Chapter 1
INTRODUCTION

This chapter briefly introduces food image classification and deep learning followed by types of deep learning techniques and their comparisons. Next, the discussion is on the motivation for this work, the research gap as observed and based on it the problem statement, objectives and scope. Research contribution of the work done and research conducted is also discussed.

1.1 OVERVIEW OF FOOD IMAGE CLASSIFICATION

Food is an essential element of everyone's life. With improved lifestyles, people nowadays are more focused on a healthy diet. People are becoming more health conscious as the globe becomes more competitive. The problem of being overweight is developing significantly nowadays due to the imbalance of calorie intake and use. It is the primary cause of serious health problems such as hypertension, diabetes, chronic illnesses, and heart disease.

According to World Health Organization in 2020, an estimated 39 million children under the age of 5 years were overweight or obese. In 2016, 39% of adults aged 18 years and over (39% men and 40% women) were overweight. Most of the world's population lives in countries where overweight and obese kill more people than underweight.

Serious health issues because of obesity can be avoided by limiting the amount and type of food consumed. Accurate methods for food identification and calorie estimation can help people to fight against obesity which is the cause of being overweight and other medical conditions. Recognition of food is the first step to a successful healthy diet and to recognize food items accurately, Image Processing can be used.

Automatic recognition of food helps people in the decision making process for the type of food eaten, calorie estimation and diet monitoring system. Classification of food images is a very challenging task as the dataset of food images is not linear. The diversity of food makes it difficult to recognize the complex features using traditional approaches and hence makes food image classification more challenging [2].

According to research, Deep learning methods for image classification give more accurate and efficient results as compared to traditional image processing methods. Many works have been done to classify different types of food like Western food, Japanese food, Fast food items, Chinese food and South Indian food using deep learning [3] [5-6] [7-9][21].

1.2 OVERVIEW OF DEEP LEARNING

Deep learning is a subset of machine learning. Unlike Machine Learning, in deep learning, basic details about the data need to be given, that process through many layers and the computer trains to recognize the patterns on its own. Deep learning is an advanced technology for image processing, speech recognition, object detection and food science and engineering [1]. It works with artificial neural networks, which is designed to imitate how humans think and learn. The availability of a large number of datasets and the availability of high Processing GPU makes deep learning techniques very successful [3].

Deep learning techniques can be divided into mainly three categories. 1. Supervised learning 2. Unsupervised learning 3. Reinforcement learning.

- **Supervised Learning:** In this type of learning, the machine learns patterns by using labeled data. It uses a training dataset and based on it makes a model and use it to do predictions about the output. A model learns from a set of labeled examples to make predictions or classify new, unseen data[1]. It can be divided into Classification & Regression. The networks in supervised learning are Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Recursive Neural Networks (RvNN). Supervised learning can be helpful in areas such as Image Recognition, speech recognition and forecasting [2].
- **Unsupervised Learning:** In this type of learning the machine is trained by unlabeled data. It just uses the input dataset and makes an analysis about the output. It can be divided into clustering, density estimation and dimensionality reduction. It uses the approach to understand patterns and discover output [45]. The networks in unsupervised learning are Deep Belief Network (DBN), Deep Boltzmann Machines (DBM), Generative Adversarial Network (GAN) and

Variational Autoencoder (VAE). Unsupervised learning can be helpful in areas such as pre-processing the data and pre-train supervised learning algorithms.

- **Reinforcement Learning:** In this type of learning there is no training dataset, instead it bound to learn from its experience. There is no predefined data available. It uses consequences or behavior to make predictions about the output. It uses the approach of following the trial-and-error methods. The network in Reinforcement learning is Markov Decision process learning. This type of learning can be helpful in areas such as warehouses, Inventory management, aircraft control, robot motion control and chess games [1,10].

If all available data in the dataset are labeled, then supervised learning gives the best results for object classification. In this research work, the images in the dataset are labeled and hence, supervised learning has been used [2].

Deep learning networks are mathematical models that work like human brains. This mathematical model is created as a neural network consisting of neurons. The neural network is divided into three major layers that are input layer (the first layer of the neural network), the hidden layer (all middle layers of the neural network) and the output layer (the last layer of the neural network). A layer in the neural network is nothing but a collection of neurons that takes an input and provides an output. The input of each of these neurons is processed through the activation function assigned to the neurons. Some of the most popular deep learning networks for supervised learning are CNN, RNN and RvNN as described as follows [1-3][45].

1.2.1 CNN: Convolutional Neural Network

CNN is the main model of Deep Learning networks to do image recognition and image classification. It takes input, processes and classifies it under predefined categories. It gives the best results in detecting patterns and features from images. It is robust to rotation, translation and scaling variance. There is no need for manual feature extraction in CNN and it can handle large amount of data with high accuracy. The main advantages of CNN are parameter sharing, sparse interactions and equivalent representations [2]. Many successful research works have been done on food object recognition through CNN proving that CNN gives the best result in terms of accuracy and error rate for object recognition [26-32].

1.2.2 RNN: Recurrent Neural Network

A very popular deep-learning model that uses recursion techniques to build models. RNN saves the output of the current layer which will be input to the next layer. Due to its internal memory, it can memorize previous inputs and hence it is especially used as a language model. As it is very necessary to memorize the previous words while reading a sentence, RNN is mostly used in natural language processing (NLP), speech processing [18, 23], text analysis and machine translation [28][39].

1.2.3 RvNN: Recursive Neural Network

RvNN can handle the inputs of different modalities [18] and has been especially successful in NLP. RvNN separates the image into different segments and forms a syntactic tree [40]. Due to the tree structure, it can manage hierarchical data, unlike RNN which can manage only sequential data. It can be best used in parsing problems.

As RvNN does not suit for image classification due to its tree structure, most of the research work has used CNN and RNN for image classification [26–30] [41-43]. To further decide which one will be suited for image classification, a comparative study of CNN & RNN has been done and it is shown in Table 1.1.

Table 1.1 Comparison between CNN and RNN

CNN	RNN
It is suitable for spatial data such as an image. Popularly used for image recognition, can be extended for NLP, speech processing and computer vision	It is suitable for temporal data, also called sequential data. Widely used for NLP and speech processing
Many research work has been done using CNN for image classification.	It can be used for image classification but theoretically, only a few research about RNN image classifier can be found.
CNN's are special for video processing and image processing.	RNN works primarily on time series information on the past influence of the customer.
The interconnection consumes a finite set of inputs and generates a finite set of outputs according to the input.	RNN can allow arbitrary input length and output length.
Considered more powerful than RNN	Less powerful than CNN

According to studies and considering the comparison in Table 1.1, It has been found that among all the deep learning networks, CNN is most suited for image classification and gives the best classification accuracy by reducing the error rate for object recognition. Hence, this research work has used CNN to classify Gujarati Food items accurately.

There are mainly two ways to classify images through CNN: 1) use the concept of transfer learning and 2) train CNN from scratch. It is not good practice to develop a CNN from scratch for smaller datasets. In such cases, transfer learning is a good solution.

1.2.4 Transfer Learning

Transfer Learning is a deep learning technique where a developed model for one task can be used for another similar task. The benefit of transfer learning is that instead of developing and training everything from scratch, the weight of the pre-trained model can be freezed and only the custom layers need to be retrained [4,5]. It can save time and is a good solution to the problem of having smaller datasets [44-45].

Transfer learning can be divided into two parts. 1) make use of a pre-trained model or 2) the model can be fine-tuned with a suitable pair of hyperparameters. In the pre-training of the model, only the classification layer needs to be retrained, keeping the feature extraction layer untouched. Hence, weights need to be changed according to application or dataset.

1.2.5 Fine-Tuning

Fine-tuning allows the feature extraction part of the model along with the classification part to be changed [4]. It is possible to fine-tune all the layers, but it is always good practice to fix the upper layers as they contain features that are more generic in nature [5].

In this technique, some of the final layers of the freezed model are initially unfreezed, by adding some additional layers, followed by training of both the unfreezed layers and additional custom layers at once.

In this research work transfer learning and fine-tuning have been implemented on some of the popular predefined CNN architectures to see how they perform on the proposed dataset.

1.3 PROBLEM DESCRIPTION

Most Gujarati Food is oily and sweet. Consuming this food daily in larger portions results in obesity and other disease affecting the healthy life of people. There are so many varieties of the same Gujarati food from common ingredients which look alike and varied by different cooking methods, texture and presentation. Also, Gujarati Food items do not have regular shapes in general. Hence it is a very crucial aspect to identify Gujarati food correctly.

Much work on food classification on different types of food datasets has been done using deep learning or CNN. CNNs can analyse the visual aspects of food images and learn to differentiate between different types of foods. As per the literature survey, it has been observed that many authors have tried to classify different types of food by either developing a CNN model from scratch or by applying transfer learning or fine tuning on Pre-built CNN model. People have proposed a model to classify Chinese, Thai, Japanese, Bengali, Malaysian and American fast-food items [106-119].

Image preprocessing is needed to improve the quality and usability of images. By removing noise, standardizing size and orientation, and enhancing contrast, image preprocessing enables more accurate and efficient analysis and interpretation of visual data. For food images, preserving the edges during image preprocessing is crucial because edges play a significant role in detecting and delineating the shape and boundaries of food items. According to literature survey, many techniques have been proposed to remove impulse noise from images but still there is further scope of improvement [47-64].

Understanding the time complexity helps assess the computational efficiency of the CNN model. By analyzing the time complexity, one can evaluate the speed and efficiency of the model, making it easier to compare and select the most suitable architecture for a given task. According to the literature study, there have been extensive studies on the architecture, optimization, and performance evaluation of CNN but no specific investigation or publication has been found that focuses on quantifying the computational complexity of CNN models [125-138] .

1.3.1 Research Finding

After conducting a thorough literature survey the following observation have been made.

- A lot of work has been done on Chinese, Japanese and American fast-food items, but the essence of Gujarati food items is missing.

- There is no dataset available for Gujarati food items.

- It has been found from the studies that different techniques have been used for efficient noise removal. But most techniques have difficulty removing salt and pepper and impulse noise while preserving edges and contours.

- No work has been done to find the CNN model's time complexity.

1.4 MOTIVATION FOR THIS WORK

The traditional food of a country resembles some part of its culture, too. One can get an idea of a country's food habits and food culture by virtue of its traditional food processing and preservation method. The people of Gujarat are known for their hospitality and most of this fame is earned by the unique tests of the traditional foods here. Until now, more research on western-style foods has been conducted through deep learning and food classification. Through literature survey, it has been found that none of the work has been done till now for Gujarati Food Classification. This analysis has motivated to focus on classifying Gujarati Food Images through Deep Learning. This concept will open doors of acceptance for Gujarati Food or cuisine at a global level. With this work, the spread of true information about the Gujarati food image globally will be enhanced through the internet and social media. It will be helpful to food bloggers and foodies to understand Gujarati cuisine.

As a resident of Gujarat, considering Gujarati Food, inspired to propose a model which can classify Gujarati Food Images accurately with less amount of time.

1.5 PROBLEM STATEMENT, OBJECTIVES AND SCOPE

1.5.1 Problem Statement

The prime idea of this research is to design and develop a model for Gujarati Food image classification with improved accuracy and performance by making the network (CNN) lightweight.

1.5.2 Objectives

- To design and develop an algorithm to remove impulse noise while preserving edges and contours.
- To design and develop a model(s) that can recognize Gujarati Food.
- To make a model lighter in terms of memory and time both so as to be used for handheld devices.
- To design a model that can give better performance with a greater number of food classes or a dataset of different food items.
- To find time complexity of a CNN model and to evaluate the performance of the model based on various parameters such as optimizers, activation function, epochs, batch size and learning rate.

1.5.3 Scope

- This research work considers single food items.
- It includes the theoretical and empirical studies of existing models of Deep Learning.
- Detailed study of different Deep Learning frameworks.
- Understand and empirical study on image classification with different parameters such as the texture of food, color, size and different types of food items.
- Understanding research issues with possible improvements.
- Detailed study of Deep Learning models and compare the proposed models with existing models.

- To preserve or increase accuracy as compared to other existing models.

1.6 Research Contribution

The following are the list of research contributions.

- A new dataset named "Traditional Gujarati Food Images Dataset (TGFD)" has been created. The dataset contains 1764 images belonging to five famous Gujarati food items classes namely Dhokla, Handvo, Khakhra, Khandvi and Patra with at least 300 images per class in the dataset, which are divided into training, validation and testing with 70%, 20%, and 10% respectively.
- Pre-processing is essential for improving the quality of an image. An algorithm has been developed on the basis of a median filter named ISMF (Improved Selective Median Filter) for the removal of impulse noise in food images. This newly developed algorithm helps to reduce MSE which ultimately improves performance. ISMF performs better than the median filter in terms of detail preservation and image denoising.
- Transfer Learning has been implemented on models, namely VGG16, VGG19, Resnet50, Inceptionv3 and Alexnet. By implementing transfer learning the validation accuracy is increased by 5% as compared with the accuracy achieved by simulation on existing models. The highest accuracy achieved is 86.22% by the Inceptionv3 model. The model contains 256,005 trainable parameters.
- To further improve the classification accuracy Fine-tuning has been implemented on all models, namely VGG16, VGG19, Resnet50, Inceptionv3 and Alexnet. By implementing Fine-tuning, the validation accuracy is increased by 8% as compared with the accuracy achieved by simulation on existing models. The Inceptionv3 model achieved the best classification accuracy of 89.36%. The model contains 22,024,357 trainable parameters.
- A model has been developed from scratch for Gujarati Food Image Classification named as "Depth Restricted Convolutional Neural Network (DRCNN)". The parameters considered for the model development are the number of convolutional layers, the number of neurons in fully connected layers, the number of filters and the filter size, which directly affect the model's accuracy. The DRCNN

model achieves a remarkable classification accuracy of 95.48%, The performance of the model is measured in terms of F1 Score, Precision and Sensitivity along with accuracy.

- The proposed DRCNN model is tested in two ways to check its effectiveness and performance.
  - In the first test case, the DRCNN is run for a higher number of Gujarati food images starting from 10 up to 20 Gujarati food items.
  - DRCNN model is also run for different food datasets and it gives outstanding performance for any type of food item proving the versatility of the model.
- The Time Complexity of the Proposed Model and CNN Model has been derived and analysed. Eight different models were developed varying by the size of filters, number of convolutional layers, number of filters, number of fully connected layers and kernel size. The result shows that factors like an optimizer, batch size, filter and neurons greatly impact the time taken by the model. From this, it has been derived that the convolutional layers, max pool and fully connected layers directly affect the performance of the model. Since DRCNN has a minimum number of convolutions and fully connected layers the model's computational complexity is less than most of the existing models.
- The work has been published in SCIE indexed, Scopus indexed and UGC Care journals and conference proceedings.

As a concluding remark, the overall research work is shown below Fig. 1.1.

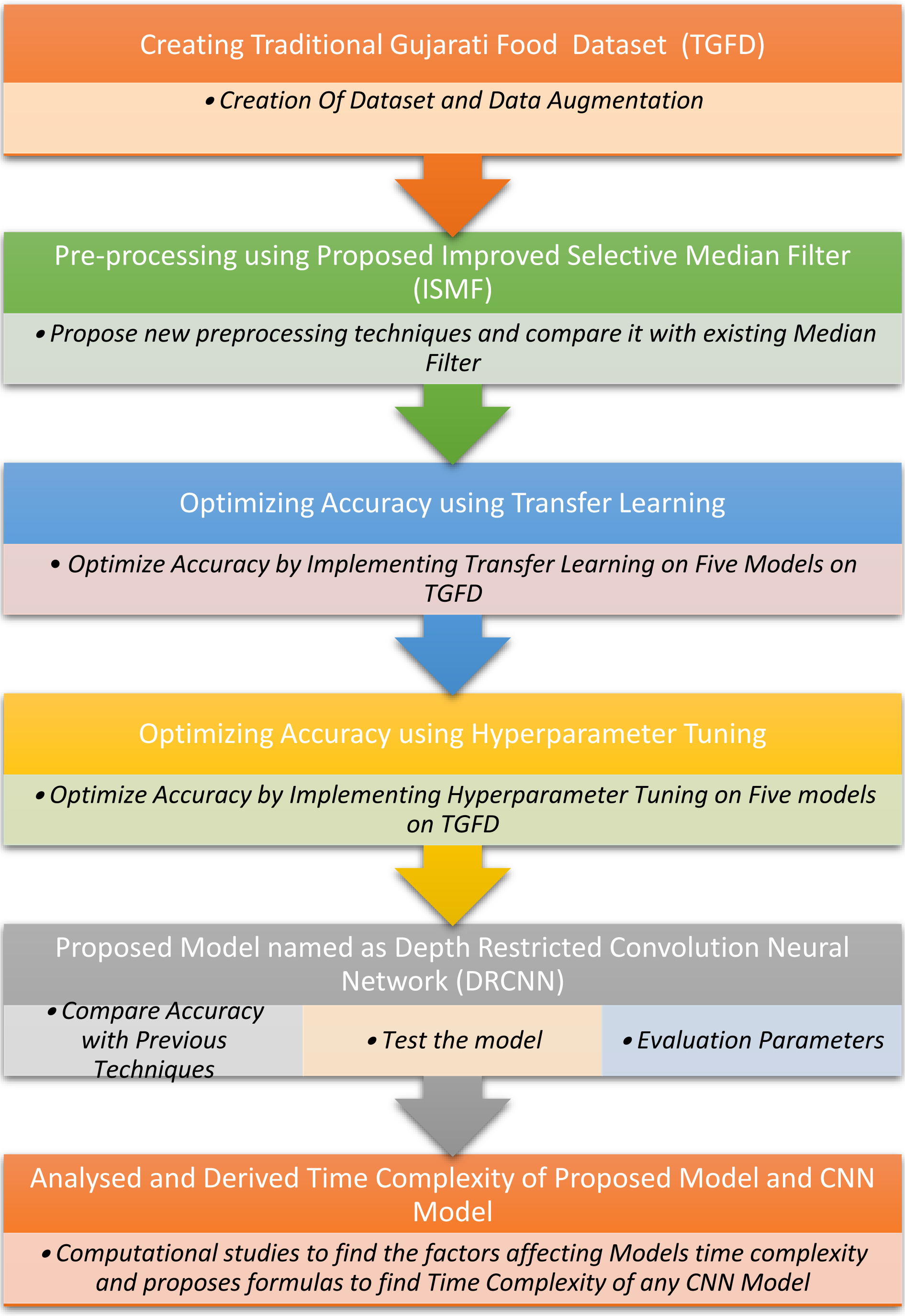

Fig 1.1 Research Flow

1.7 Thesis Organization

Chapter 2: This section presents a literature review. Different papers on preprocessing algorithms have been discussed here. Detailed studies on different techniques for food image classification have been discussed based on the approach used, the dataset used and the accuracy achieved. Further, a detailed investigation of different factors affecting on time complexities of the CNN model has been studied and analyzed .

Chapter 3: Gives introduction to different types of existing food dataset available. It also gives a detailed explanation on how the dataset has been created by collecting images from different sources. Data augmentation has been applied on images of the dataset has been discussed with its parameters.

Chapter 4: Describes the importance of preprocessing along with the existing median filter technique. It covers a new preprocessing algorithm named ISMF with a detailed analysis on noisy images.

Chapter 5: This chapter discusses simulation, transfer learning and fine-tuning technique. It also covers a comparison of accuracy achieved from transfer learning and fine-tuning.

Chapter 6: Describes the proposed model named "Depth restricted convolutional neural network". The hyperparameter selection for the proposed model has been discussed, which has been selected from the findings from empirical analysis. The results such as the accuracy of the model, number of parameters and number of layers have been compared with pre-built CNN models. Apart from accuracy, the results are also compared with Precision, Sensitivity and F1 Score. It also discusses performance analysis of model on extended TGFD with different types of food classes.

Chapter 7: This section describes the time complexity of the model and decides which are the crucial parameters for determining the time complexity of the model.

Chapter 8: Concludes the research carried out and discusses the future direction