guided and modified 3D Sobel filter.” SEG Technical Program Expanded Abstracts 2011. Society of Exploration Geophysicists, 2011. 999-1003.

11. Sharifrazi, Danial, et al. ”Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images.” Biomedical Signal Processing and Control 68 (2021): 102622.

12. Van Dyk, David A., and Xiao-Li Meng. ”The art of data augmentation.”Journal of Computational and Graphical Statistics 10.1 (2001): 1-50.

13.Basavaraj Anami and Vishwanath C. Burkpalli, “Recognition and Clas- sification of Images of Fruits Juices Based on 3-Sigma Approach”, International Journal of Computational Vision and Robotics, Vol. 1, No. 2, pp. 206- 217, 2010