Abstract

People today pay close attention to their diet and overall wellness. Consuming food items with a lot of calories can be unhealthy and can contribute to major medical disorders like obesity, heart disease, chronic diseases, and hypertension. The essential step for maintaining a successful healthy diet is to keep track of the type of food and amount of food taken. Accurate identification of food is very important as based on which calories can be counted and that helps to reduce the risk of serious health conditions. This research work is mainly concentrated on the identification of food accurately.

People in Gujarat are very fond of eating different varieties of food. Gujarati food is very vivid in taste and texture due to its rich heritage and history. As most Gujarati food is oily and sweet, consuming this food daily in larger portions results in various health problems. A lot of work has been done on other kinds of food but no previous efforts have been made to classify Gujarati food. Therefore, being Gujarati, this study focuses on the classification of Gujarati food images.

In this research, a new dataset has been created and it named as "Traditional Gujarati Food Images Dataset (TGFD)". Initially dataset has been created with five popular Gujarati food items namely Dhokla, Handvo, Khakhra, Khandvi, and Patra consisting of 1764 images. The dataset is created by the images collected from the internet, pictures taken using mobile phones, and real images captured by visiting different restaurants. The total number of images after applying data augmentation techniques is 35,280.

Since the images collected in this way may contain noise and in order to remove noise and improve the quality of the images collected in TGFD, an algorithm was implemented which named as Improved Selective Median Filter (ISMF). The output of this algorithm has been used further to process food images in order to classify them correctly.

Simulation has been done on the existing five models namely VGG16, VGG19, Resnet50, Inceptionv3 and Alexnet to analyze their performance on Gujarati food images.Training these models on TGFD, the highest accuracy of 81.58% is obtained with Inceptionv3.

To further improve the classification accuracy, Transfer Learning has been implemented on all five models. To do so more precision fine-tuning has been applied to all five models. The results of the simulation, Transfer learning, and Fine-tuning on all the models have been compared in terms of accuracy.

From the simulation and empirical analysis, it has been observed that the factors that directly impact the model's accuracy include the number of convolutional layers, the number of neurons in fully connected layers, the number of filters, and the size of the filters. Taking these parameters into consideration in this research a model for Gujarati food image classification has been developed and named "Depth Restricted Convolutional Neural Network (DRCNN)". The classification accuracy achieved with DRCNN for TGFD is 95.48% which is 14% higher than the highest accuracy achieved with the existing model. The proposed model size is 48 times smaller than the Inception v3 model in terms of parameters and hence takes half the time to run compared to other pre-built models with millions of parameters.

The performance of DRCNN has been tested by increasing Gujarati food image classes. TGFD initially consists of 5 classes and then increased up to 20 Gujarati Food Classes. The performance of DRCNN has improved with a large number of images showing the versatility of the model. In the second case, DRCNN is tested against different food datasets and it gives outstanding accuracy for any type of food dataset.

The computational complexity of the model has been derived by considering the factors that affect the CNN model's performance. The time it takes for each layer to run has also been taken into consideration to see its effect on the model's overall performance. It has been found from the results that 90% of computational time is taken by convolutional layers and 5-10% of the time is taken by fully connected layers.

The research done consists of basically designing and implementing the DRCNN model with improved preprocessing, better output and efficient time complexity. The work has been published in reputed journals and conference proceedings.