




Proceeding Paper

# DenseMobile Net: Deep Ensemble Model for Precision and Innovation in Indian Food Recognition <sup>†</sup>

Jigarkumar Ambalal Patel <sup>1,\*</sup>, Gaurang Vinodray Lakhani <sup>1</sup>, Rashmika Ketan Vaghela <sup>2</sup>  
and Dileep Laxmansinh Labana <sup>3</sup>

<sup>1</sup> Government Polytechnic, Bhuj 370001, India; gvlakhani1@gmail.com

<sup>2</sup> Government Polytechnic, Ahmedabad 380006, India; rashmivaghela.rv@gmail.com

<sup>3</sup> R C Technical Institute, Ahmedabad 380060, India; dileep.labana@gmail.com

\* Correspondence: jigarsharp@gmail.com; Tel.: +91-9662032881

<sup>†</sup> Presented at the 5th International Electronic Conference on Applied Sciences, 4–6 December 2024;

<https://sciforum.net/event/ASEC2024>.

**Abstract:** Precision and efficacy are vital in the constantly advancing field of food image identification, particularly in the domains of medicine and healthcare. Transfer learning and deep ensemble learning techniques are employed to enhance the accuracy and efficiency of the Indian Food Classification System. The ensemble model effectively captures various patterns and correlations within the information by employing many machine learning techniques. The ensemble method we employ utilizes the MobileNetV3 and DenseNet-121 transfer learning models to construct a robust model. The ensemble model benefits from the integration of model predictions, resulting in enhanced recognition of Indian food. The study utilized a dataset consisting of 6000 photographs of Indian cuisine, categorized into 26 distinct groups. The picture dataset is divided into two subsets: 80% is allocated for training and 20% is reserved for testing. The experimental results demonstrate that DenseNet-121 surpasses MobileNetv3 in terms of testing accuracy, achieving a rate of 90%. The MobileNetV3 model achieves an accuracy of 87.64% on the Indian food image dataset. The integration of both models in ensemble learning yields a model accuracy of 92.38%, surpassing the performance of each individual model. This research revolutionizes our food relationship with the use of state-of-the-art technologies. By utilizing the most advanced transfer learning algorithm specifically designed for the precise classification of Indian cuisine, our aim is to establish a new standard in both technology and gastronomy. This will facilitate innovation in food perception, comprehension, and engagement.



Academic Editor: Francesco Dell'Olio

Published: 7 February 2025

**Citation:** Patel, J.A.; Lakhani, G.V.; Vaghela, R.K.; Labana, D.L.

DenseMobile Net: Deep Ensemble Model for Precision and Innovation in Indian Food Recognition. *Eng. Proc.* **2025**, *87*, 3. <https://doi.org/10.3390/engproc2025087003>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** ensemble learning; transfer learning; DenseNet-121; MobileNetV3; Indian food images

## 1. Introduction

There have been significant developments in the field of food image recognition in recent times. The quest for precision and innovation in recognizing diverse and intricate culinary creations, especially within the context of Indian cuisine, has become imperative. This research endeavours to transcend the existing paradigms by harnessing the transformative power of deep ensemble learning for the elevated recognition of Indian food. The importance of food image classification and recognition is recognized from multiple perspectives [1]. The predominant determinant of individuals' health difficulties, such as diabetes and obesity [2], is commonly recognized to be their dietary habits. Evaluating food consumption is a vital method in the control of obesity employed in fitness programmes.

With its wide range of tastes, ingredients, and regional differences, Indian food presents a special challenge for automated food recognition systems [3]. Because of their complex composition, which is shaped by centuries of cultural legacy, meals require a nuanced approach that goes beyond traditional methods. Taking into account this requirement, we conducted research on using deep ensemble learning—a state-of-the-art method that combines the best features of several models—to advance the field of Indian cuisine recognition.

Based on the existing DenseNet-121 [4] and MobileNetV3 [5], the proposed model was developed. DenseNet-121 is chosen for its high accuracy and efficient feature reuse, making it ideal for tasks requiring deep feature extraction. MobileNetV3 is selected for its lightweight design, optimized for real-time performance and efficiency, especially on mobile or resource-constrained devices. An ensemble approach combines these models to improve accuracy and robustness by leveraging their complementary strengths, reducing errors, and enhancing performance across various conditions. The following summarizes the proposed model's significance:

- Using pre-trained weights of Densenet-121 and MobileNetV3 on the ImageNet dataset, the suggested method for Indian cuisine detection works well even when the dataset is unbalanced.
- Deep ensemble learning leads to improved accuracy in recognizing a wide variety of Indian foods and could also facilitate new discoveries in how we engage with and comprehend food.
- The proposed approach demonstrates enhanced accuracy compared to the current state-of-the-art methods for food classification.

## 2. Related Work

Several classification and recognition approaches are at our disposal, including support vector machine [6], artificial neural networks [7], and random forest. Convolutional neural networks [8] are highly suited for the problem of food recognition because of their exceptional performance and efficacy in the area of image feature extraction. Nevertheless, a large dataset with millions of images is usually required to train a CNN model that satisfies high requirements.

We reviewed the work of several researchers, including Md Tohidul Islam, Yuzhen Lu, VijayaKumari G, Chang Liu, Michele De Bonis, Rajayogi, and N. Hnoohom, to gain a deeper understanding of the performance of different CNN models on different types of cuisine. A summary of the literature review is shown in Table 1.

**Table 1.** Summary of literature review.

Researcher	Algorithm	Dataset	Accuracy
Md Tohidul Islam [9]	CNN (Self-Designed)	Food-11	74.70%
Yuzhen Lu [10]	BoF with SVM	Custom-10 Classes	90%
VijayaKumari G [11]	EfficientNetb0	Food-101	80%
Chang Liu [12]	CNN	Food-101	77.4%
Michele De Bonis [13]	GoogleNet	UPMC 101	70%
	SqueezeNet	ETHZ 101	60%
Rajayogi [14]	InceptionV3	Indian Food	87.9%
N. Hnoohom [15]	GoogleNet	Thai Fast Food	88.33%

Based on the literature review, it has been found that there is a lack of extensive research on the classification of Indian cuisine products. This paper introduces a deep learning-based approach for classifying Indian cuisine. In this study, transfer learning techniques

were employed to enhance the performance of the system. For the purpose of performance evaluation, we employed two models, namely DenseNet-121 and MobileNetV3, that were trained on the Indian cuisine dataset. To enhance performance, we employed ensemble learning by utilising these two models [16].

### 3. Methodology

This section outlines the proposed transfer learning and ensemble learning for classifying and recognizing Indian foods. Here, we used MobileNetV3 and DenseNet-121 models, which are pre-trained on the ImageNet dataset. A diagram of the proposed model is shown in Figure 1.

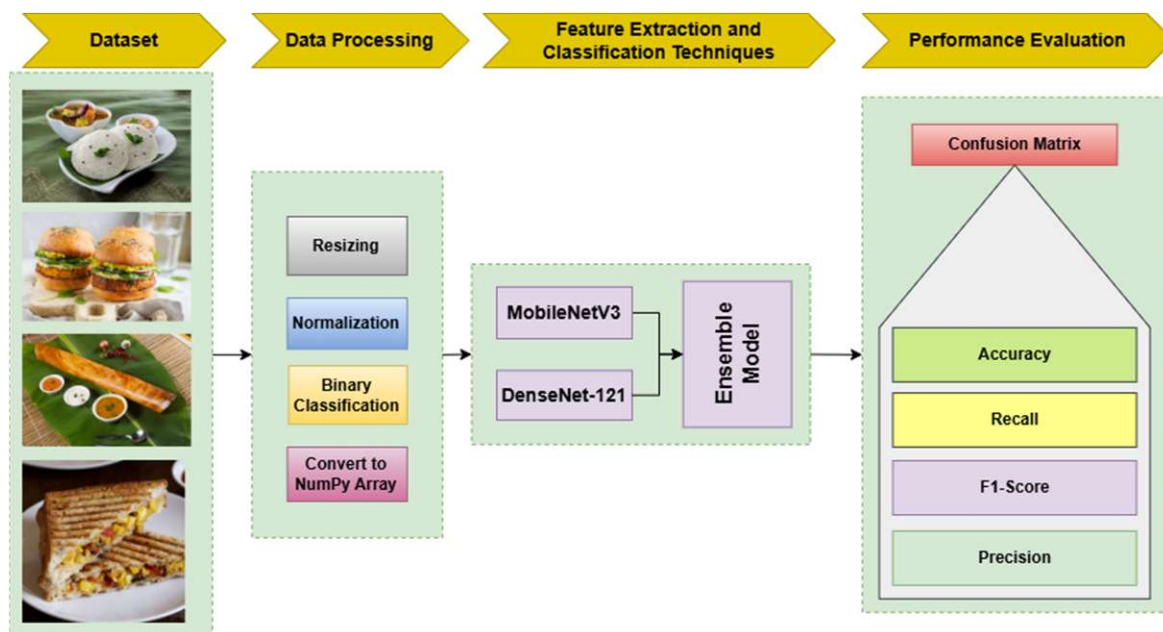


Figure 1. Indian food classification system architecture.

#### 3.1. Indian Food Dataset

The research utilized the “26\_Indian\_food dataset”. The collection has more than 6000 images. The dataset comprises an extensive assortment of Indian food images, spanning over 26 unique categories. Every category in the collection consists of a significant number of images, from 100 to 300 images per category. This study will primarily concentrate on the food products that are native to our local region. To address the problem of overfitting, 20% of the training data are set aside for validation.

#### 3.2. Data Pre-Processing

After obtaining the dataset, we performed preprocessing on the dataset’s photos. Image preprocessing can significantly enhance the quality of images for subsequent analysis and processing. Normalization, contrast enhancement, and image rescaling are all crucial preprocessing techniques. To obtain an understanding of how the quantity of images is distributed among different categories, a list was created consisting of 26 categories, each accompanied by the corresponding number of images. Furthermore, the photos were subjected to rescaling throughout the conversion procedure, accomplished by dividing each individual pixel value by 255.

### 3.3. Feature Extractions and Classification

We employed the convolutional network method for the purpose of extracting features and performing classification on the photos. For transfer learning, we employed the pre-trained CNN models MobileNetV3 and DenseNet-121. Subsequently, we constructed an ensemble learning model by employing these techniques.

#### 3.3.1. Transfer Learning with MobileNetV3

MobileNet is a streamlined and highly efficient convolutional neural network designed specifically for mobile vision tasks. It is characterized by its simplicity and low computational requirements. MobileNetV3 is a convolutional neural network structure specifically created for mobile applications that require efficiency and low computational requirements. The fundamental concept underlying MobileNetV3 is to employ a blend of diverse methodologies in order to attain an optimal balance between the size of the model, the computational expense, and the level of accuracy.

##### Architecture of MobileNetV3

The MobileNetV3 model consists of five primary components: input layers, convolution block, activation function, squeeze-and-excitation (SE) blocks and fully connected layers. These components are described in detail below.

1. **Input Layer:** It is consisting of an image or a tensor that serves as the data to be processed.
2. **Convolution Block:** MobileNetV3 employs inverted residual blocks featuring linear bottlenecks in its convolutional architecture. The blocks are composed of a sequence of depth-wise separable convolutions and pointwise convolutions, which effectively decrease computational complexity while preserving the capacity to capture significant features.
3. **Activation Function:** It plays a crucial role in deep learning models. MobileNetV3, incorporates sophisticated activation functions, including the hard-swish activation function. This particular activation function introduces non-linearity to the network, enhancing its ability to capture complex patterns and improve overall performance.
4. **Squeeze-and-Excitation (SE) Blocks:** It allows for the adaptive recalibration of feature maps by learning the relative importance of different channels. This enhances the overall representation power of the model.
5. **Fully Connected Layer:** Before the fully connected layer is employed, the first layer flattens to decrease the spatial dimensions of an image and generate a feature vector of a consistent length, regardless of the size of the input image.

MobileNetV3 was employed for the task of extracting features. Afterwards, the food images were categorized using the dense layer, which had a number of units equal to the entire number of food categories in the dataset, which was 26 in this case. The SoftMax function was utilized in this specific instance due to the presence of a multiclass classification challenge.

#### 3.3.2. Transfer Learning with Densenet-121

The latest architectural innovation in the field is the DenseNet, which is a Densely Connected Convolution Network. The high level of interconnectivity in the DenseNet architecture is crucial as it allows for effective reuse of features and reduction in parameters. DenseNet has superior efficiency when compared to other models, thanks to its streamlined structure and decreased number of parameters. We employ the DenseNet model that has undergone pre-training on the ImageNet dataset and further fine-tuning on our own dataset.

### Architecture of DenseNet

1. The first layers: Before images are fed into the dense block layers, a significant amount of information is extracted from them using the convolution and pooling layers. To reduce the total number of parameters, down sampling is applied to the feature maps.
2. Dense Block Layers: To make information flow between layers easier, the DenseNet architecture uses a dense connection. The convolution layer, activation function, and batch normalization make up the dense block. Each layer's input is sent to the layers that follow it. As a result, all feature maps from earlier layers are sent to the nth layer. The inputs can be written as follows:

$$X_n = H_n [(X_0, X_1, \dots, X_{n-1})] \quad (1)$$

In this case,  $X_n$  stands for the nth layer's output.

3. Interfacial layers, another name for transition layers, are areas that lie between two distinct phases. These layers contribute to down sampling in convolutional neural network architecture. The network design is divided into four dense blocks to reduce the feature map's dimension. Transition layers are the middle layers that are located in the network design between the dense blocks. Batch normalization, a rectified linear unit (ReLU) activation function, a  $1 \times 1$  convolution layer, and a  $2 \times 2$  max pooling layers are some of the components that make up these transition layers. It is preferable to use maximum pooling rather than average pooling when down sampling. This decision was made in order to detect important characteristics, like edges, and efficiently reduce the dimensionality of features. It should be mentioned that average pooling is a method that gradually and continuously retrieves features. Compression is used to reduce the number of feature maps, reduce the possibility of overfitting, and improve the generalizability of the model.
4. Fully Connected Layers: Two fully connected layers make up the model's last section. By using global average pooling, the first layer flattens the feature maps, producing a linear array with 1024 nodes. The fully connected layer, which serves as a classifier, then receives this array as input. Each of the 26 neurons that make up the second layer represents a different food class.

#### 3.3.3. Ensemble Learning with MobileNetV3 and DenseNet-121

Deep ensemble learning involves creating an ensemble of numerous deep neural networks to collectively create predictions and make definitive conclusions. This phenomenon has the potential to lead to increased accuracy, durability, and expanded capacity to apply acquired knowledge to unfamiliar information. In this research, we used densenet-121 and mobileNetV3 for ensemble learning. We used Max-voting for classification work. The idea of Max voting, sometimes known as majority voting, is a commonly used and uncomplicated strategy in ensemble learning, especially for classification tasks. The fundamental concept of max voting is to reduce the possibility of inaccurate predictions that can occur from individual models by combining the predictions of multiple models. Facilitating the improvement of the overall accuracy and robustness of the predictive model.

#### 3.4. Performance Evaluation

The scientific community has widely adopted various criteria to assess the effectiveness of classification systems. The evaluation of research success involves employing a confusion matrix that encompasses crucial parameters like true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN). Validity metrics such as Accuracy, Recall, F1-score, and Precision can be derived from these parameters.

## 4. Implementation and Results

In this research, we used Kaggle.com (accessed on 27 July 2023), an online platform specifically designed for deep learning purposes, to conduct the experiment. The present study utilizes a GPU P100, which has been accessed via a portal, to optimize and accelerate the computational procedures. The implementation of this study was conducted utilizing the Tensorflow and Keras libraries.

### 4.1. Results with MobileNetV3

The training accuracy was determined to be 98%. The testing accuracy achieved a value of 87.64% when evaluated on the testing dataset with transfer learning using the pre-trained MobileNetV3 model. Figure 2 shows Results with MobileNetV3 model.

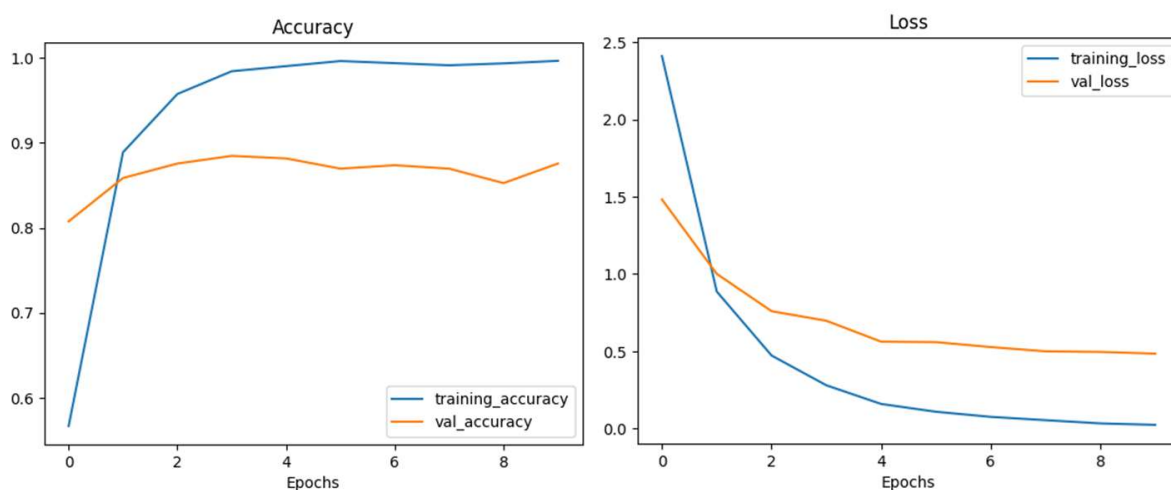


Figure 2. Accuracy and loss graph of system using MobileNetV3.

### 4.2. Results with DenseNet-121

The testing accuracy achieved a value of 89.98% when evaluated on the testing dataset with transfer learning using the pre-trained DenseNet-121 model. Figure 3 shows Results with DenseNet-121 model.

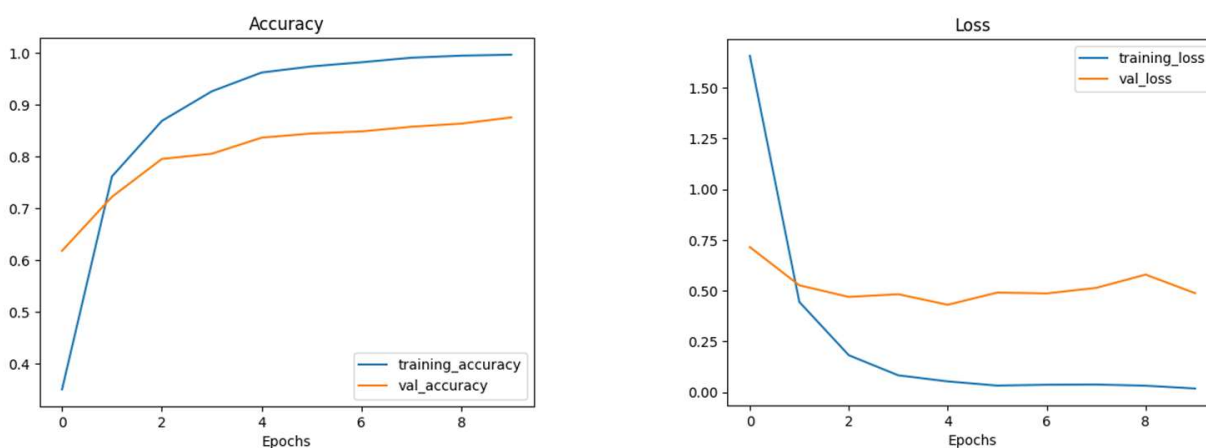


Figure 3. Accuracy and loss graph of system using DenseNet-121.

### 4.3. Results with Ensemble Model

By employing this methodology, we achieved a test accuracy of approximately 92.38%, which exceeds a 2–3% difference from the individual model. Figure 4 shows Results comparison of DenseNet-121, MobileNet and Ensemble model.

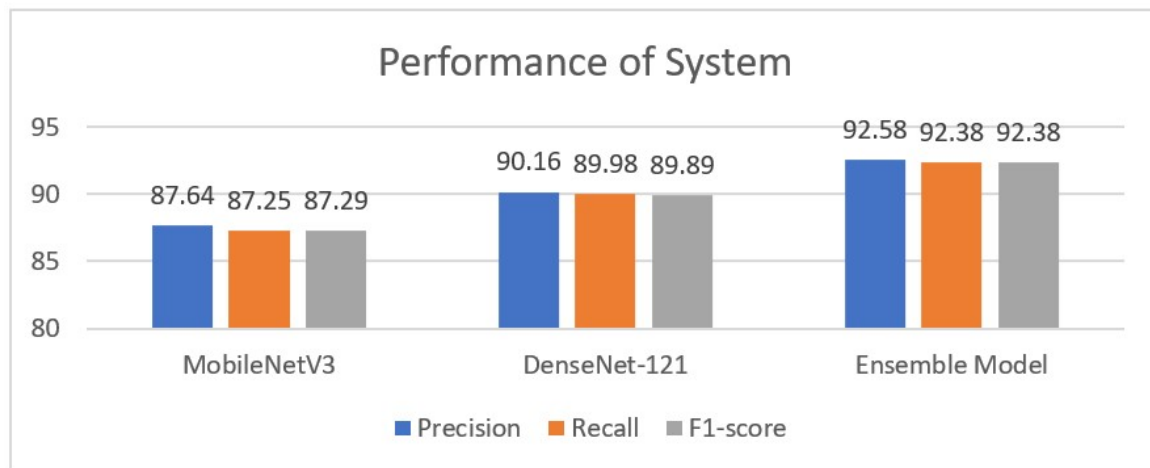


Figure 4. Comparison of results.

## 5. Conclusions

This research presents an innovative method for categorising Indian food by employing transfer learning techniques with MobileNetV3 and DenseNet-121 architectures. The study demonstrates favorable results, with DenseNet-121 obtaining a notable accuracy of 90% and MobileNetV3 at 87.25%. By incorporating ensemble learning techniques with these models, the system's accuracy experiences a substantial improvement, reaching 92.38%. The suggested models outperform earlier studies on Indian food datasets, surpassing the previous highest accuracies of 88% attained using different transfer learning techniques. MobileNetV3, specifically created for limited-resource platforms such as mobile devices, demonstrates its efficiency in this particular scenario.

The research goes beyond classification and envisions potential uses in food systems, calorie estimation, and healthcare specifically designed for the Indian people. Although DenseNet-121 demands somewhat higher computing resources compared to MobileNetV3, it exhibits a 3% increase in accuracy. The integration of ensemble learning significantly improves system performance. Managing multiple models sometimes complicates training, storage, and inference, while an ensemble of similar models may not enhance generalization and could lead to overfitting.

Acknowledging the dataset is imbalanced, future work aims to refine the system by applying data augmentation techniques. This strategic step is anticipated to address imbalances and contribute to the continuous improvement of the proposed model's effectiveness.

**Author Contributions:** Conceptualization, methodology, software, visualization—J.A.P., R.K.V. and D.L.L.; validation writing, supervision—G.V.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Dataset is available at <https://www.kaggle.com/datasets/jigarsharp/26-indianfood> (accessed on 27 July 2023) and the trained proposed model is available at [https://github.com/jigarsharp/Indian\\_food\\_classification.git](https://github.com/jigarsharp/Indian_food_classification.git) (accessed on 27 July 2023).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Sadler, C.R.; Grassby, T.; Hart, K.; Raats, M.; Sokolović, M.; Timotijevic, L. Processed food classification: Conceptualisation and challenges. *Trends Food Sci. Technol.* **2021**, *112*, 149–162. [[CrossRef](#)]
2. Lee, K.-S. Multispectral Food Classification and Caloric Estimation Using Convolutional Neural Networks. *Foods* **2023**, *12*, 3212. [[CrossRef](#)] [[PubMed](#)]
3. Shah, B.; Kanani, P.; Joshi, P.; Pandya, G.; Kulkarni, D.; Patil, N.; Kurup, L. Traditional Indian Food Classification Using Shallow Convolutional Neural Network. *Int. J. Intell. Syst. Appl. Eng.* **2024**, *12*, 769.
4. Hiremath, G.; Mathew, J.A.; Boraiah, N.K. Hybrid Statistical and Texture Features with DenseNet 121 for Breast Cancer Classification. *Int. J. Intell. Eng. Syst.* **2023**, *16*, 24–34.
5. Patel, J.; Modi, K. Indian Food Image Classification and Recognition with Transfer Learning Technique Using MobileNetV3 and Data Augmentation. *Eng. Proc.* **2023**, *56*, 197. [[CrossRef](#)]
6. Lo, F.P.W.; Sun, Y.; Qiu, J.; Lo, B. Image-based food classification and volume estimation for dietary assessment: A review. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 1926–1939. [[CrossRef](#)] [[PubMed](#)]
7. Hameed, K.; Chai, D.; Rassau, A. Texture-based latent space disentanglement for enhancement of a training dataset for ANN-based classification of fruit and vegetables. *Inf. Process. Agric.* **2023**, *10*, 85–105. [[CrossRef](#)]
8. Ren, X.; Wang, Y.; Huang, Y.; Mustafa, M.; Sun, D.; Xue, F.; Wu, F. A CNN-Based E-Nose Using Time Series Features for Food Freshness Classification. *IEEE Sens. J.* **2023**, *23*, 6027–6038. [[CrossRef](#)]
9. Islam, M.T.; Siddique, B.N.K.; Rahman, S.; Jabid, T. Food image classification with convolutional neural network. In Proceedings of the 2018 International Conference on Intelligent Informatics and Biomedical Sciences, Bangkok, Thailand, 21–24 October 2018; pp. 257–262.
10. Lu, Y. Food Image Recognition by Using Convolutional Neural Networks (CNNs). *arXiv* **2019**, arXiv:1612.00983.
11. VijayaKumari, G.; Vutkur, P.; Vishwanath, P. Food classification using transfer learning technique. *Glob. Transit. Proc.* **2022**, *3*, 225–229.
12. Liu, C.; Cao, Y.; Luo, Y.; Chen, G.; Vokkarane, V.; Ma, Y. 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, Wuhan, China, 25–27 May 2016*; Springer: Cham, Switzerland, 2016; Volume 14, pp. 37–48.
13. De Bonis, M.; Amato, G.; Falchi, F.; Gennaro, C.; Manghi, P. Deep learning techniques for visual food recognition on a mobile app. In *Multimedia and Network Information Systems: Proceedings of the 11th International Conference MISSI*; Wrocław, Poland, 12–14 September 2018, Springer: Cham, Switzerland, 2019; Volume 11, pp. 303–312.
14. Rajayogi, J.R.; Manjunath, G.; Shobha, G. Indian food image classification with transfer learning. In Proceedings of the 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, 20–21 December 2019; pp. 1–4.
15. Hnoohom, N.; Yuenyong, S. Thai fast food image classification using deep learning. In Proceedings of the 2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), Chiang Rai, Thailand, 25–28 February 2018; pp. 116–119.
16. Mehta, R.; Singh, K.K. An efficient ear recognition technique based on deep ensemble learning approach. *Evol. Syst.* **2023**, *15*, 771–787. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.