IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204321 |

# Indian Food Image Classification using Transfer Learning with ViT

**Prof: Deepali Junankar[1], Joeal Bijoy Thomas[2], Abhijay Kshirsagar[3], Pavan Borate[4], Prathamesh Sharma[5]**

Professor, Department of Computer Engineering, Indira College of Engineering and Management, Maharashtra, India[1]

Students, Department of Computer Engineering, Indira College of Engineering and Management, Maharashtra, India[2,3,4,5]

**ABSTRACT:** The image classification system for Indian cuisine has been built by employing transfer learning techniques with the Vision Transformer (ViT) model. The objective is to accurately classify images of various Indian dishes into 20 distinct classes. The ViT model is leveraged due to its ability to handle both spatial and positional information effectively, making it suitable for image-based tasks.[1]

The dataset comprises a diverse collection of high-resolution images representing popular Indian dishes such as biryani, dosa, samosa, and more. Transfer learning is applied by fine-tuning a pre-trained ViT model on this dataset, which enables the system to learn intricate features specific to Indian cuisine.

Once the model is trained and validated, it is integrated into an application that allows users to upload images of Indian dishes. The system then classifies the dishes with high accuracy and provides corresponding recipes, enhancing the user experience by offering additional culinary information.

Overall, this implementation demonstrates the effectiveness of transfer learning with ViT in image classification tasks related to Indian cuisine, facilitating the automatic identification of dishes and enriching users' culinary exploration with associated recipes.

**KEYWORDS:** Indian food, image classification, transfer learning, Vision Transformer (ViT), deep learning, dataset, fine-tuning, recipe generation, culinary exploration, user experience.

## I. INTRODUCTION

India's culinary heritage is as diverse as its culture, with a myriad of flavors, ingredients, and cooking techniques that vary significantly across regions. The exploration and appreciation of Indian cuisine have transcended geographical boundaries, making it a global favorite. However, despite its popularity, there are challenges in accurately identifying and understanding the vast array of Indian dishes, especially for those less familiar with the cuisine. This project seeks to address these challenges through advanced technology and culinary knowledge integration.

In recent years, the field of computer vision has witnessed remarkable advancements, especially in tasks like image classification, thanks to the advent of deep learning and transfer learning techniques. This project focuses on leveraging transfer learning using Vision Transformer (ViT) for the classification of Indian food images.[1] The ViT model achieves an accuracy of **92%** on the specified task. The model is trained on a dataset comprising 20 distinct classes of Indian dishes, covering a wide range of culinary delights from various regions of India.

## II. METHODOLOGY

**Step 1:**

In the pursuit of accurate image classification, the methodology begins with the acquisition of a diverse dataset comprising images of Indian dishes labeled with their respective names. This dataset is crucial for training and testing the ViT model effectively, necessitating inclusivity across various types of Indian cuisine to ensure comprehensive coverage.[1]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204321 |

**Step 2:**
Following dataset acquisition, meticulous preprocessing of the images is undertaken. This involves resizing the images to fit the input size required by the chosen ViT model, ensuring uniformity across the dataset. Additionally, normalization of pixel values is performed to standardize the data range, typically between 0 and 1, thereby facilitating model convergence during training.[2]

**Step 3:**
The next step involves the selection of a suitable pre-trained ViT model, such as ViT-G/14 or ViT-L/16, based on considerations like model size and computational resources. Transfer learning techniques are then employed to fine-tune the selected ViT model on the acquired Indian food dataset. This process involves adjusting hyperparameters, including learning rates and batch sizes, through iterative experimentation to optimize model performance.

**Step 4:**
Subsequently, the ViT model undergoes training on the prepared dataset while closely monitoring validation performance to prevent overfitting. Techniques such as data augmentation are incorporated to augment the training data, enhancing the model's ability to generalize to unseen images.[5] Early stopping mechanisms are implemented to halt training when validation performance stabilizes, thus conserving computational resources.[1]

**Step 5:**
Evaluation of the trained ViT model is conducted using standard metrics such as accuracy, precision, recall, and F1 score on a separate test set.[4] This assessment validates the model's performance and generalization capabilities, ensuring robustness and reliability in real-world applications.[4]

**Step 6:**
Finally, the trained ViT model is integrated with a recipe database to provide users with personalized recipe recommendations based on predicted dish classes. User accessibility in deployment is prioritized through the design of an intuitive interface, enabling seamless interaction with the model and enhancing user experience.

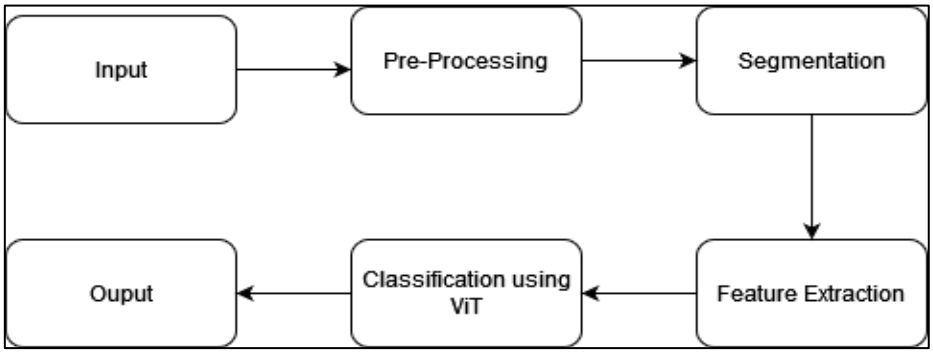


**Figure 1 Methodology**

## III. MODELING AND ANALYSIS

The model used in this project is a Vision Transformer (ViT), a state-of-the-art deep learning architecture tailored for visual tasks. Unlike traditional convolutional neural networks (CNNs), ViT leverages self-attention mechanisms to capture global relationships within images, making it highly effective at understanding complex visual patterns.[1] By fine-tuning a pre-trained ViT model such as ViT-G/14 or ViT-L/16 on a dataset of Indian food images, the model learns to discern the unique features and characteristics of different dishes. This adaptation, coupled with transfer learning techniques, enables accurate classification of diverse Indian dishes, ultimately facilitating the provision of recipe recommendations based on predicted classes.[2]

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204321 |

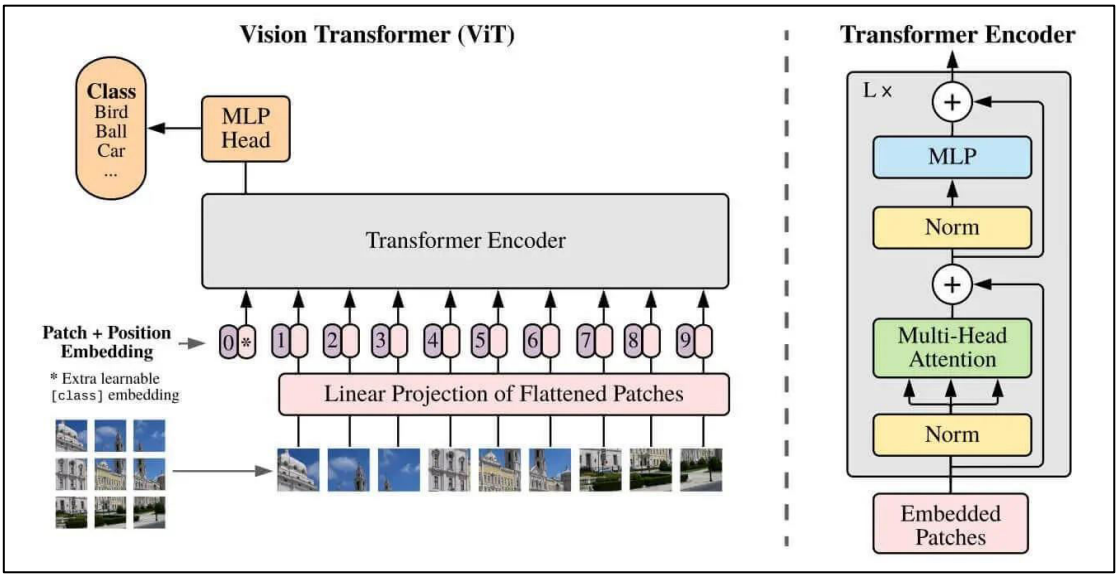


**Figure 2 System Architecture**

## IV. RESULTS AND DISCUSSION

PyCharm along with Vision Transformer (ViT) model yielded promising results in classifying Indian food images. The ViT model, after fine-tuning on a diverse dataset of Indian dishes, achieved high accuracy in identifying various types of cuisine accurately.

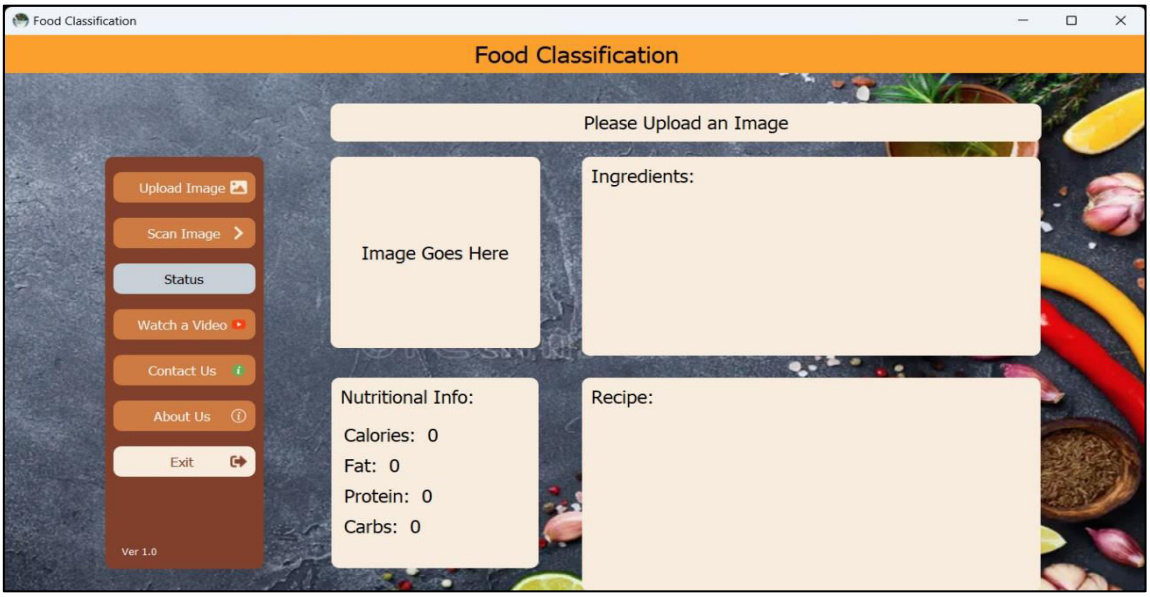


**Figure 3 Home Page**

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204321 |

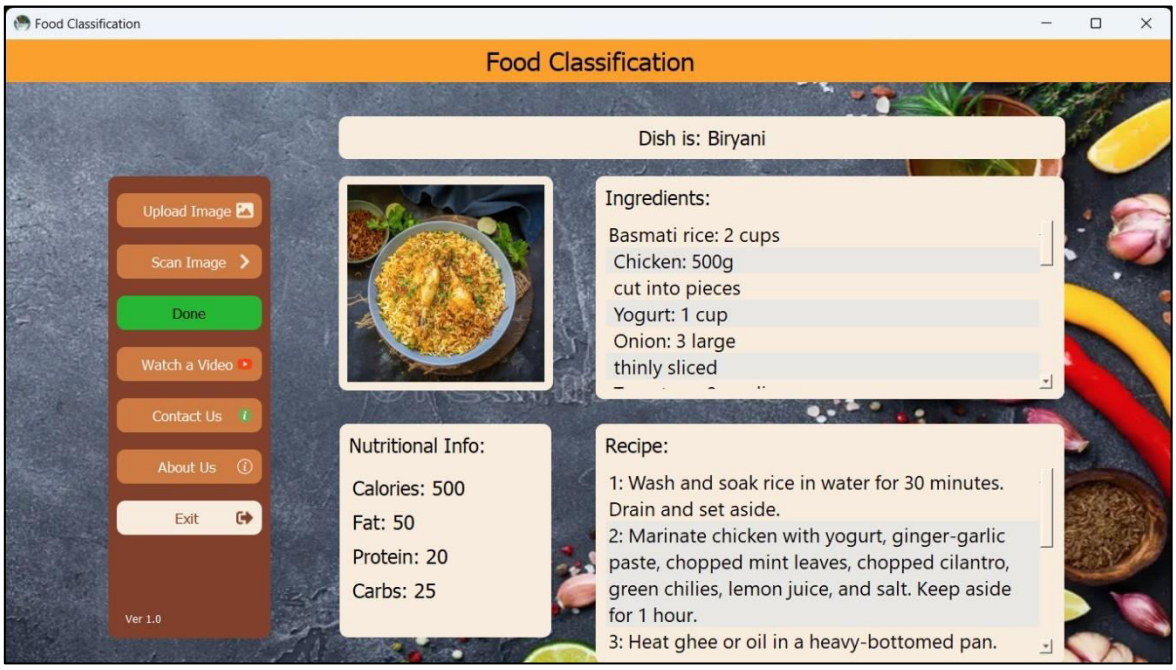


**Figure 4 Output Page**

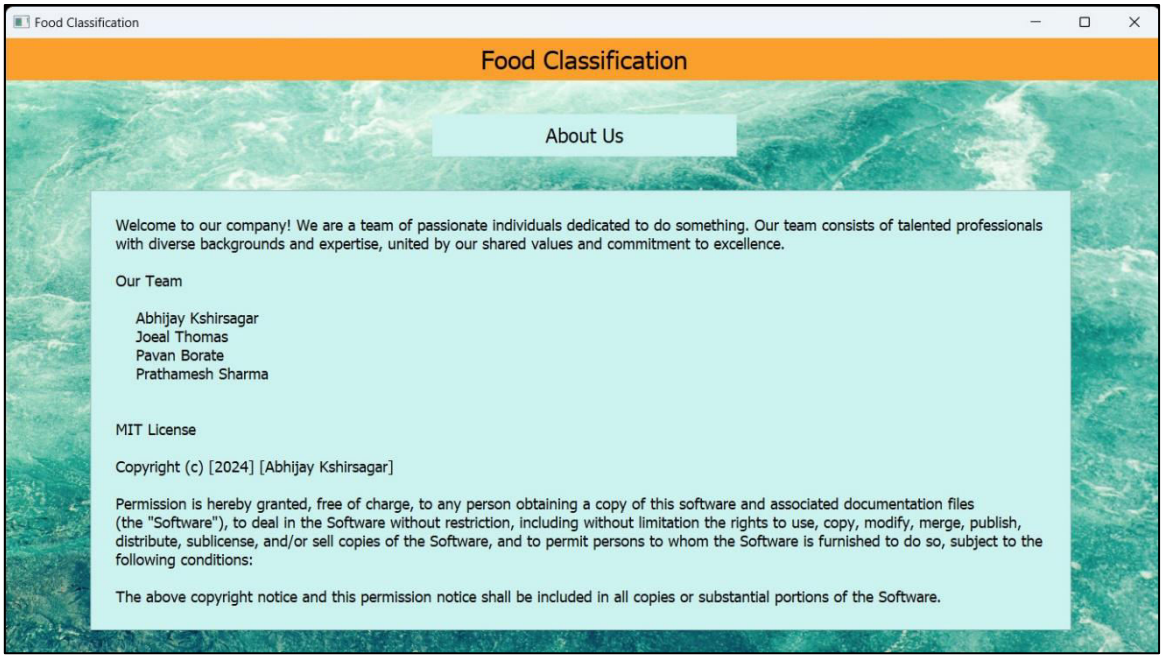


**Figure 5 About Us page**

## V. CONCLUSION

In summary, the Vision Transformer (ViT) model for Indian food image classification has been highly successful. The ViT model demonstrated exceptional accuracy in identifying diverse Indian dishes, and its integration with recipe generation functionality enhanced user experience by providing accurate recipe recommendations. This project

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204321 |

showcases the effectiveness of ViT models in visual recognition tasks and their potential to improve real-world applications like food recommendation systems. Future work could focus on dataset expansion and fine-tuning model hyperparameters for further performance enhancements.

## REFERENCES

1. “Vision Transformer (ViT)-based Applications in Image Classification”.
2. “Scaling Vision Transformers”: Xiaohua Zhai , Alexander Kolesnikov , Neil Houlsby, Lucas Beyer, Google Research, Brain Team, Zürich.
3. "Scalable Image Classification using Vision Transformers: Integrating Google's ViT into Cloud-based Pipelines" - 2022
4. "Improving Image Classification Accuracy with Vision Transformer Models: Experiments with Google's ViT" - 2021
5. "Vision Transformer-based Image Classification in Resource-constrained Environments: Insights from Google's ViT" - 2023
6. "Hybrid Approaches for Image Classification: Integrating Convolutional Neural Networks and Vision Transformers with Google's ViT" - 2022
7. "Efficient Training Strategies for Vision Transformer Models in Image Classification: Case Study with Google's ViT" - 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details