A CONVOLUTION NEURAL NETWORK BASED CLASSIFICATION APPROACH FOR RECOGNIZING TRADITIONAL FOODS OF BANGLADESH FROM FOOD IMAGES

BY

ROMYULL ISLAM
ID: 152-15-5833
AND

NISHAT TASNIM
ID: 152-15-5709

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Shaon Bhatta Shuvo
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Itisha Nowrin
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

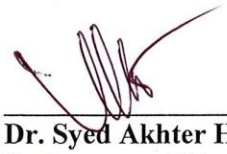
DHAKA, BANGLADESH

MAY 2019

APPROVAL

This Project titled **"A Convolution Neural Network based Classification Approach for Recognizing Traditional Foods of Bangladesh from Food Images"**, submitted by Md. Romyull Islam, ID No: 152-15-5833 and Nishat Tasnim, ID No: 152-15-5709 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on May 3, 2019.

BOARD OF EXAMINERS



Dr. Syed Akhter Hossain — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Tarek Habib — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

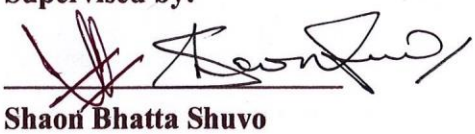
Moushumi Zaman Bonny — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Swakkhar Shatabda — **External Examiner**
Associate Professor
Department of Computer Science and Engineering
United International University

DECLARATION

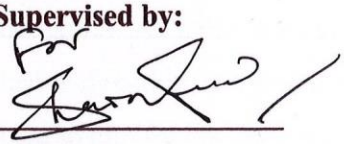
We hereby declare that, this thesis has been done by us under the supervision of **Mr. Shaon Bhatta Shuvo, Senior Lecturer, Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Shaon Bhatta Shuvo
Senior Lecturer
Department of CSE
Daffodil International University

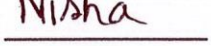
Co-Supervised by:



Itisha Nowrin
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Romyull Islam
ID: -152-15-5833
Department of CSE
Daffodil International University



Nishat Tasnim
ID: -152-15-5709
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First and foremost, admirations and thanks to the Almighty, for His showers of blessings all the way through our research work to accomplish the thesis successfully.

The work appears in its present form due to the support and assistance of numerous people. We would like to offer our earnest thanks to all of them.

At the very outset, we express our inmost sense of gratefulness to our esteemed supervisor **Shaon Bhatta Shuvo**, **Senior Lecturer**, Department of CSE Daffodil International University, our cordial thanks for his warm encouragement, constant and energetic supervision, insightful decision, critical comments and correction of the thesis. His special interest and knowledge in the field of "*Computer Vision*" enabled him to give us thoughtful guidance and provided us with much-needed motivation. Without his invaluable direction, this thesis would not have been achievable and we shall eternally be indebted to him for his guidance.

We take pride in admitting the insightful advice of **Dr. Syed Akhter Hossain, Professor and Head,** Department of CSE, for his valuable and constructive suggestions during the planning and development of this thesis. We owe our deep sense of gratitude to other faculty member and the staff of the CSE department of Daffodil International University.

Sincere thanks to all our friends at Daffodil International University for their kindness and moral support during the course work.

DEDICATION

We would like to dedicate this dissertation to our beloved parents who are the biggest source of our strength. We have to be grateful for their affection and care throughout our life. Thank you for giving us the stamina to reach for the stars and pursue our dreams.

ABSTRACT

Research has become one of the most talked buzzwords in the modern era due to knowledge development and practical improvement obtained by research papers. However, choosing the exact research topic from the vast ocean of knowledge fields for an unaware individual person is quite hard. Sometimes a beginner researcher/student cannot specify the research interest, or they are unfamiliar with trending topic and technology, which are most impactful in the near future. Therefore, we think computer vision is such an important thing that already helps the human being in a various way and it will help us in the future.

The process of identifying food items from an image is one of the promising applications of visual object recognition in computer vision. However, analysis of food items is a particularly challenging work due to the characteristic of their images, which is why a low classification accuracy has achieved by traditional approaches in the field. Deep neural networks have exceeded such solutions. With a goal to successfully applying computer vision techniques to classify food images founded on Inception-v3 model of TensorFlow platform, we conduct the transfer learning technique to retrain the food category datasets. Our methodology demonstrates auspicious outcomes with an average precision of nearly 95.2% in properly identifying among seven traditional Bangladeshi foods. This research paper intends to give details about various classification techniques and process by using computer vision methods that are being used in today's research for classifying food items.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBREVIATIONS	EXPLANATIONS
CNN	Convolutional Neural Network
ReLU	Rectified Linear Unit
SVM	Support Vector Machine
AAM	Active Appearance Model
ASM	Active Shape Model

CHAPTER 1

Introduction

1.1 Introduction

Traditional foods play a significant role in a country's culture, which reflects its unique history, identity, heritage, lifestyle, values, and beliefs, as well as helps to easily understand people's humor. Some traditional foods have geographical indications. Most of our traditional foods have evolved over centuries, and it is a gift from a generation to the next generation. Knowledge about the processing and preservation techniques of traditional foods has been known for many generations in a country. Traditional foods have historical precedence in a countries national dish. Regional cuisine or local cuisine refers to specific ingredients, and the location of the production. These traditional foods vary significantly from each other and use locally available spices, herbs, fruits, and vegetables. Almost every country has unique recipes to represent their respective traditional foods. Like other countries, Bangladesh also has its own tradition of foods. People of Bangladesh tend to identify themselves with their food.

1.2 Motivation

The process of identifying food items and the extraction of information from food image is quite interesting as well as a challenging issue. With the growth of computer and digital image processing tools, individuals started to search the technique of automatic food items recognition by computer. For this reason, we decided to do a research on this topic and we think this research knowledge will help us in future to make a better and accurate support system which will help our people and all over the world. The ultimate goal of our research is to develop a new strategy to recognize food items automatically by utilizing new techniques of Computer Vision and Machine Learning and improving accuracy. During this thesis, the key technological innovation is the deep learning-based food image recognition algorithms.

1.2 Rationale of the Study

In our thesis, we use the transfer learning method to retrain the Inception-v3 [1] model of Tensor Flow [2] on the dataset of 7 food items, traditionally known as Kalavuna, Khichuri, Murgir-Roast, Panta-Ilish, Sorse-Ilish, Gorur-Rezala, and Tehari.

1.3 Research Questions

How to choose the best topic that may suits best to one's ability and desire and how to involve students with new research trending topics?

1.4 Expected Output

From this research, universities and research institutions can discover the papers utmost related to their research plans or thesis. Hence, in search of the correct papers to read turn out to be an immensely essential segment of their educational lives and introducing with trending and upcoming research topic impacts on their thought. 'A Convolution Neural Network based Classification Approach for Recognizing Traditional Foods of Bangladesh from Food Images' advantage these individuals in serving the best-related papers and preserving their valuable time. In addition, it can explore more talked buzz, used the topic in recent times, which motivates them to work with new problems, finding new solutions, and help the students who are confused about their field of interest. In computer vision, there are lots of techniques and algorithms to extract the best knowledge and information from a huge number of data. However, when we use it for classification we need to find out the accurate result. For this research, we are using here one model for our data set and find out the best result but this is a continuous process and the prediction or result totally depends on the attribute and the data set. We prepared a dataset that is accessible by all and utilizing that dataset the model that is created will be an uncovered source. It will support researchers for further improvement. Researchers will enhance features in the future.

1.5 Report Layout

There are 5 chapters in this report. In this chapter, we have discussed the introduction of traditional foods, motivation, the rationale of the study, research questions and the outcome of the thesis. Later followed by the report layout.

The residual of the report arrangement is as per the following:

In chapter 2, we will discuss the background of our research topic. This chapter also deals with the literature review, research summary, scope of the problem and challenges.

In chapter 3, we will discuss the methodologies employed in our study.

In chapter 4, we will discuss the obtained results and discussion.

Finally, some future work scopes, conclusion and recommendations are explained in chapter 5.

Follows the 5 units and at the conclusion, there is a reference of the following resources we have stated in our research.

CHAPTER 2

Background

2.1 Introduction

We have used Inception-v3 [1] model of TensorFlow [2] platform and CNN [4]. TensorFlow is an accessible source library, established by Google Brain Team within Google's Machine Learning Intelligence analysis association, for numerical computation that makes machine learning simpler and faster. TensorFlow joins the computational algebra of compilation optimization techniques, making simple the calculation of many mathematical expressions where the problem is the time needed to accomplish the computation.

Inception-v3 is the 2015 repetition of Google's Inception structural design and a broadly used image identification network model that has been presented to obtain more than 78.1% accurateness on the ImageNet dataset. Inception is a remarkable structural design and it is the outcome of numerous cycles of test and error. The model is the climax of several concepts advanced by many researchers over the years. The original paper on which Inception-v3 is based on is "Rethinking the Inception Architecture for Computer Vision" by Szegedy, et al. [5].

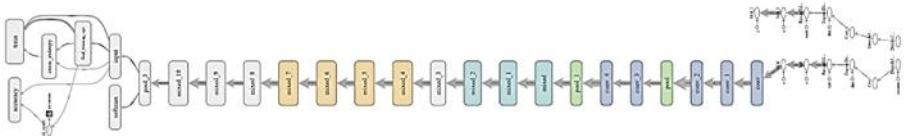


Figure 1: Main graph of Inception-v3 model from tensor board

Transfer learning [3] is the application of knowledge obtained from completing one task to help in solving a different but related problem. The development of algorithms that assist transfer learning processes has become a goal of machine learning technicians as they attempt to make machine learning as human-like as possible [3]. For example, the knowledge obtained by a machine learning algorithm to recognize trucks could later be transferred for use in other machine learning model being developed to recognize other types of vehicles, such as buses. In comparison with the customary neural network, it

merely needs to use a minor quantity of data to train the model, and attain high precision with short training time.

Convolutional neural networks (CNN) [4] are deep artificial neural network architectures, which are utilized to categorize images, bunch them by resemblance, and to accomplish object identification within scenes. A CNN consists of single or additional convolutional layers and then followed by single or additional completely associated layers as in a standard multilayer neural network [6]. It learns especially from images. A CNN can be prepared to do image analysis tasks including object recognition, segmentation, classification, and image processing.

Many types of layers are used to build ConvNet architectures like Convolutional Layer, Non-Linearity Layer, Rectification Layer, Rectified Linear Units (ReLU), Pooling Layer, Fully Connected Layer, Dropout Layer.

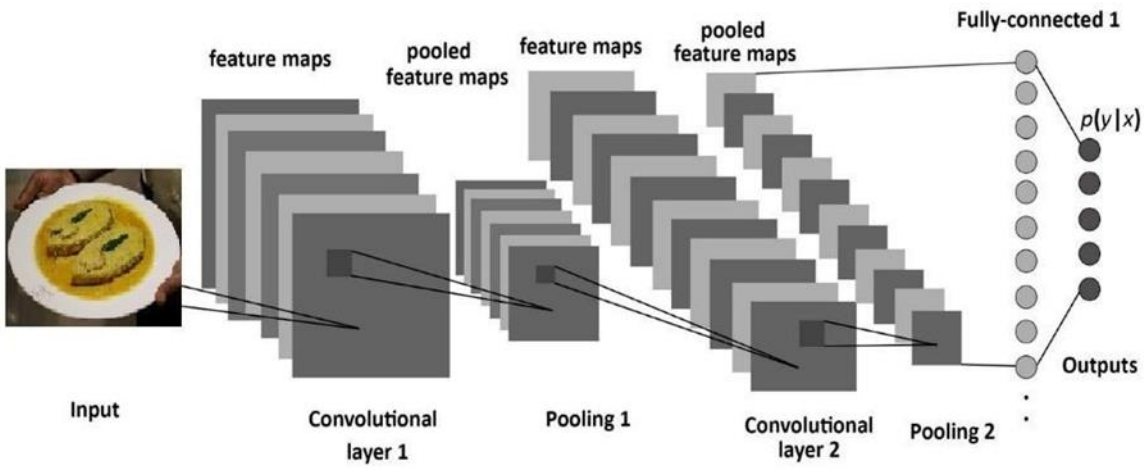


Figure 2: The Architecture of Convolutional Neural Network (CNN) [7]

The core task of the convolutional layer is to identify local conjunctions of features from the preceding layer and mapping their presence to a feature map. As a consequence of convolution in neuronal networks, the image is divided into perceptrons, making local capacious arenas and lastly compressing the perceptrons in feature maps of size $m_2 \times m_3$. Thus, this map reserves the info where the feature takes place in the image and how fine it matches the filter. Consequently, each filter is trained spatial concerning the location in the volume it is applied to [8].

In every layer, there is a bank of m_1 filters. The amount of how many filters are applied in a single phase is equal to the depth of the volume of output feature maps. Every filter identifies a specific feature at each position on the input [6]. The output $Y_i^{(l)}$ of layer l comprises of $m_1^{(l)}$ feature maps of size $m_2^{(l)} \times m_3^{(l)}$. The i^{th} feature map symbolized $Y_i^{(l)}$, is calculated as

$$Y_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K * Y_j^{(l-1)} \quad (1)$$

where $B_i^{(l)}$ is a bias matrix and $K_{i,j}^{(l)}$ is the filter of size $2h_1^{(l)} + 1 \times 2h_2^{(l)} + 1$ joining the j^{th} feature map in layer $(l-1)$ with i^{th} feature map in layer.

The pooling layer is accountable for decreasing the spatial size of the activation maps [8]. The pooling layer l has two hyperparameters, the spatial extent of the filter $F^{(l)}$ and the stride $S^{(l)}$. It takes an input volume of size $m_1^{(l-1)} \times m_2^{(l-1)} \times m_3^{(l-1)}$ and delivers an output volume of size $m_1^{(l)} \times m_2^{(l)} \times m_3^{(l)}$ where;

$$m_1^{(l)} = m_1^{(l-1)} \quad (2)$$

$$m_2^{(l)} = m_2^{(l-1)} - F^{(l)}) / S^{(l)} + 1 \quad (2)$$

$$m_3^{(l)} = (m_3^{(l-1)} - F^{(l)}) / S^{(l)} + 1 \quad (3)$$

The objective of the fully associated structure is to tune the weight parameters to generate a stochastic possible illustration of every class based on the activation maps created by the concatenation of convolutional, non-linearity, rectification and pooling layers [8].

If $l-1$ is a fully associated layer;

$$y_i^{(l)} = f\left(z_i^{(l)}\right) \text{ with } z_i^{(l)} = \sum_{j=1}^{m_1^{(l-1)}} w_{i,j}^{(l)} y_i^{(l-1)} \quad (5)$$

Else;

$$y_i^{(l)} = f(z_i^{(l)} \; with \; z_i^{(l)} = \sum_{j=1}^{m_1^{(l-1)}} \sum_{r=1}^{m_2^{(l-1)}} \sum_{s=1}^{m_3^{(l-1)}} w_{i,j,r,s}^{(l)} \left(Y_i^{(l-1)}\right) r, s \quad (4)$$

2.2 Related Works

A plethora of research and development efforts have been made in the area of computer vision over the last few years to ease the effort of automatic object recognition, among which food image recognition has recently gained much importance. However, none of the work has found recognizing traditional foods.

Inception-v3 [1] model is used in many types of research of different categories. One of the work by Xiaoling Xia and Cui Xu from College of Computer Science, Donghua University used the transfer learning method to retrain the Inception-v3 model of TensorFlow on the flower category datasets [9] of Oxford-I7 and Oxford-102 for Flower Classification in 2017. The classification precision of the model was 95% on Oxford-I7 flower dataset and 94% on Oxford-102 flower dataset [10].

Alwyn Mathew, Jimson Mathew et al. from bVuelogix Technologies Pvt Ltd utilized Google's TensorFlow is a deep learning framework to train, test and validate the network for Intrusion Detection in 2017 [11]. The precision was 95.3%. However, the recommended network is discovered to be tougher to train due to disappearing gradient and degradation issues. Brady Kieffer, Morteza Babaie et al. used CNN and Inception-v3 model for Histopathology Image Classification in 2017 [12]. All experiments were done on Kimia Path24 dataset. The precision was 56.98%.

Xiao-Ling Xia, Cui Xu et al. worked for Facial Expression Recognition based on the Inception-v3 model of TensorFlow platform in 2017. They used CK+ dataset [13] and selected 1004 images of facial expression. The precision was 97% but it was not based on dynamic sequences.

Bat-Erdene.B and Ganbat.Ts worked on Effective Computer Model for Recognizing Nationality from the Frontal Image in 2016 [14]. They used SVM [15], AAM [16], ASM [17]. The precision was 86.4%. The analysis was worked manually and images must be the frontal face image that has smooth lighting and does not have any rotation angle.

The aim of our research is to provide a suitable methodology for accurate automation of traditional food images classification as the first work of its kind.

2.3 Research Summary

For doing our research, we studied several articles, research paper, conference paper, and book. In this section, we discussed on the premises of other research and their output. As well as a discussion about the scope of the problem and challenges and an overview of the background. In computer vision, many researchers applied the same algorithms on the same dataset. Sometimes the same dataset and same algorithms can give different types of result. Therefore, it is very confusing. Many researchers used Inception-v3 model on distinct datasets.

Few authors did not provide enough information about their procedure on which algorithm they used. It creates a problematic situation for re-used or re-implement the algorithm. Subsequently, various implementations are used in the same procedure that might create differences in output. A small change in datasets, procedures or user inputs predictably creates a bigger change in the performance of the approaches. Therefore, determining the perfect approaches is an important task.

2.4 Scope of the Problem

When we were doing our research, we found that choosing the exact research field from the vast ocean of knowledge fields for an unaware individual person is quite hard. There are many fields when a researcher wants to do research. However, he/she could not choose most of the time. In addition, in research institutions, colleges and universities, professors, graduate students, and other researchers need to discover the papers utmost related to their research plans. In their educational lives, looking for the correct papers to read turn out to be an immensely essential segment. This research paper will advantage these individuals in serving the best-related papers and preserving their valuable time. There is a lot of related work. However, no works have taken enough relevant information about research papers into the recommendation. In addition, most of the researcher research on the same dataset.

2.5 Challenges

Automatic food classification from images is an exciting challenge for Computer Vision researchers with uses in diverse spheres. In specific, food classification is turning more and more significant for the crucial part that it plays in health economies. In our paper, we address the study of food image classification from the viewpoint of Computer Vision. The food recognition is a challenging task as the food renders high irregularity and an inherent deformability. The image illustration used to automatically comprehend food images plays the greatest significant role. To discover an appropriate illustration of food images it is essential to have illustrative datasets with a great diversity of dishes. Practically a research paper/field like classification system has to use a big dataset.

In our project, we have to process an enormous quantity of image data by hand and it is extremely challenging.

Numerous food datasets are comprised of images gathered through the Internet (e.g., downloaded from Social Networks) have generally a low resolution and have been processed by the operators with creative or improvement filters. The big challenges for this research are collecting data from the internet, food shop or kitchen. Most of the raw data is noisy. And there are also some unusable data.

After collecting the data, the main challenge is processing the data.

CHAPTER 3

Research Methodology

3.1 Introduction

Different country has distinctive traditional foods. Any kind of food classification is very tough because the computer system cannot identify the images easily. All researcher struggles with great force and applies a distinct technique to identify the food items with the computer system. In the recommended system, we applied a certain technique to interact with the computer system. This method identifies 700 food images.

Models are the most important fact for classification. Choosing the appropriate model is very difficult. In this chapter, we are going to describe the models used for this project in the implementation requirements part. We will also discuss the phases of our working process and data collection procedure. In addition, the chapter elaborates the statistical analysis. Research subject and instrumentation are also included in this chapter. The working procedure of the two models is shown with simple block diagrams.

3.2 Research Subject and Instrumentation

This experimentation is based on the Inception-v3 [1] model and MobileNet-v1 model of Tensor Flow [2] platform. It is the arena of the computer vision system.

Computer vision has some advantages-

- Simpler and faster processes
- Reliable
- Higher accuracy
- A wide range of use
- Reduction of costs

Efficiency and scalability of computer vision models-

They are-

- insanely small
- insanely fast
- remarkably accurate
- easy to tune for resources vs. accuracy

Tools or instrument that we use-

Up to now, we have explained the theoretical notions and procedures. Now a list of requirements of instruments are given below-

Hardware and software instruments-

- 2GHz Intel i3 processor
- 4GB memory
- 1600MHz DDR3
- 64-bit operating system
- x-64 based processor
- Canon 600 D (18/55 Lens)

Developing tools-

- Python 3
- Keras
- Scikit Learn
- Numpy
- OS
- Python OpenCV

3.3 Steps of Working Process

We used two retrain model named Inception-v3 and MobileNet-v1 in a specific way. Our models have applied certain process for execution.

The following diagrammatic representation explains the methodology of our proposed models:

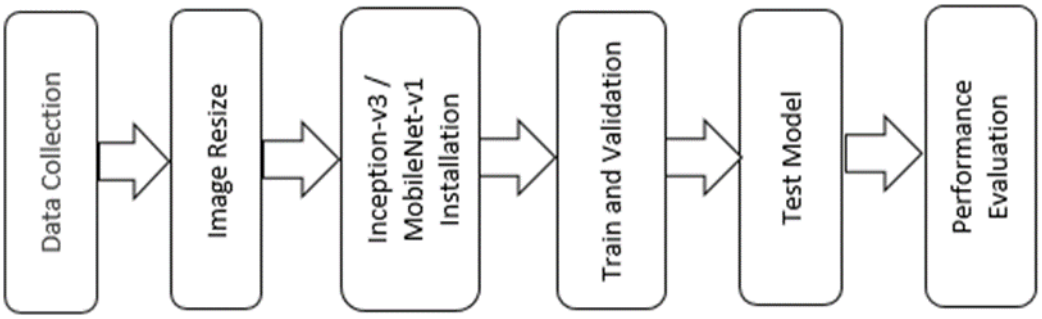


Figure 3: Flowchart of the proposed models

The whole working procedure is same for both the models except the third phase. For Inception-v3 and MobileNet-v1, we resized images 300x300px and 224x224px respectively and made separate dataset with same label.

3.4 Data Collection Procedure

Dataset

Bangladesh has plenty of traditional foods. Our dataset is composed of seven traditional food items. They are locally known as, Kalavuna, Khichuri, Murgir-Roast, Panta-Ilish, Sorse-Ilish, Gorur-Rezala, and Tehari. We collected the data from the internet, food shop and kitchen. The following images represent a portion of our dataset.

Kalavuna (10) Kalavuna (36) Kalavuna (34) Khichuri (2) Khichuri (6)

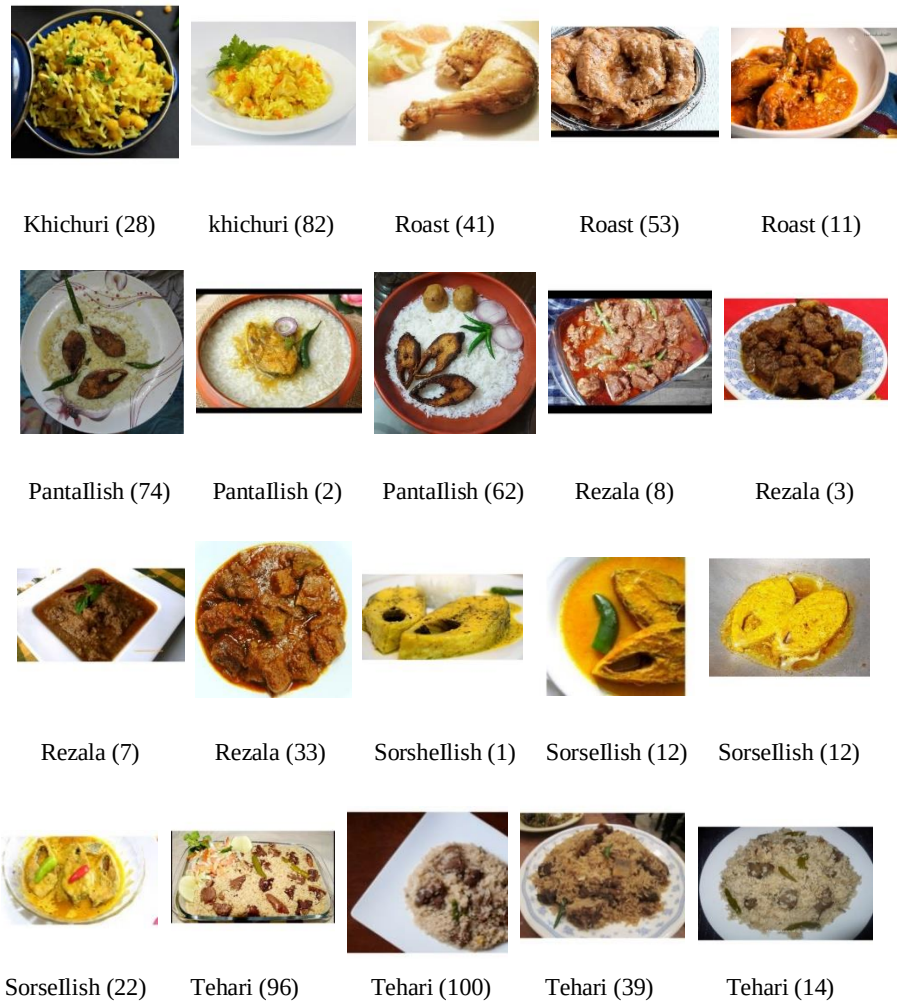


Figure 4: A portion of our dataset of traditional foods of Bangladesh

Data pre-processing

Image pre-processing is an important stage to promote the effect of image classification. The learning method of CNN directs the execution of our activity in machine learning, and that's why for the image pre-processing phase we have labeled and resized the images for training and testing from selected clear images. The entire data collection procedure has been separated into 5 distinct phases. All those phases are explained in Figure 5.

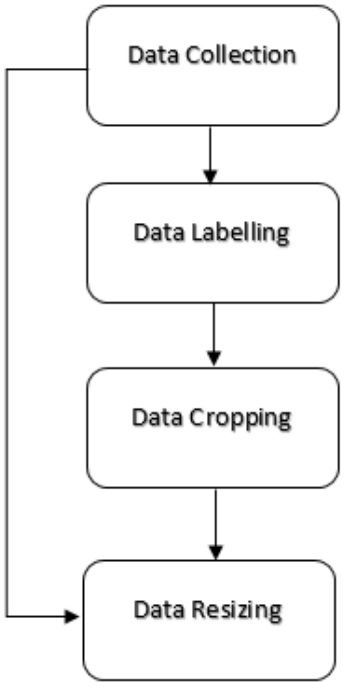


Figure 5: Flowchart of image preprocessing steps

3.5 Statistical Analysis

Statistical analysis is a component of data analytics. For recognizing different traditional food, we gathered 700 images of seven traditional foods. Each of the items contains 100 images. All images are kept in .jpg-formatted datatype.

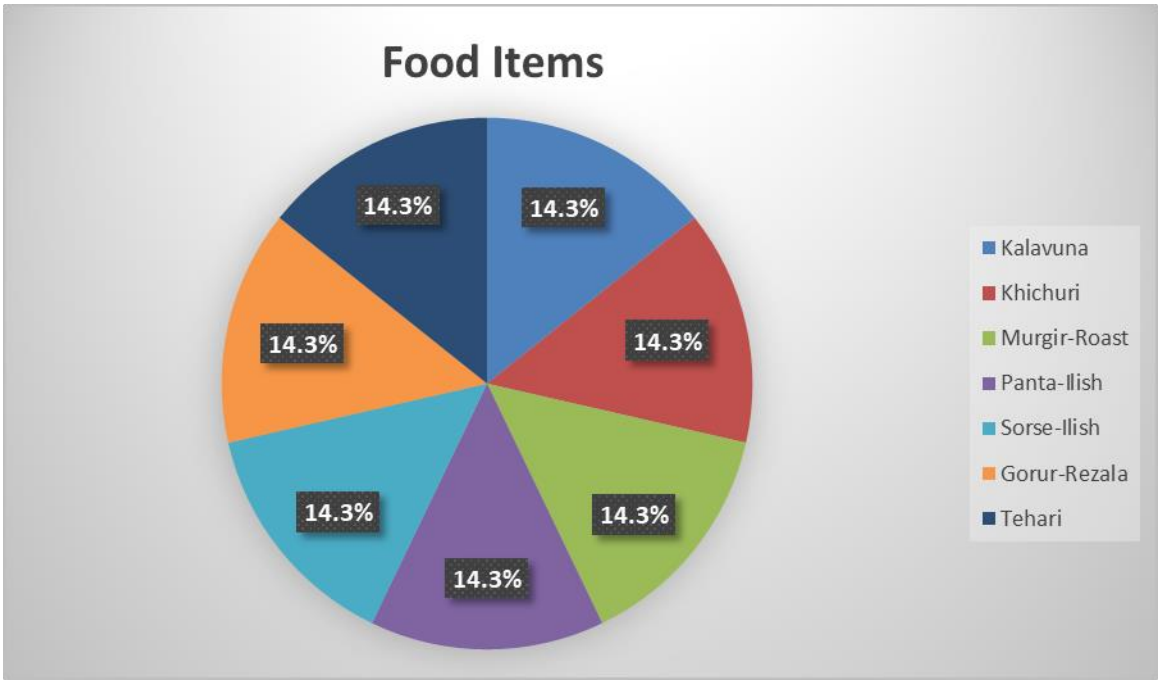


Figure 6: Image percentage per label

3.6 Implementation Requirements

Model installation

For the implementation, first, we have downloaded TensorFlow. Then, we have downloaded Inception-v3 model and Mobilenet-v1 model. We have also used the transfer learning technique that keeps the parameters of the earlier layer and have expelled the final layer of the Inception-v3 model and MobileNet-v1 model, then retrains a final layer.

Model Dissection

To build and run the models we need to use some Library and modules of "Python" with "Tensorflow" Framework such as "numpy", "hashlib", "OS", "datetime", "sys" etc. The use of those library and module on the python code are shown on the figure given below:

```
93  from __future__ import absolute_import
94  from __future__ import division
95  from __future__ import print_function
96
97  import argparse
98  import collections
99  from datetime import datetime
100 import hashlib
101 import os.path
102 import random
103 import re
104 import sys
105 import tarfile
106
107 import numpy as np
108 from six.moves import urllib
109 import tensorflow as tf
110
111 from tensorflow.python.framework import graph_util
112 from tensorflow.python.framework import tensor_shape
113 from tensorflow.python.platform import gfile
114 from tensorflow.python.util import compat
```

Figure 7: Used libraries and modules

To separate “training” “testing” and “validation” data from our dataset used a function named “create_image_list()”. It analyzes the subfolders of the data directory, parts them with stable training, validation and testing sets. It provides a data structure portraying the arrangement of images of every label and their paths in return.

Code of separating directories and Image formats for reading images are given below:

```
def create_image_lists(image_dir, testing_percentage, validation_percentage):
  if not gfile.Exists(image_dir):
    tf.logging.error("Image directory '" + image_dir + "' not found.")
    return None
  result = collections.OrderedDict()
  sub_dirs = [
    os.path.join(image_dir,item)
    for item in gfile.ListDirectory(image_dir)]
  sub_dirs = sorted(item for item in sub_dirs
                    if gfile.IsDirectory(item))
  for sub_dir in sub_dirs:
    extensions = ['jpg', 'jpeg', 'JPG', 'JPEG']
    file_list = []
    dir_name = os.path.basename(sub_dir)
    if dir_name == image_dir:
      continue
    tf.logging.info("Looking for images in '" + dir_name + "'")
    for extension in extensions:
      file_glob = os.path.join(image_dir, dir_name, '*.' + extension)
      file_list.extend(gfile.Glob(file_glob))
    if not file_list:
      tf.logging.warning('No files found')
      continue
    if len(file_list) < 20:
      tf.logging.warning('WARNING: Folder has less than 20 images, which may cause issues.')
    elif len(file_list) > MAX_NUM_IMAGES_PER_CLASS:
      tf.logging.warning('WARNING: Folder {} has more than {} images. Some images will '
          'never be selected.'.format(dir_name, MAX_NUM_IMAGES_PER_CLASS))
    label_name = re.sub(r'[^a-z0-9]+', ' ', dir_name.lower())
    training_images = []
    testing_images = []
    validation_images = []
    for file_name in file_list:
      base_name = os.path.basename(file_name)

      hash_name = re.sub(r'_nohash_.*$', '', file_name)
      hash_name_hashed = hashlib.sha1(compat.as_bytes(hash_name)).hexdigest()
      percentage_hash = ((int(hash_name_hashed, 16) %
                          (MAX_NUM_IMAGES_PER_CLASS + 1)) *
                         (100.0 / MAX_NUM_IMAGES_PER_CLASS))
      if percentage_hash < validation_percentage:
        validation_images.append(base_name)
      elif percentage_hash < (testing_percentage + validation_percentage):
        testing_images.append(base_name)
      else:
        training_images.append(base_name)
    result[label_name] = {
        'dir': dir_name,
        'training': training_images,
        'testing': testing_images,
        'validation': validation_images,
    }
  return result
```

Figure 8: Dataset splitting and Image format defining

"create_image_lists(image_dir, testing_percentage, validation_percentage)" this function reflect the following scenario.

Arguments:

Image directory: String path of a directory carrying the subfolders of given images.

Testing percentage: Integer number percentage of the photographs to order for tests.

Validation percentage: Integer level of images are stored for the validation.

Returns:

An ordered repository containing a section of every name of subfolder which divides images with training, testing, and validation sets under every label.

Train model

In this progression, we must maintain the parameters of the past layer, at that time expel the ultimate layer and info our data to retrain the new ultimate layer. The ultimate layer of this model is trained by backpropagation algorithmic system [18], besides this, the cost function cross-entropy [19] is utilized to integrate the weight parameter by shrewd the loss or blunder within the yield of the softmax layer and label vector of the provided test class [10] [13].

Model evaluation

The datasets of our models were separated into some sections- training data, validation data, and test data. The models were trained with the training data and then validated with the validation data. The following table describes the whole scenario of the models.

TABLE 1: DATASET DISTINGUISHING FOR TRAINING, TEST AND VALIDATION

Experiment	Total Data	Training Data	Validation Data	Test Data
Inception-v3	700	560	70	70
MobileNet-v1	700	560	70	70

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

We applied an efficient food items recognition model spending a short training time and achieved a greater precision. By the end of this chapter, one will understand the reason behind choosing the proposed model and its function to this project. We can find out the best model with the best accurate result by using different types of datasets. Hope next researchers will follow it and research on new data.

4.2 Experimental Results

The variations in accuracy based on cross-entropy in our training dataset for Inception-v3 model are presented in Figure 9 and Figure 10 respectively and for MobileNet-v1 model is displayed in Figure 11 and Figure 12 accordingly.

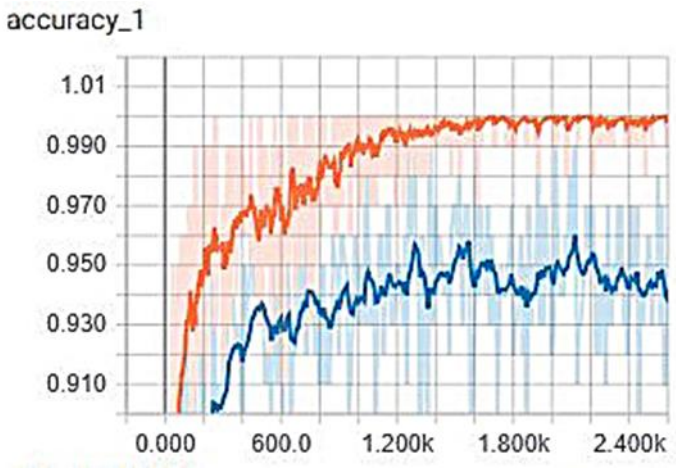


Figure 9: The variation of accuracy on training and validation set for Inception-v3

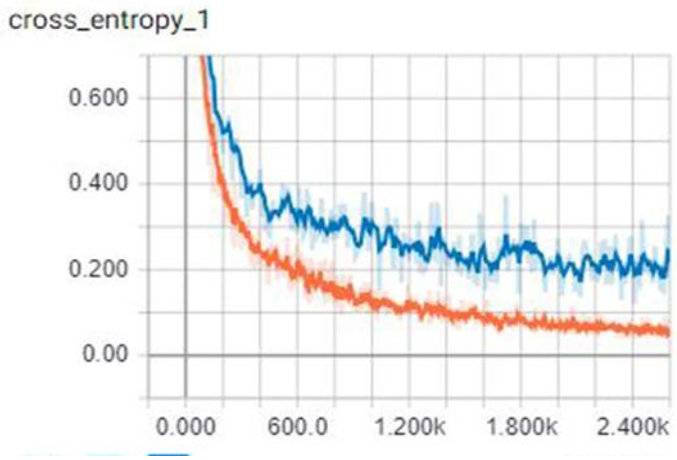


Figure 10: The variation of cross-entropy for Inception-v3

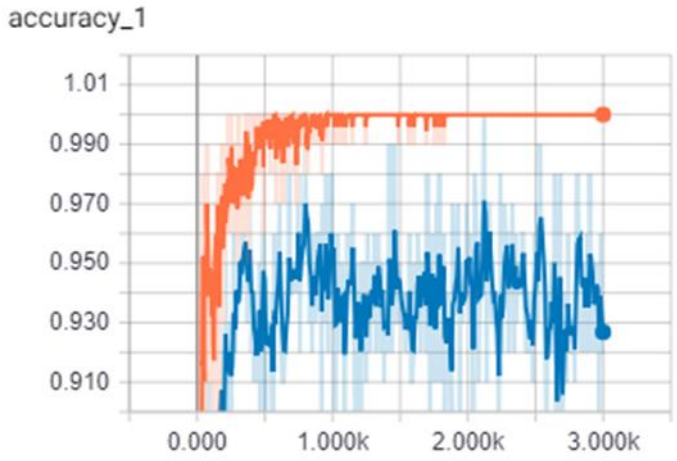


Figure 11: The variation of accuracy on training and validation set for Mobilenet-v1

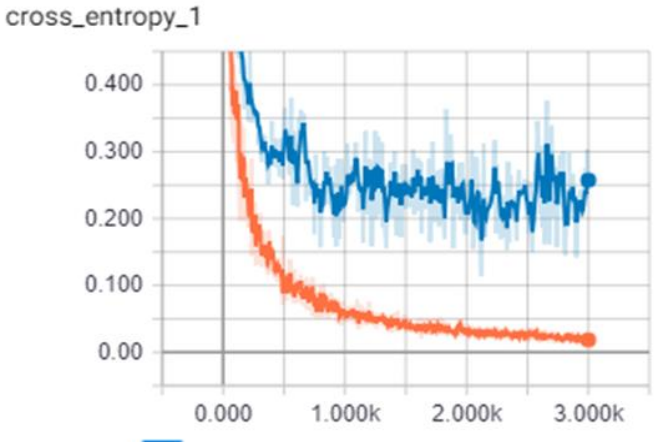


Figure 12: The variation of cross-entropy for MobileNet-v1

The training set is represented by the orange line, and the validation set is represented by the blue line.

The description of the figures is shown in the following table.

TABLE 1: DESCRIPTION OF THE FIGURES

Dataset	Index	Performance of Inception-v3	Performance of MobileNet-v1
Dataset	The precision of the training set	100%	100%
	The precision of the validation set	92% to 95%	90.5 to 96%
	The cross entropy of the training set	0.06	0.03
	The cross-entropy of the validation set	22	26

4.3 Descriptive Analysis

For our dataset, the training accuracy for Inception-v3 reached 100%, and the validation accuracy maintained between 92% to 95%. On the contrary, training accuracy for MobileNet-v1 model reached up to 100% but the validation accuracy wavers between 90.5% to 96%. However, cross entropy of MobileNet-v1 model is less than Inception-v3 model for training set but cross entropy of validation set is higher.

MobileNet-v1 gives higher accuracy than Inception-v3 but analyzing the validation accuracy and fluctuation, we can say that Inception-v3 is a better model for this dataset to classify and recognize traditional foods than MobileNet-v1 model.

4.4 Summary

At the end of the experiment, we can decide that the Inception-v3 model performing better than MobileNet-v1 model. And hopefully, this research paper will help the student and the researcher who wants to research more on this topic.

CHAPTER 5

Summary, Conclusion, Recommendation and Implication for Future Research

5.1 Summary of the Study

In our paper, we studied how to apply a convolutional neural network to the task of identifying and classifying images of foods. We focussed on classification models of computer vision used in data recognition. Different classification techniques of computer vision have merits and demerits for data classification and knowledge extraction. Furthermore, Inception-v3 and MobileNet-v1 models are helpful in classification.

In this section, we will like to describe the conclusion, recommendations and further improvement ideas of this research.

5.2 Recommendations

It is recommended:

- that appropriate model selection is an essential part of any classification;
- that a fresh and healthy image should be used;
- that an increase of data diversity will help to predict more accurately;
- that the creation of bottleneck is important.

5.3 Conclusions

We have demonstrated a comprehensive pathway to classify traditional foods of Bangladesh from food images, which is so far the first work of its kind. As a first research work on this domain, the result is quite satisfactory as well as encouraging. We also believe this work will inspire researchers from various countries to work on their traditional items.

5.4 Implication for Further Study

We proposed the classification model based on the Inception-v3 model for seven different food items. This model has been developed with future improvement possibilities. In data science project here data is the most important fact for working and making an applicable model. Hopefully, in the future, we could extend the work with a larger dataset having more varieties of items. We will try to establish our model more efficient in the future. We also have the plan to implement some other CNN based models to compare the accuracy on the same dataset.

Appendices

Appendix A: Research Reflection

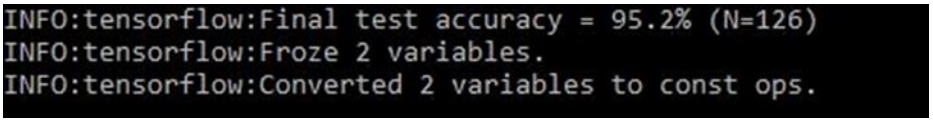


Figure A1: Testing accuracy of Inception-v3

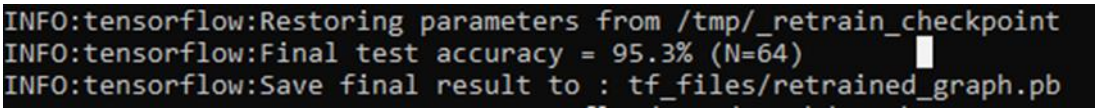


Figure A2: Testing accuracy of MobileNet-v1

References

[1] Inception-v3, available at << https://cloud.google.com/tpu/docs/inception-v3-advanced>>, last accessed on 01-09-2018 at 1:00pm.

[2] TensorFlow, available at << https://www.packtpub.com/mapt/book/big_data_and_business_intelligence/9781786468574/1/ch01lvl1sec9/tensorflow--a-general-overview>>, last accessed on 02-09-2018 at 5:00pm.

[3] Transfer learning, available at << https://searchcio.techtarget.com/definition/transfer-learning>>, last accessed on 03-09-2018 at 8:30pm.

[4] M. S. Islam, F. A. Foysal, and N. Neeha, "InceptB: A CNN Based Classification Approach for Recognizing Traditional Bengali Games", Procedia computer science, 143, pp. 595-602, January 2018.

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception ar-chitecture for computer vision," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826, 2016.

[6] Convolutional neural network architecture, available at << http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/>>, last accessed on 05-09-2018 at 7:30am.

[7] The Architecture of Convolutional Neural Network, available at << https://www.mdpi.com/10994300/19/6/242/htm>>, last accessed on 12-09-2018 at 2:00pm.

[8] Convolutional Neural Network, available at << https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+Network>>, last accessed on 05-09-2018 at 4:45pm.

[9] Flower category datasets, available at << https://datascience.stackexchange.com/questions/15989/microaverage-vs-macro-average-performance-in-a-multiclass-classification>>, last accessed on 06-09-2018 at 9:00pm.

[10] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," 2017 2nd International Conference on Image, Vision and Computing (ICIVC), pp. 783-787, June 2017.

[11] A. Mathew, J. Mathew, M. Govind, and A. Mooppan, "An Improved Transfer learning Approach for Intrusion Detection," Procedia computer science, 115, pp. 251-257, January 2017.

[12] B. Kieffer, M. Babaie, S. Kalra, and H. R. Tizhoosh, "Convolutional Neural Networks for Histopathology Image Classification: Training vs. Using Pre-Trained Networks," In 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1-6, November 2017.

[13] X. L. Xia, C. Xu, and B. Nan, "Facial expression recognition based on tensorflow platform," ITM Web of Conferences, vol. 12, p. 01005, EDP Sciences, 2017.

[14] B. E. Batsukh, and G. Tsend, "Effective Computer Model For Recognizing Nationality From Frontal Image," March 2016.

[15] Support vector machine, available at << https://en.wikipedia.org/wiki/Support_vector_machine>>, last accessed on 13-09-2018 at 12:00am.

[16] Active appearance model, available at << https://en.wikipedia.org/wiki/Active_appearance_model>>, last accessed on 15-09-2018 at 10:00am.

[17] T. Cootes, E. R. Baldock, and J. Graham, "An introduction to active shape models," Image processing and analysis, pp. 223-248, 2000.

[18] Backpropagation algorithm, available at << https://en.wikipedia.org/wiki/Backpropagation>>, last accessed on 11-09-2018 at 1:30pm.

[19] Cross-entropy, available at << https://en.wikipedia.org/wiki/Cross_entropy>>, last accessed on 08-09-2018 at 6:00pm.