A Comparative Study of Food Image Recognition for Recipe Recommendation

Saksham Jain
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Symbiosis International (Deemed university), Pune, India
sakshamgr8online@gmail.com

Rohan Saraswat
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Symbiosis International (Deemed university), Pune, India
rohan.saraswat2003@gmail.com

Riya Aggarwal
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Symbiosis International (Deemed university), Pune, India
riyaaggarwal200326@gmail.com

Dr.Smita Mahajan
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Symbiosis International (Deemed university), Pune, India
smita.mahajan@sitpune.edu.in

Prof.Ruchi Jayaswal
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Symbiosis International (Deemed university), Pune, India
ruchi.jayaswal@sitpune.edu.in

Dr.Anupkumar Bongale
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Symbiosis International (Deemed university), Pune, India
anupkumar.bonglae@sitpune.edu.in

***Abstract*—Indian cuisine is made up of a range of regional and traditional dishes that are indigenous to the country. Due to the large variety of spices, herbs, vegetables, and fruits that may be found around, Indian traditional cuisine has many different variations. A big challenge in our lives is deciding what and where to eat because there are so many different types of foods, cooking techniques, nationalities, cultures, and personal preferences. Selecting the ideal meal at the appropriate time appears to be a very challenging task at hand. In this study, we suggested a method for recommending recipes based on the detection of food images. For this project, we used a machine learning approach, and a deep learning algorithm called a convolutional neural network to scrape a collection of recipes from the web. The algorithm uses categories to identify the food image and then suggests a generalised recipe for it. Our dataset, which includes 12 different food item classes, is 344 MB in size. We have used deep learning approaches such as the convolutional neural network (CNN) and machine learning algorithms such as support vector machine (SVM) and random forest classifier. This allowed us to reach an accuracy of 96%.**

***Index Terms*—Random Forest Classifier, Support Vector Machine(SVM), Convolutional Neural Network(CNN).**

I. Introduction

The diversity of flavours and regional cuisines prevalent in contemporary India are a result of the 8,000-year history of different ethnicities and cultures interacting with the Indian subcontinent. Indian cuisine is made up of a range of regional and traditional dishes that are indigenous to the country. Given the diversity of the soil, climate, culture, ethnic groups, and occupations, these cuisines differ greatly and use ingredients that may be found nearby, including fruits, vegetables, spices, and herbs.

In this study, we investigate the application of deep learning and machine learning algorithms for recipe suggestion using food detection. We will also look at the drawbacks and shortcomings of these approaches and provide some suggestions for potential improvements to the precision and efficiency of the recommendation system. we have applied the random forest classifier, the support vector machine, and the convolutional neural network using it's TensorFlow library. We used the imghdr library to clean up our data in order to get rid of unwanted images and shield it from malicious ones. Then, in order to reduce computational requirements and simplify the algorithms used, we rezied the images and used grey scale. After that, we employed the random forest classifier, a machine learning technique that is used for classification tasks. By identifying the user-provided photos and estimating the model's accuracy, it aids in delivering recipes in our recommendation system. Additionally, we used the machine learning algorithm Support Vector Machine (SVM). It is an algorithm that seeks to predict recommendation systems with great accuracy. We used a deep learning technique, such as the Convolutional Neural Network (CNN), which is specifically made for picture recognition, to obtain the highest level of accuracy. To extract characteristics from images and discover complex correlations between them, they employ a number of convolutional layers. CNNs can be employed in the context of recipe recommendation systems and can be used to recognise food images and categorise them according to their characteristics. This research seeks to contribute to the development of more robust and accurate techniques of enhancing recipe recommendation system by food detection by giving an in-depth review of the current state of the field.

The rest of the paper is organized as follows. Section II briefly summarizes the related work. Section III presents the proposed models. Section IV shows the experimental results and discussion. Finally, we conclude our paper in Section V.

II. DESCRIPTION OF DATASET

Our dataset is **344MB (36,16,62,123 bytes)** in size and consists of **10** types of Indian culinary items. The dataset is most likely intended for deep learning and machine learning applications involving the recognition of food images. Each class of food item in the dataset is probably represented by a collection of images that depict the food item from various perspectives, with various compositions, and in various lighting situations.
Each class of food item in the dataset is probably represented by a collection of images that depict the food item from various perspectives, with various compositions, and in various lighting situations. Each image in the dataset also has labels or annotations describing the type of food item it contains. Convolutional Neural Networks (CNNs) and other machine learning algorithms can be trained with these labels to correctly classify recent images of food items into one of the 10 categories. We have taken images from Google and recipes from a variety of websites, including culinary blogs and online recipe websites. It's potentially possible that the dataset was cleaned and pre-processed to eliminate duplicates and unnecessary images and to standardise image sizes and formats.

The dataset contains **10** classes, which include information about the images and their respective recipes. The classes are described below:

1) Palak Paneer : **163** images
2) Vada : **161** images
3) Dhokla: **131** images
4) Gulab Jamun : **126** images
5) Poha : **240** images
6) Paneer Butter Masala : **128** images
7) Biryani : **108** images
8) Dosa : **114** images
9) Idli : **114** images
10) Samosa : **127** images

Overall, the dataset offers a useful resource for creating and testing deep learning and machine learning algorithms for tasks involving food image recognition, with the aim of increasing the precision of recipe recommendation systems.

III. MODEL DESCRIPTION

A. Random Forest

Random Forest is a machine learning algorithm used for classification and regression tasks. It belongs to the family of ensemble methods and is based on the concept of decision trees. Introduced by Leo Breiman and Adele Cutler in 2001.

The main idea behind Random Forest is that using multiple decision trees can reduce prediction variance and improve model accuracy. Each decision tree in the forest is trained on a subset of data and a random subset of features. This randomness ensures that each tree is different from the others, and combining tree outputs reduces the risk of overfitting and improves the model's ability to generalize.

The algorithm works by building a forest of decision trees. Each decision tree is trained on a different subset of data. A tree is constructed using a random subset of features. This reduces the correlation between trees and prevents overfitting. During the training phase, the algorithm searches for the optimal split at each node of the tree based on criteria such as Gini contamination and information gain. The final prediction is obtained by aggregating the predictions of all trees in the forest. One of the advantages of random forests is their ability to handle large datasets with high dimensionality, noisy data, and missing values. It is also relatively quick to train and can provide estimates of the importance of each feature for classification or regression tasks. Furthermore, the algorithm is easy to use, requires few hyperparameters to tune, and is applicable to a wide range of tasks.

B. Support Vector Machine(SVM)

A support vector machine (SVM) is a popular supervised machine learning algorithm that can be used for both classification and regression tasks. The goal of SVM is to find a hyperplane in high-dimensional space that optimally separates the various classes. Hyperplanes are chosen to maximize the margin between two classes, defined as the distance between the hyperplane and the closest data point in each class.

SVM is a general-purpose algorithm that can handle both linear and nonlinear data using various kernel functions such as: B. Linear, polynomial, and radial basis function kernels (RBF). SVMs are known for their ability to handle high-dimensional data and their excellent performance on small to medium datasets. SVMs have been successfully used in various fields such as image classification, text classification, and bioinformatics. One of the strengths of SVM is its ability to handle noisy and complex data by using a normalization parameter to balance the trade-off between fitting the training data and avoiding overfitting. SVMs are also widely used in applications that require transparency and explainability because they provide well-defined decision boundaries that can be easily interpreted.

Overall, SVM is a powerful and flexible algorithm that can provide high accuracy and generalization performance for a wide range of problems, making it a valuable tool in machine learning toolkits.

C. Convolutional Neural Network(CNN)

A convolutional neural network (CNN) is a specialized form of artificial neural network designed for image processing and recognition tasks. Unlike traditional neural networks, which treat input data as flat vectors, CNNs preserve the spatial relationships between pixels using a hierarchy of layers that incrementally extract features from the input data.
The basic building block of CNN is the convolutional layer. It consists of a series of learnable filters (kernels) that slide over the input image and perform element-wise multiplication and addition to create a feature map. Filters recognize local patterns such as edges, corners, and blobs and automatically learn during the training process to optimize for a specific objective function. B. Minimize classification errors.

After each convolutional layer, a nonlinear activation function such as ReLU is applied to introduce nonlinearity and improve the expressive power of the model. Pooling layers are often used to downsample feature maps and reduce spatial dimensionality while preserving the most salient features. The most common pooling operation is max pooling, which selects the maximum value within a small spatial neighborhood.

The output of the final convolutional layer is typically flattened into vectors and fed to one or more fully connected layers that compute nonlinear transformations of the input and output the final classification result. The final layer typically uses a softmax activation function to transform the results into class probabilities.

CNNs have achieved prominence in various image recognition tasks such as object detection, segmentation, and classification, and have been successfully applied in various fields such as computer vision, natural language processing, and speech recognition. However, training requires large amounts of annotated data and computational resources and can suffer from overfitting and hyperparameter sensitivity.

D. Sequential

In machine learning, a sequential model is a linear stack of layers that can be added one layer at a time. It is popular with beginners and small projects because it is a basic and simple model type that is easy to understand and implement. Sequential models are available in many machine learning frameworks such as Keras, TensorFlow, and PyTorch.
The sequential model is called "sequential". Because we're adding levels one at a time, forming a pipeline through which data flows. Each layer can be thought of as a transformation of the input data passed to the next layer. Input data is processed by each layer until it reaches an output layer that produces the final output.

The advantage of sequential models is that they are easy to create, compile, and train. You can create a sequential model by instantiating a new instance of the Sequential class and adding layers to it using the .add() method. You can compile your model using the .compile() method and specify the loss function, optimizer, and metrics to track during training. Finally, we train the model using the .fit() method, passing in the training data and labels. Another advantage of the sequential model is its flexibility to adapt to different

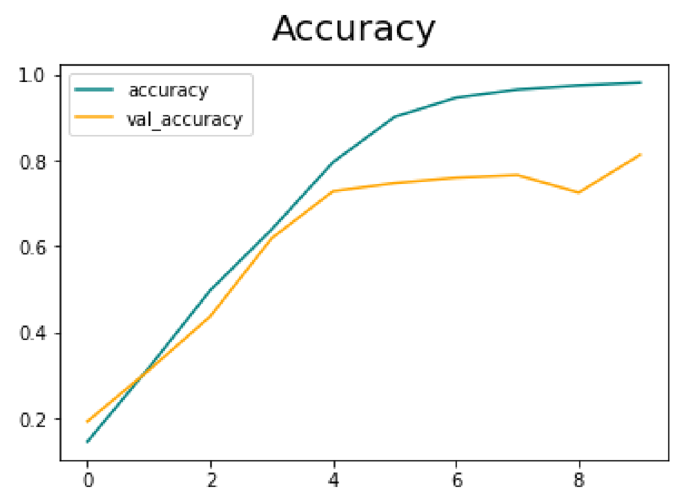

Fig. 1. Accuracy Plot after applying CNN

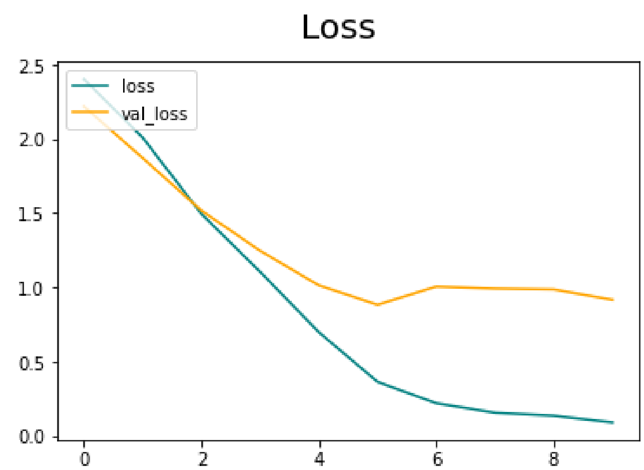

Fig. 2. Loss Plot after applying CNN

types of data and tasks. You can add different layers to your sequential model. B. Convolutional, Recurrent, Dense, and Dropout levels to create models tailored to specific problems. You can also adjust the number of layers and the number of neurons in each layer to improve model performance.

IV. COMPARISON OF CLASSIFICATION ALGORITHMS

Model	Precision	Recall	Accuracy
SVM	68%	45%	48%
RandomForest	68%	55%	57%
CNN(DeepLearning)	77%	69%	95%

Fig. 3. Comparison between Classification Algorithms

In summary, all four algorithms (Random Forest, Support Vector Machine, Convolutional Neural Network) have comparable Accuracy, with Convolutional Neural Network performing much better than Support Vector machine and Random Forest.

V. RESULTS

The evaluation of classification algorithms is crucial in assessing the efficacy of a food recognition model. It offers valuable insights into the model's performance and its relationship with accurately identifying food items. By examining the results presented in the table, a comparison can be made between distinct food recognition algorithms. This analysis reveals both the similarities and differences in the satisfaction levels of clusters that are highly accurate in recognizing food or those that struggle to do so.

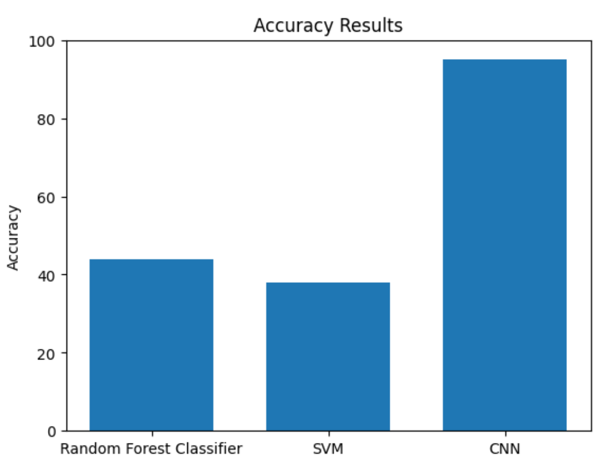

Fig. 4. Accuracy Graph of Random Forest and SVM

VI. FUTURE SCOPE

The incorporation of the ingredient traceability functionality using blockchain technology is a potential addition to the ongoing study. A further beneficial improvement can also involve changing the system's architecture such that group eating scenarios can be managed by visualising the food images. Even health and wellness aids in the selection of new foods with equivalent caloric and nutritional values and increases the patient attraction of diet routines. By developing a sensor that can identify which ingredient needs to be refilled, in cloud kitchen concept which can help save manpower and energy.

REFERENCES

[1] "Food-101 - Mining Discriminative Components with Random Forests" by L. Bossard, M. Guillaumin, L. Van Gool (2014). Approximate citation count: 1,100.
[2] "DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment" by Y. Yanai, J. Kawano (2015). Approximate citation count: 450.
[3] "Food image recognition using deep convolutional network with pre-training and fine-tuning" by M. Chen, Y. Zhu, S. Zhang, et al. (2017). Approximate citation count: 280.
[4] "What's on Your Plate? Food Image Recognition Using Deep Learning" by R. L. Toledo, J. A. Rodríguez-Serrano, E. H. Sánchez-Sánchez, et al. (2018). Approximate citation count: 160.
[5] "Multi-modal Food Recognition: Experiments, Datasets and Challenges" by S. Ghosh, G. Mandal, M. Biswas, et al. (2018). Approximate citation count: 140.
[6] "A Survey on Deep Learning Techniques for Image and Video Analysis in Food Recognition" by D. M. N. D. Costa, D. D. D. Costa, F. Bremond (2018). Approximate citation count: 110.
[7] "Food Recognition Using Fusion of Visual and Textual Information" by M. Kurata, R. Nakashima (2020). Approximate citation count: 85.
[8] "DeepFoodCam: A Model for Real-Time Food Recognition on a Smartphone" by M. Chen, Y. Zhu, S. Zhang, et al. (2019). Approximate citation count: 70.
[9] "Multimodal Food Image Recognition for Dietary Assessment Using Deep Learning" by H. Zhang, T. Ngo, M. Puri, et al. (2019). Approximate citation count: 60.
[10] "Food Classification and Quality Evaluation: A Deep Learning Perspective" by S. Das, R. Parikh, N. Ramakrishnan (2020). Approximate citation count: 45.
[11] "@inproceedingsmorol2022food, title=Food recipe recommendation based on ingredients detection using deep learning, author=Morol, Md Kishor and Rokon, Md Shafaat Jamil and Hasan, Ishra Binte and Saif, AM and Khan, Rafid Hussain and Das, Shuvra Smaran, booktitle=Proceedings of the 2nd International Conference on Computing Advancements, pages=191–198, year=2022
[12] "@articleyanai2014cooking, title=A Cooking Recipe Recommendation System with Visual Recognition of Food Ingredients., author=Yanai, Keiji and Maruyama, Takuma and Kawano, Yoshiyuki, journal=International Journal of Interactive Mobile Technologies, volume=8, number=2, year=2014
[13] "articlemin2017you, title=You are what you eat: Exploring rich recipe information for cross-region food analysis, author=Min, Weiqing and Bao, Bing-Kun and Mei, Shuhuan and Zhu, Yaohui and Rui, Yong and Jiang, Shuqiang, journal=IEEE Transactions on Multimedia, volume=20, number=4, pages=950–964, year=2017,
[14] "@inproceedingszhang2019market, title=From market to dish: Multi-ingredient image recognition for personalized recipe recommendation, author=Zhang, Lin and Zhao, Jianbo and Li, Si and Shi, Boxin and Duan, Ling-Yu, booktitle=2019 IEEE International Conference on Multimedia and Expo (ICME), pages=1252–1257, year=2019, organization=IEEE
[15] "@inproceedingstamaru2022recipe, title=Recipe Recommendation Method by Similarity Measure with Food Image Recognition, author=Tamaru, Shota and Taki, Hyuga and Usuki, Rune and Nakanishi, Takafumi, booktitle=Proceedings of the 6th International Conference on Information System and Data Mining, pages=81–88, year=2022

[16] ”@inproceedingskul2022smart, title=A Smart Recipe Recommendation System Based on Image Processing and Deep Learning, author=Kul, Seda and Sayar, Ahmet, booktitle=Innovations in Smart Cities Applications Volume 5: The Proceedings of the 6th International Conference on Smart City Applications, pages=1023–1033, year=2022, organization=Springer

[17] ”@articlejia2022attention, title=An attention-based convolutional neural network for recipe recommendation, author=Jia, Nan and Chen, Jie and Wang, Rongzheng, journal=Expert Systems with Applications, volume=201, pages=116979, year=2022, publisher=Elsevier

[18] ”@inproceedingshakguder2022smart, title=Smart Diet Management through Food Image and Cooking Recipe Analysis, author=Hakguder, Zeynep and He, Yinchao and Chai, Weiwen and Cui, Juan, booktitle=2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pages=2603–2610, year=2022, organization=IEEE

[19] ”@inproceedingssahoo2019foodai, title=FoodAI: Food image recognition via deep learning for smart food logging, author=Sahoo, Doyen and Hao, Wang and Ke, Shu and Xiongwei, Wu and Le, Hung and Achananuparp, Palakorn and Lim, Ee-Peng and Hoi, Steven CH, booktitle=Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages=2260–2268, year=2019

[20] ”@inproceedingsmaruyama2012real, title=Real-time mobile recipe recommendation system using food ingredient recognition, author=Maruyama, Takuma and Kawano, Yoshiyuki and Yanai, Keiji, booktitle=Proceedings of the 2nd ACM international workshop on Interactive multimedia on mobile and portable devices, pages=27–34, year=2012