AI-BASED FOOD RECOGNITION FOR THE VISUALLY IMPAIRED

Nikhil M[1] Harshita S H[2] Venugeetha Y[3]

Don Bosco Institute of Technology,Samsung Innovation Campus

Bengaluru,vgydbit@gmail.com

***Abstract*— This project presents an innovative food item classification system designed to assist visually challenged individuals in identifying various types of foods. The system utilizes the capabilities of deep learning through the implementation of two distinct approaches: a custom convolutional neural network (CNN) and the InceptionV3 pre-trained model. The integration of InceptionV3 results in a notable improvement in classification accuracy and loss. One of the key features of this system is its accessibility component, which provides auditory output for visually challenged users.**

I. INTRODUCTION

In this paper we present an "AI-based Food Recognition for Visually impaired" project is a cutting-edge application of technology designed to assist individuals with visual impairments in identifying and recognizing different types of food. The main objective is to provide real-time audio descriptions of food items, empowering visually challenged individuals to make informed choices about their meals without relying on sight.

In this innovative project, computer vision and machine learning technologies are harnessed to process images of food and convert the results into audio descriptions. By uploading food images, users can access detailed audio information, enhancing their dining experiences and aiding those with visual impairments .This technology not only enhances the independence of visually impaired individuals in daily food-related activities but also promotes healthier dietary choices and inclusivity, ensuring they can participate more fully in social and culinary experiences.

In a world that heavily relies on visual cues for food selection and presentation, this project serves as a significant breakthrough for the visually challenged community. With its real-time audio-based food recognition, it not only offers practical assistance but also fosters a sense of empowerment and autonomy. The system's core features include seamless image capture, robust image preprocessing, and a powerful food recognition model based on deep learning. This model is capable of identifying a wide range of foods, spanning various cuisines . The recognized food items are then transformed into clear and articulate audio descriptions using text-to-speech technology.

This system is designed to bridge the accessibility gap for visually impaired individuals and provide a seamless, user-friendly experience. By analyzing food images, our platform offers audio descriptions, making it easier for users to identify and appreciate different dishes while dining out or ordering takeout.

The implementation of this computer vision and machine learning technologies are harnessed to process images of food and convert the results into audio descriptions. By uploading food images, users can access detailed audio information, enhancing their dining experiences and aiding those with visual impairments .This technology not only enhances the independence of visually impaired individuals in daily food-related activities but also promotes healthier dietary choices and inclusivity, ensuring they can participate more fully in social and culinary experiences. In this paper, This technology not only promotes inclusivity but also enhances the overall dining experience for everyone, ensuring that food choices are informed and enjoyable.

II. LITERATURE SURVEY

[1]. In April 2021, Dr. Gurbakash Phonsa presented a paper on Content-Based Image Retrieval (CBIR) using algorithms to retrieve images from a database based on features. The paper demonstrates image classification using a Convolutional Neural Network (CNN) on the cifar-10 dataset, achieving 94% accuracy for the classes aeroplane, bird, and car with a batch size of 64 in a Jupyter notebook.

[2]. In February 2019, Mingyuan Xin and Yong Wang presented research on an image classification model based on deep convolutional neural networks. They introduced an innovative training criterion for maximum interval minimum classification error, combining cross entropy and M3CE for improved results. Their proposed model, M3 CE-CEc, was tested on MNIST and CIFAR-10 databases, demonstrating its effectiveness as an enhancement to cross-entropy, achieving good results in both datasets.

[3]. In October 2018, Neha Sharma, Vibhor Jain, and Anju Mishra conducted an empirical analysis of popular Convolutional Neural Networks (CNNs) like AlexNet, GoogLeNet, and ResNet50 for real-time object identification in video feeds. The study focused on assessing the performance of these CNNs, widely recognized for object detection and classification in images, using various image datasets to evaluate their effectiveness

[4]. In the research conducted by Lihua Luo in August 2021, the significance of pictures as carriers of substantial information is highlighted. In scenarios involving a vast number of images, the timely extraction of valuable information becomes paramount. Consequently, the efficacy of image classification algorithms plays a pivotal role in determining classification outcomes. Image classification involves inputting an image and employing a specific classification algorithm to ascertain its category. The primary stages of image classification encompass image preprocessing, image feature extraction, and the design of a classifier. In contrast to the manual feature extraction processes in traditional machine learning, Convolutional Neural Networks (CNNs) within the deep learning paradigm offer a distinctive advantage. They possess the ability to automatically extract local features and share weights,

streamlining the process and contributing to enhanced performance in image classification tasks .

[5]. In September 2018, Farhana Sultana, Abu Sufian, and Paramartha Dutta presented a paper on "Advancements in Image Classification using Convolutional Neural Network." The paper discusses the state-of-the-art role of Convolutional Neural Networks (CNNs) in image classification, providing insights into various CNN components. The authors explore different CNN architectures, ranging from LeNet-5 to the latest SENet model, detailing each model's description and training specifics. The paper includes a comprehensive comparison among these models, showcasing the advancements in CNN technology.

III. METHODOLOGY

The project aims to provide text-to-speech functionality enhances the overall accessibility of the system, making it a valuable tool for visually challenged individuals seeking food recognition autonomy.

The key features of the project are as follows

A. Data Collection:

- The algorithm begins with a dataset of food-related images, including a diverse range of food items from the 101 specified food classes.

B. Feature Extraction:

- Convolutional Neural Networks (CNNs) are often used for feature extraction in image recognition tasks. The algorithm employs Custom model which uses CNN architecture as a feature extractor. Visualization.

C. Prediction:

- With a trained model, the algorithm can make predictions on new, unseen food images.

D. Real-Time Recognition (if applicable):

- If designed for real-time recognition, the algorithm continuously captures or accesses food images and feeds them to the model for immediate classification. This mode enables dynamic recognition of food items as they are presented in real life.

IV. DATASET

The food recognition dataset is dynamically retrieved from a live image feed or online database based on user input. Unlike conventional datasets that are sourced from static files, this dataset relies on real-time image capture or access to an online repository of food-related images. The system leverages specialized libraries and APIs to capture or access images dynamically and conduct food recognition tasks in realtime. It continuously updates the dataset as new images become available, ensuring that it reflects the latest and most relevant visual data for food recognition. This dynamic nature of the dataset allows for more accurate and up-to-date recognition of various food items and culinary creations.

The attributes in each dataset is as follows

- Images: The heart of the dataset comprises a vast collection of images, each depicting various food items. These images vary in terms of content, composition, quality and resolution close.
- Metadata: Metadata contains supplemental information associated with each image in the dataset.

Key elements of metadata may include:

  - Labels or Food Classes: Metadata should specify which of the 101 food classes each image belongs to.
  - Image File Information: Metadata can include details about the file, such as file paths, image file formats.

- Training Data: The training data subset is dedicated to training machine learning models. It includes a substantial portion of the dataset's images paired with their corresponding food class labels.
- Test Data: The test data subset is reserved for evaluating your trained models. It contains images that the model hasn't seen during training.
- Class Distribution: To comprehend the dataset thoroughly, it's vital to examine the distribution of images across the 101 food classes.

V. LIBRARIES

The Libraries used for the project is as follows:

- Streamlit : Streamlit is a Python library used to create web applications for data science and machine learning. We chose Streamlit for our project because it offers a user-friendly and efficient way to build data-driven web applications with minimal coding effort. Streamlit's simplicity, rapid prototyping capabilities, and excellent support for data visualization made it the ideal choice for our project.
- gTTS (Google Text-to-Speech) or pyttsx3:gTTS is a Python library and CLI tool that interfaces with Google Translate's text-to-speech API. pyttsx3 is a cross-platform text-to-speech conversion library. With gTTS, you can easily add speech functionality to your Python scripts, allowing you to create audio files from any text.
- Pillow (PIL): Pillow is a Python Imaging Library that allows you to open, manipulate, and save image files. It's useful for image processing and manipulation. incorporates lightweight image processing tools that aids in editing, creating and saving images.
- TensorFlow : TensorFlow is an open-source machine learning framework developed by Google that is widely used for building and training machine learning and deep learning models. TensorFlow is an open-source machine

learning framework developed by Google that is widely used for building and training machine learning and deep learning models.

- Matplotlib: Matplotlib is a popular Python library for creating static, animated, or interactive visualizations. It enables us to create a wide range of charts, graphs, and plots to effectively communicate data insights.
- NumPy: NumPy, short for "Numerical Python," is a fundamental library for numerical computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with an extensive collection of mathematical functions to operate on these arrays.

VI. RESULTS

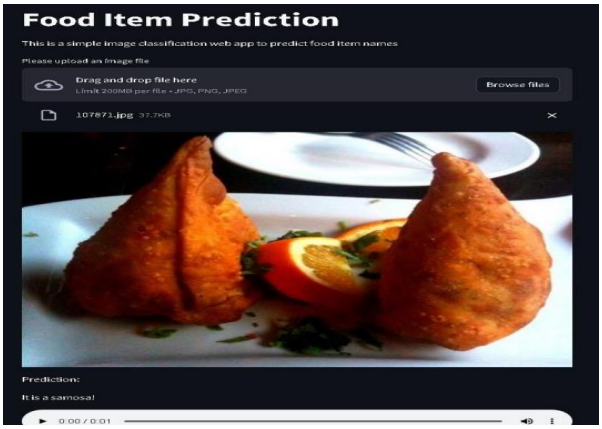

Figure 1-Image of Samosa.

Figure 1 Shows the image as SAMOSA and the audio option will convert the text ‘samosa’ to speechformula.

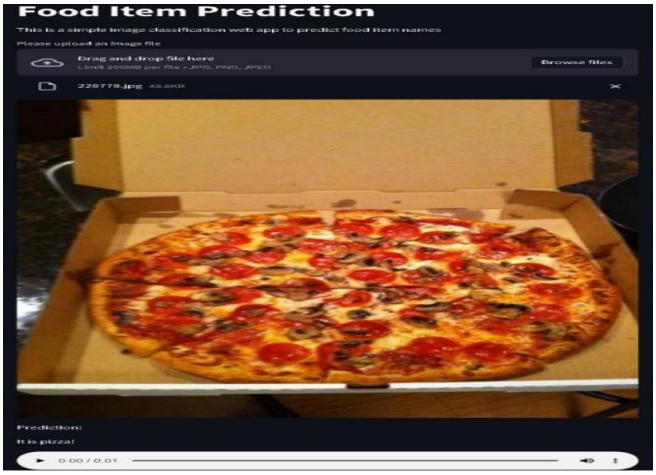

Figure 2.Image of Pizza.

Figure 2 Shows the image as PIZZA and the audio option will convert the text ‘Pizza’ to speech.

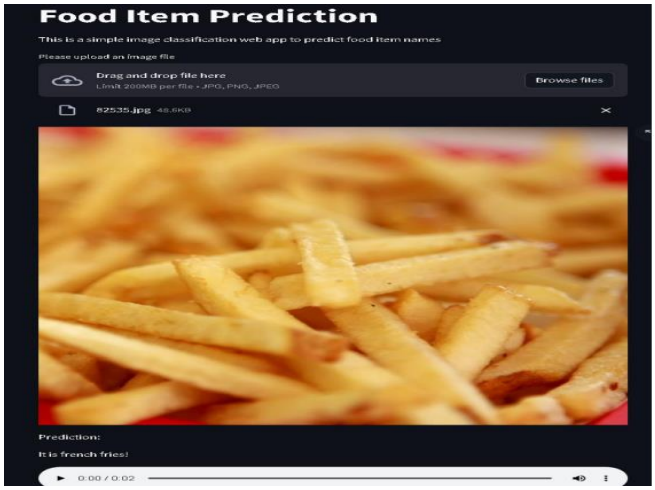

Figure 3-Image of French Fries

Figure 3 Shows the image as FRENCH FRIES and the audio option will convert the text ‘French fries’ to speech.

REFERENCES

[1] A. Sharma and G. Phonsa, "Image Classification Using CNN," Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021, 27 Apr 2021.

[2] Mingyuan Xin & Yong Wang, "Research on image classification model based on deep convolution neural network," EURASIP Journal on Image and Video Processing, volume 2019, Article number: 40 (2019).

[3] Neha Sharma, Vibhor Jain, Anju Mishra, "An Analysis Of Convolutional Neural Networks For Image Classification," Procedia Computer Science, Volume 132, 2018, Pages 1571-1577. doi: 10.1016/j.procs.2018.05.198. This paper is available under a Creative Commons license.

[4] Farhana Sultana, Abu Sufian, Paramartha Dutta, "Advancements in Image Classification using Convolutional Neural Network," Publisher: IEEE.

[5] Hitesh Kumar Sharma, et al., "Deep Learning based Binary lassification of Paintings and Photographs using CNN Model," University of Petroleum & Energy Studies, India.

[6] arhana Sultana, Abu Sufian, Paramartha Dutta, "Advancements in Image Classification using Convolutional Neural Network," Publisher: IEEE.