A project report on

DEEP SUPERMARKET: TRANSFER LEARNING APPROACH FOR CLASSIFICATION OF INDIAN SUPERMARKET PRODUCTS

Submitted in partial fulfillment for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

by

VARUN VERMA (20BCE1506)

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

March, 2024

A project report on

DEEP SUPERMARKET: TRANSFER LEARNING APPROACH FOR CLASSIFICATION OF INDIAN SUPERMARKET PRODUCTS

Submitted in partial fulfillment for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

by

VARUN VERMA (20BCE1506)

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

March, 2024

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)
CHENNAI

DECLARATION

I hereby declare that the thesis entitled "DEEP SUPERMARKET: TRANSFER LEARNING APPROACH FOR CLASSIFICATION OF INDIAN SUPERMARKET PRODUCTS" submitted by me, for the award of the degree of Bachelor of Technology in Computer Science and Engineering, Vellore Institute of Technology, Chennai is a record of bonafide work carried out by me under the supervision of Dr. Noel Jeygar Robert V.

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Chennai

Date: **Signature of the Candidate**

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

School of Computer Science and Engineering

CERTIFICATE

This is to certify that the report entitled **"Deep Supermarket: Transfer Learning Approach for Classification of Indian Supermarket Products"** is prepared and submitted by **Varun Verma** (**20BCE1506**) to Vellore Institute of Technology, Chennai, in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology in Computer Science and Engineering programme** is a bonafide record carried out under my guidance. The project fulfills the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Signature of the Guide:

Name: Dr./Prof.

Date:

Signature of the Internal Examiner	Signature of the External Examiner
Name:	Name:
Date:	Date:

Approved by the Head of Department,
B.Tech. CSE
Name: Dr. Nithyanandam P

Date:

(Seal of SCOPE)

ABSTRACT

The project titled "Deep Supermarket: Transfer Learning Approach to Classify Indian Supermarket Products" intends to conduct a comparative analysis of different deep neural network models in order to effectively categorize Indian supermarket products. This will be achieved by employing both traditional and transfer learning (TL) techniques. The study centers on the necessity of precise product classifier that can identify Indian supermarket / grocery products for retail and self-checkout kiosks, as well as massive automated warehouse storages. The investigation revealed a dearth of image dataset specifically for Indian products, hence necessitating the creation of a customized image dataset (Indian Grocery Image dataset_v3). Initially, Convolutional Neural Network (CNN) and VGG19 models were developed, but their performance was suboptimal. Transfer learning techniques were employed, leveraging pre-trained (weights from ImageNet dataset) models like EfficientNetB7, InceptionResNetV2, DenseNet169, and DenseNet201. The transfer learning models significantly outperformed non-transfer learning models, with the InceptionResNetV2 and DenseNet family of DNN showing exceptional performance. Among the architectures, the DenseNet201 model showed the highest performance, with training accuracy of 99.36% and a validation accuracy of 80.47% making it the most optimal among them for the research problem of classifying Indian supermarket products.

ACKNOWLEDGEMENT

It is my pleasure to express with deep sense of gratitude to Dr. Noel Jeygar Robert V, Assistant Professor Sr Grade 1, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, for his constant guidance, continual encouragement, understanding; more than all, he taught me patience in my endeavor. My association with him is not confined to academics only, but it is a great opportunity on my part of work with an intellectual and expert in the field of Artificial Intelligence (AI) and Networks.

It is with gratitude that I would like to extend my thanks to the visionary leader Dr. G. Viswanathan our Honorable Chancellor, Mr. Sankar Viswanathan, Dr. Sekar Viswanathan, Dr. G V Selvam Vice Presidents, Dr. Sandhya Pentareddy, Executive Director, Ms. Kadhambari S. Viswanathan, Assistant Vice-President, Dr. V. S. Kanchana Bhaaskaran Vice-Chancellor, Dr. T. Thyagarajan Pro-Vice Chancellor, VIT Chennai and Dr. P. K. Manoharan, Additional Registrar for providing an exceptional working environment and inspiring all of us during the tenure of the course.

Special mention to Dr. Ganesan R, Dean, Dr. Parvathi R, Associate Dean Academics, Dr. Geetha S, Associate Dean Research, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai for spending their valuable time and efforts in sharing their knowledge and for helping us in every aspect.

In jubilant state, I express ingeniously my whole-hearted thanks to Dr. Nithyanandam P, Head of the Department, B.Tech. CSE and the Project Coordinators for their valuable support and encouragement to take up and complete the thesis.

My sincere thanks to all the faculties and staffs at Vellore Institute of Technology, Chennai who helped me acquire the requisite knowledge. I would like to thank my parents for their support. It is indeed a pleasure to thank my best friend Somya Kumari who encouraged me to take up and complete this task.

Place: Chennai

Date: **Varun Verma**

CONTENTS

ABSTRACT ..... **i**
ACKNOWLEDGEMENT ..... **ii**
CONTENTS ..... **iii**
LIST OF FIGURES ..... **v**
LIST OF TABLES ..... **vi**
LIST OF ACRONYMS ..... **vii**
CHAPTER 1
INTRODUCTION
1.1 OVERVIEW 1
1.1.1 STATEMENT OF PROBLEM 2
1.1.2 BACKGROUND CONTEXT OF PROBLEM 3
1.1.3 SIGNIFICANCE OF STUDY 4
1.2 DATASET CHALLENGE 5
1.3 DATA COLLECTION STRATEGIES 6
1.3.1 ADVANTAGES OF STRATEGY 7
1.3.2 DISADVANTAGES OF STRATEGY 8
1.3.3 ETHICAL CONSIDERATIONS / POTENTIAL RISKS 9
CHAPTER 2
LITERATURE REVIEW
2.1 PAPERS 10
2.1.1 THEMES DISCOVERED 21
2.1.2 GAPS IDENTIFIED 21
2.2 PROBLEM STATEMENT 22
2.3 CHALLENGES ADDRESSED 23
2.4 RESEARCH OBJECTIVES 24
CHAPTER 3
SYSTEM ARCHITECTURE AND DESIGN
3.1 OVERVIEW OF DEEP LEARNING 26
3.1.1 INTRODUCTION TO DEEP NEURAL NETWORKS 26
3.1.2 IMAGE CLASSIFICATION USING DNN 28
3.2 OVERVIEW OF TRANSFER LEARNING 30
3.2.1 INTRODUCTION TO PROCESS OF TRANSFER LEARNING 30
3.2.2 ADVANTAGES OF TRANSFER LEARNING 31
3.2.3 TRANSFER LEARNING OVER TRADITIONAL APPROACH 31
3.3 OVERVIEW OF THE MODELS USED IN RESEARCH 32

3.3.1 TRADITIONAL CNN MODEL 32
3.3.2 VGG19 ARCHITECTURE 33
3.3.3 EFFICIENTNETB7 MODEL 34
3.3.4 INCEPTIONRESNETV2 MODEL 35
3.3.5 DENSENET169 MODEL 36
3.3.6 DENSENET201 MODEL 37
3.4 METHODOLOGY 39
3.4.1 DATASET CREATION METHODOLOGY 39
3.4.2 MODEL CREATION AND COMPARISION 40

CHAPTER 4

PROPOSED SOLUTION

4.1 IMPLEMENTATION 42
4.2 PROPOSED SYSTEM 43
4.2.1 DATASET CREATION 43
4.2.2 INITIAL DATA ACQUISITION 43
4.2.3 DATASET CLEANING AND ORGANIZATION 44
4.2.4 CLASS BALANCING 46
4.2.5 PROPOSED SYSTEM FOR MODEL COMPARISION 48

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 RESULTS 58
5.1.1 PERFORMANCE OF CNN 58
5.1.2 PERFORMANCE OF VGG19 60
5.1.3 PERFORMANCE OF EFFICIENTNETB7 60
5.1.4 PERFORMANCE OF INCEPTIONRESNETV2 61
5.1.5 PERFORMANCE OF DENSENET169 62
5.1.6 PERFORMANCE OF DENSENET201 62
5.1.7 COMPARITIVE RESULTS 63
5.2 DISCUSSION 64

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION 65
6.2 FUTURE WORKS 65

REFERENCES 67

LIST OF FIGURES

FIGURE 3.1: VGG19 ARCHITECTURE 34
FIGURE 3.2 : METHODOLOGY OF STUDY 39
FIGURE 4.1: PROPOSED SYSTEM FOR DATASET CREATION 43
FIGURE 4.2: CLASS DISTRIBUTION OF DATASET_V2 45
FIGURE 4.3 : PERCENTAGE CLASS DISTRIBUTION OF DATASET_V2 46
FIGURE 4.4 : CLASS DISTRIBUTION OF DATASET_V3 47
FIGURE 4.5 : PERCENTAGE CLASS DISTRIBUTION OF DATASET_V3 47
FIGURE 4.6: PROPOSED SYSTEM FOR MODEL CREATION AND COMPARISON 48
FIGURE 5.1: ACCURACY OF TRADITIONAL CNN 58
FIGURE 5.2: LOSS OF TRADITIONAL CNN 58
FIGURE 5.3: ACCURACY OF TRADITIONAL VGG19 60
FIGURE 5.4: LOSS OF TRADITIONAL VGG19 60
FIGURE 5.5: ACCURACY OF ENB7 - MODEL1 60
FIGURE 5.6: LOSS OF ENB7 - MODEL1 60
FIGURE 5.7: ACCURACY OF ENB7 - MODEL2 61
FIGURE 5.8: LOSS OF ENB7 - MODEL2 61
FIGURE 5.9: ACCURACY OF INCEPTIONRESNETV2 61
FIGURE 5.10: LOSS OF INCEPTIONRESNETV2 61
FIGURE 5.11: ACCURACY OF DENSENET169 62
FIGURE 5.12: LOSS OF DENSENET169 62
FIGURE 5.13: ACCURACY OF DENSENET201 62
FIGURE 5.14: LOSS OF DENSENET201 62
FIGURE 5.15:TRAINING ACCURACIES OF ALL MODELS 63
FIGURE 5.16: TESTING ACCURACIES OF ALL MODELS 63
FIGURE 5.17:TRAINING LOSS FOR ALL MODELS 63
FIGURE 5.18: TESTING LOSS FOR ALL MODELS 63

LIST OF TABLES

4.1	ARCHITECTURE OF TRADITIONAL CNN	49
4.2	ARCHITECTURE OF TRADITIONAL VGG19	49
4.3	ARCHITECTURE OF ENB7 MODEL 1	51
4.4	ARCHITECTURE OF ENB7 MODEL2	52
4.5	ARCHITECTURE OF INCEPTIONRESNETV2	54
4.6	ARCHITECTURE OF DENSENET169	55
4.7	ARCHITECTURE OF DENSENET201	56
5.1	PERFORMANCE SCORES OF TRADITIONAL CNN	59
5.2	PERFORMANCE SCORES OF TRADITIONAL VGG19	60
5.3	PERFORMANCE SCORES OF ENB7 - MODEL1	61
5.4	PERFORMANCE SCORES OF ENB7 - MODEL2	61
5.5	PERFORMANCE SCORES OF INCEPTIONRESNETV2	62
5.6	PERFORMANCE SCORES OF DENSENET169	62
5.7	PERFORMANCE SCORES OF DENSENET201	63
5.8	TABLE OF COMPARISON BETWEEN ALL MODELS	64

LIST OF ACRONYMS

ACRONYM	ABBREVIATION
TL	Transfer Leaning
DNN	Deep Neural Networks
CNN	Convolutional Neural Network
VGG19	Visual Geometry Group
ENB7	EfficientNetB7 model
NO.	Number
RGB	Red, Blue, Green colors that a pixel of an image can be
FC_1, FC_2	Fully Connected Layers
SGD	Stochastic Gradient Descent
FAIR	Facebook AI Research
GPU	Graphics Processing Unit
US	United States
UK	United Kingdom
AI	Artificial Intelligence

Chapter 1

Introduction

1.1 OVERVIEW

In today's digital age, the advanced supermarkets with self-checkouts and automated large warehouses for shipping of products has transformed the way we shop by providing convenience and access to a diverse range of products. The growing need for intelligent technologies for autonomous product categorization in retail settings, particularly in automated retail warehouse storage and self-checkout terminals, necessitates the development of specialized product classifiers that can efficiently categorizing and classifying products in order to streamline the purchasing experience for customers and warehouse workers alike. While much study has been undertaken in image classification, there is a significant void in the literature about the classification of Indian supermarket / grocery merchandise.

To close this gap, this research article presents a thorough study titled "Deep Supermarket: Transfer Learning Approach to Classify Indian Supermarket Products." The primary goal of this work is to study various deep neural network models for accurately classifying Indian supermarket products using Deep learning and transfer learning (TL) techniques. The foundation of this research lies in the creation of a custom image dataset comprising 15 distinct classes, totalling 37,310 images. Given the absence of existing datasets containing Indian supermarket / grocery product images, the creation of this custom dataset was essential to facilitate model training and evaluation for the Indian market applications.

The study investigates the effectiveness of different deep learning models in categorizing Indian retail goods. Initially, two models are created: a customized Convolutional Neural Network (CNN) and the VGG19 architecture. Despite numerous attempts to enhance these models through architectural changes, hyperparameter tuning, and learning rate tweaks, both models performed suboptimal.

The paper then looks into transfer learning, using pre-trained models such as EfficientNetB7, InceptionResNetV2, DenseNet169, and DenseNet201. Transfer learning, a process in which knowledge gained from one problem is applied to another, has shown to be a game changer. The transfer learning models performed much better than the non-transfer learning models, demonstrating their capacity to use prior information to increase classification accuracy, specifically using the weights from the ImageNet dataset.

While the EfficientNetB7 model was projected to be the best performer, it only slightly outperformed the traditional CNN and VGG19 models. In contrast, the

InceptionResNetV2 and DenseNet family (including DenseNet169 and DenseNet201) models performed admirably, but with signs of possible overfitting.

Moving forward, this research paper project will delve deeper into the dataset, analysing its characteristics and resolving the issue of class imbalance. Furthermore, it will do a thorough assessment of the model architectures, compare their performance using empirical results and graphical analysis, and eventually make recommendations on the best models for the specific Indian supermarket items dataset.

1.1.1 STATEMENT OF PROBLEM

The modernization of supermarkets with advanced technologies like self-checkouts and automated warehouses has becoming a trend, offering convenience and access to a wide range of products in minimum amount of time. Examples of applications like Amazon Go and Walmart's Intelligent Retail Lab showcasing the potential of unmanned retail stores driven by deep learning technologies.

However, the lack of intelligent technologies for autonomous product categorization in retail settings poses a challenge, particularly in automated sectors mentioned above. The absence of specialized product classifiers capable of efficiently categorizing Indian grocery merchandise exacerbates this issue, leading to inefficiencies in both customer shopping experiences and warehouse operations.

This can be a serious hindrance for the Indian subcontinent because of the massive population. In India all most all of the cashiers and checkout as human operated. Though not always but many a times because of human nature, this results in long lines and unnecessary hindrances to customers. Similarly in the management of huge number of products and human error can lead to disorganization or misplacement of the products. Hence there is a need in India for making Image classifiers which can

Existing literature on image classification has predominantly focused on generic datasets which are of products from foreign countries like US and UK. They lack a comprehensive exploration of Indian supermarket products. This lack can be one of primary causes that there is a gap of research in trying to address this issue are there is no proper dataset of Indian product images. This gap highlights the need for tailored solutions that can be used to train classifiers to on Indian grocery items, taking into account their unique characteristics and cultural nuances using a dataset which contains the images of products relevant to the Indian supermarket landscape.

Thus, the primary problem addressed by this research is the creation of a custom image dataset and development of robust deep learning models for the classification of Indian supermarket products. By leveraging transfer learning techniques and creating a custom image dataset, this study aims to fill the void in the literature and provide insights into effective approaches for categorizing Indian retail goods. Through empirical evaluation and analysis, the research seeks to identify the most suitable deep learning architectures for enhancing the classification accuracy of Indian supermarket products, ultimately contributing to the optimization of retail operations and customer experiences in the Indian market in the future.

1.1.2 BACKGROUND CONTEXT OF PROBLEM

In the contemporary landscape of retail, modern supermarkets have undergone a remarkable evolution, embracing advanced technologies to redefine the shopping experience. These supermarkets are characterized by features such as self-checkouts, automated warehouses, and the seamless integration of digital tools to streamline operations and enhance customer convenience.

The significance of advanced technologies in modern supermarkets cannot be overstated. These technologies play a pivotal role in improving efficiency, optimizing inventory management, and elevating customer satisfaction levels. Barcode scanners, inventory management systems, and automated checkout counters are just a few examples of the technologies that have revolutionized the retail sector, offering unparalleled convenience and speed to consumers.

In parallel with the advancements in modern supermarkets, the concept of unmanned retail stores has gained traction, showcasing the potential of deep learning technologies to revolutionize the retail landscape further. Two prominent examples of such innovative ventures are Amazon Go and Walmart Intelligent Retail Labs, both of which exemplify the seamless integration of deep learning technologies into retail operations.

Amazon Go, introduced by e-commerce giant Amazon, represents a pioneering approach to unmanned retail stores. Leveraging a sophisticated combination of computer vision, sensor fusion, and deep learning algorithms, Amazon Go enables customers to enter the store, pick up items, and walk out without the need for traditional checkout counters. The system automatically detects and adds items to the customer's virtual cart, leveraging advanced machine learning models to accurately identify products and track customer movements in real time. This innovative approach not only streamlines the shopping experience but also showcases the transformative potential of deep learning technologies in retail environments.

Similarly, Walmart Intelligent Retail Labs is at the forefront of leveraging deep learning technologies to enhance retail operations. With the aim of optimizing store operations and improving customer experiences, Walmart has deployed advanced AI-powered systems in select stores. These systems utilize computer vision, machine learning, and data analytics to monitor inventory levels, analyse customer behaviour, and optimize product placement. By harnessing the power of deep learning, Walmart Intelligent

Retail Labs aims to increase operational efficiency, reduce costs, and deliver personalized shopping experiences to customers.

Both Amazon Go and Walmart Intelligent Retail Labs serve as compelling examples of the transformative potential of unmanned retail stores driven by deep learning technologies. By automating routine tasks, optimizing inventory management, and personalizing customer experiences, these innovative ventures are reshaping the retail landscape and setting new standards for efficiency, convenience, and customer satisfaction. As deep learning continues to evolve, the possibilities for unmanned retail stores are boundless, offering exciting opportunities for further innovation and disruption in the retail sector.

Nevertheless, despite the progress made in technology, Indian supermarkets continue to encounter many obstacles in their operations, namely in automated areas such as self-checkout terminals and warehouse storage systems. This can be attributable to the continuously expanding and immense population. Annually, the Indian population is increasing by 1% and now stands at 146.86 Crore. Indian supermarkets require automation to minimize the time wasted in lengthy waits and prevent mishandling of bulk products at large warehouses owing to human error, as a significant number of manual labourers in India lack basic education. Supermarkets in the digital age have significant obstacles such as product classification, inventory management, and providing a smooth consumer experience. These challenges can only be overcome by utilizing advanced technologies.

An evident problem in retail environments is the lack of dedicated classifiers literature for autonomous classification of products, particularly when it comes to Indian grocery items. The absence of intelligent technologies worsens inefficiencies and obstructs operations in supermarkets, resulting in below-optimal performance and customer unhappiness. The issue is exacerbated by the vast number of Indian consumers. The lack of specialized technologies designed for the Indian supermarket industry may be a major factor contributing to the complete absence of automated stores on the Indian subcontinent.

1.1.3 SIGNIFICANCE OF STUDY

Situated within the domain of AI applications for automated retail, this research addresses the growing demand for intelligent technologies to streamline product categorization processes in retail settings. The significance of this study lies in its dual contributions.

Firstly, the creation of a custom dataset tailored to Indian supermarket products fills the void in existing resources, enabling more accurate and culturally relevant classification models.

Secondly, by exploring and comparing various deep learning architectures such as CNN, InceptionResNetV2, EfficientNet and DenseNet, the research aims to identify the optimal classifier architecture for Indian retail environments.

The expected outcomes include the identification of the most effective model for image classification of Indian supermarket products, the construction of a specialized dataset to support this endeavour. The significance of this research paper aligns with provisioning of practical insights for the development and implementation of automated systems in superstores, aligning with the evolving trend towards minimal human interaction and maximal efficiency in retail operations.

1.2 DATASET CHALLENGE

The primary challenges encountered during the course of this research was the unavailability of adequate datasets tailored to the specific requirements of classifying Indian supermarket products. In order to train a model capable of accurately categorizing the diverse range of products found in Indian supermarkets, it was essential to have access to a comprehensive dataset comprising labeled images representing various product categories. However, existing datasets available online did not fulfill these criteria, posing a significant hurdle in the initial stages of the research.

Although efforts were made to identify potential datasets containing Indian product images, the available options were limited and did not meet the necessary labeling and organization standards required for training an effective classifier.

For instance, a dataset containing Indian product images was identified; however, the images were not labeled or organized according to product categories, rendering them unsuitable for the research purposes.

Consequently, it was decided to create a custom dataset specifically tailored to the research objectives. This undertaking presented its own set of challenges, primarily stemming from the need to collect and organize a large volume of relevant images representing Indian supermarket products. An automated Python script was developed to streamline the process of downloading images from online sources. However, due to limitations in the image search engine utilized by the script (Bing), the results obtained were not always accurate or relevant to the search query.

As a result, a significant portion of the downloaded images were found to be irrelevant or unrelated to Indian supermarket products, necessitating manual intervention to filter out the undesirable images. This manual curation process added considerable time and effort to the dataset creation process, highlighting the complexity of obtaining high-quality labeled datasets for specialized research domains such as Indian supermarket product classification.

Despite these challenges, the creation of a custom dataset was deemed essential to facilitate the training and evaluation of classification models tailored to the Indian market context.

1.3 DATA COLLECTION STRATEGIES

In the pursuit of creating a custom dataset tailored to the classification of Indian supermarket products, several data collection strategies were devised to overcome logistical constraints and tight deadlines. Given the impracticality of physically capturing images due to time constraints, online sources were explored as the primary means of data acquisition.

Here's an overview of the data collection strategies employed:

Selection of Online Cloud Storage:
Google Drive was chosen as the online cloud storage platform for the dataset, given its integration with Google Collaboratory where the model notebooks were initially planned to be developed.

Utilization of Online Image Scrapers Package:
Initially, efforts were made to utilize the google_images_download Python package for scraping images from Google Images. However, due to structural updates to the Google Images website, this approach became impractical. Subsequently, the bing_image_downloader Python package was employed to scrape and download images from Microsoft Bing, despite its limitations in terms of search accuracy.

Python Script Development:
A custom Python script was developed to automate the process of downloading images from online sources. The script was designed to create the necessary directory structure, with each product class having its own folder. Within each class folder, subfolders corresponding to specific search queries were created to organize the downloaded images.

Identification of Product List:
To determine the list of product images required for the dataset, various online grocery websites such as BigBasket and JioMart were consulted. The product list obtained from these sources served as the basis for conducting search queries and downloading relevant images.

Folder Structure Inconsistencies:
One major challenge encountered during the data collection process was the inability

to modify the folder structure created by the bing_image_downloader package. This resulted in a nested folder structure, which required manual adjustments to align with the desired organization.

Relevance of Downloaded Images:
Another significant challenge was the lack of relevance in the images downloaded from Bing due to the limitations of its image search capabilities. Approximately 50-60% of the downloaded images were deemed relevant for each search query, necessitating additional manual filtering to remove irrelevant images.

Data Cleaning and Augmentation:
Despite the challenges faced, all downloaded images were stored in their raw format without undergoing any data cleaning or augmentation processes. This ensured the preservation of the original dataset integrity for subsequent analysis.

Dataset Versioning:
Upon completion of the data collection phase, the collected dataset comprising approximately 30,000 images was labeled as "dataset_v1" (version 1), signifying the initial iteration of the custom dataset.

Data Cleaning, Organizing and Balancing:
The data in dataset_v1 was subjected to meticulous human cleaning and categorization, resulting in a reduced but usable dataset called "dataset_v2". By employing data augmentation and random sampling procedures, a new version of the dataset called dataset_v3 was created. This version is characterized by a balanced distribution of data and is very suitable for the study's classification challenge.

1.3.1 ADVANTAGES OF STRATEGY

- Efficiency: Utilizing online sources for data acquisition allowed for the rapid collection of a large volume of images within a short timeframe, addressing the tight deadlines of the research project.

- Cost-Effectiveness: Online image scraping methods eliminated the need for costly equipment or resources associated with physical image capture, resulting in cost savings.

- Accessibility: Leveraging online cloud storage platforms like Google Drive ensured easy access to the dataset for all collaborators, facilitating seamless collaboration and sharing of resources.

- Automation: The development of a custom Python script enabled the automation of the data collection process, reducing manual effort and increasing overall efficiency.

- Diverse Data Sources: Consulting multiple online grocery websites provided access to a diverse range of Indian supermarket products, ensuring the dataset's comprehensiveness and representativeness.

1.3.2 DISADVANTAGES OF STRATEGY

- Not really Automated: Thought the python package was helping in downloading 20-40 Images of a product all in one go, the package did not work by using List (array) as in input. Which basically meant that each and every search query needed to manually enter into the function and run the script over and over again. Because of this, the script had to be modified even more so as to run multiple instances of the same function just to get the job done more efficiently as a single instance was causing a major bottleneck.

- Internet Speed: This method was heavily dependent on the internet speed. Even after having a significantly high internet speed (at least for India), the downloading process for each query took about 1-3 mins. This resulted in a lot of time needed to completely download the entire version 1 of the dataset approximately 2-3 weeks.

- Search Accuracy Limitations: The reliance on online image scraping tools such as bing_image_downloader resulted in limitations in search accuracy, leading to the retrieval of irrelevant images and potential dataset contamination.

- Folder Structure Constraints: Inflexibilities in modifying the folder structure generated by the image scraping packages posed challenges in organizing the dataset according to the desired hierarchy, requiring additional manual adjustments. These manual adjustments also resulted in a lot of time wasted in just organizing the data later.

- Quality Control Issues: The lack of manual oversight during the data collection process may have resulted in the inclusion of low-quality or irrelevant images in the dataset, potentially affecting the model's performance. Also, because many of the products belonging to different classes but look the same in terms of packets or colors may lead to poor performance.

- Dataset Cleaning Inconsistency: The dataset though was cleaned and balanced later, there might be some products which are more than one class due to human errors and restricted deadlines. This could also lead to poor performance of the models.

- Ethical Concerns: The automated scraping of images from online sources raises ethical considerations regarding copyright infringement and intellectual property rights, necessitating careful adherence to usage policies and guidelines.

1.3.3 ETHICAL CONSIDERATIONS / POTENTIAL RISKS

- Copyright Compliance: It is essential to ensure that the images collected from online sources adhere to copyright laws and usage permissions. Proper attribution and obtaining permission, if required, are crucial ethical considerations in data collection.

- Privacy Concerns: Respect for user privacy is paramount when scraping images from online sources. Care must be taken to avoid the collection of personal or sensitive information inadvertently, and measures should be implemented to anonymize data where necessary.

- Data Transparency: Transparency in data collection practices involves clearly communicating the sources and methods used to collect the dataset. Researchers should document the data collection process comprehensively to ensure reproducibility and accountability.

- Data Security: Safeguarding the integrity and confidentiality of collected data is essential to prevent unauthorized access or misuse. Implementing robust security measures and protocols helps mitigate risks associated with data breaches or unauthorized access.

The dataset creation process involved a combination of automated scraping, manual cleaning, and organizational adjustments to ensure the dataset's integrity and suitability for model training and evaluation. Through iterative improvements and versioning, the dataset would be refined to mitigate class imbalances and optimize performance for image classification tasks.

Chapter 2

Literature Review

2.1 PAPERS

- "Store product classification using CNN" by I Made Wiryana, Suryadi Harmanto, Alfharizky Fauzi, Imam Bil Qisthi, Zalita Nadya Utami [1]

The research paper "Store product classification using Convolutional Neural Network" by I Made Wiryana, Suryadi Harmanto, Alfharizky Fauzi, Imam Bil Qisthi, and Zalita Nadya Utami aims to utilize a CNN architecture to enhance the efficiency and reduce the costs associated with sorting products in stores. The study delves into the realm of deep learning-based retail product recognition, emphasizing the significance of CNN models in detecting various objects for product identification. The paper provides a comprehensive literature review, discusses challenges, techniques, and available datasets for deep learning-based product recognition, offering valuable insights for researchers and engineers entering this field.

The research highlights the evolution from traditional computer vision methods to deep learning approaches in retail product recognition. It outlines the process of image capture, preprocessing, feature extraction, feature classification, and output recognition in product image recognition systems. Deep learning methods have revolutionized computer vision by enabling more precise feature extraction through deeper layers, leading to significant advancements in image classification and object detection tasks.

Moreover, the paper sheds light on the increasing interest in deep learning-based retail product recognition within the industry, with notable applications like Amazon Go and Walmart's Intelligent Retail Lab showcasing the potential of unmanned retail stores driven by deep learning technologies. Despite the progress in this field, there is a scarcity of comprehensive reviews or surveys summarizing existing achievements and current advancements in deep learning-based retail product recognition.

In conclusion, this research paper contributes by providing a detailed overview of computer vision methods for product recognition, addressing challenges in detecting grocery products in retail stores, presenting current techniques to solve complex problems, discussing available datasets, and outlining future research directions in the domain of deep learning-based retail product identification.

- "Convolutional Neural Network (CNN) of Resnet-50 with Inceptionv3 Architecture in Classification on X-Ray Image" by Muhathir, M. F. Dwi Ryandra, R. B. Y. Syah, N. Khairina, and R. Muliono. [2]

The research paper "Convolutional Neural Network (CNN) of Resnet-50 with

Inceptionv3 Architecture in Classification on X-Ray Image" by Muhathir, M. F. Dwi Ryandra, R. B. Y. Syah, N. Khairina, and R. Muliono focuses on utilizing advanced CNN architectures like Resnet-50 and Inceptionv3 for the classification of X-ray images. The Inceptionv3 architecture, known for its efficiency in computational power, incorporates techniques like factorized convolutions, smaller convolutions, asymmetric convolutions, auxiliary classifiers, and grid size reduction to optimize network performance and reduce computational costs. This architecture was proposed in the paper "Rethinking the Inception Architecture for Computer Vision" by Christian Szegedy et al. in 2015.

Inceptionv3 was trained on the ImageNet dataset and compared with other contemporary models, demonstrating superior performance when augmented with techniques like an auxiliary classifier, factorization of convolutions, RMSProp, and Label Smoothing. The research highlights the importance of these architectural modifications in achieving lower error rates compared to other models.

Overall, this paper contributes to the field of deep learning by showcasing the effectiveness of combining Resnet-50 and Inceptionv3 architectures for X-ray image classification, emphasizing the significance of architectural enhancements in improving model performance and efficiency.

- "An effective CNN and Transformer complementary network for medical image segmentation," An effective CNN and Transformer complementary network for medical image segmentation – ScienceDirect [3]

The research paper "An effective CNN and Transformer complementary network for medical image segmentation" explores the integration of Convolutional Neural Network (CNN) and Transformer encoders to enhance medical image segmentation. The motivation behind combining CNN and Transformer encoders lies in their complementary nature, where CNN encoders excel in capturing local features while Transformer encoders are adept at modeling long-range dependencies for global representations.

The study introduces a hybrid architecture named ConvFormer, which combines CNN and Transformer components to leverage the strengths of both approaches in learning local and global representations for improved medical image segmentation. ConvFormer incorporates a feed-forward module called Enhanced DeTrans to introduce local information, a residual-shaped hybrid stem to capture both local and global features effectively, and an encoder that generates feature maps in different scales to exploit multi-scale features for enhanced segmentation performance.

Experiments conducted with ConvFormer demonstrate its superiority over various CNN- or Transformer-based architectures, achieving state-of-the-art performance in medical image segmentation tasks. This research contributes significantly to advancing the field of medical image analysis by leveraging the synergies between CNN and Transformer architectures to enhance the accuracy and efficiency of image segmentation processes.

- “Comparison of CNN-based deep learning architectures for rice diseases classification,” Comparison of CNN-based deep learning architectures for rice diseases classification – ScienceDirect [4]

The research paper "Comparison of CNN-based deep learning architectures for rice diseases classification" by Ahmed, T., Rahman, C. R., and Abid, M. F. M. aims to compare the performance of various CNN architectures in the classification of rice diseases using deep learning models. The study focuses on the effectiveness of different CNN architectures, including GoogleNet, ResNet-18, SqueezeNet-1.0, and DenseNet-121, in accurately identifying rice diseases. The authors also discuss the limitations of CNN-based rice disease recognition and provide insights into future research directions.

The paper highlights the importance of using deep learning models for rice disease classification, as they can automatically extract image properties and advanced features that cannot be extracted manually. However, the authors also acknowledge the challenges associated with rice disease recognition, such as the requirement for massive computational resources and the need for a large amount of data to train the model effectively.

The research contributes to the field of rice disease classification by providing a comprehensive comparison of different CNN architectures and their performance in identifying rice diseases. The findings of this study can be useful for researchers and practitioners working in the agricultural domain, as they can help in selecting the most suitable deep learning model for rice disease classification tasks.

- "Deep Learning Model Based on ResNet-50 for Beef Quality Classification" by S. E. Abdallah, W. M. Elmessery, M. Y. Shams, and N. S. A. Al-Sattary [5]

The research paper "Deep Learning Model Based on ResNet-50 for Beef Quality Classification" by S. E. Abdallah, W. M. Elmessery, M. Y. Shams, and N. S. A. Al-Sattary presents a novel approach to classify beef images into healthy and rancid categories using a deep learning model based on ResNet-50 architecture. The study addresses the challenge of distinguishing between healthy and rancid beef images, which is difficult due to the texture analysis of the beef surface.

In this research, a limited number of beef images were initially collected, including eight healthy images and ten rancid beef images, which were insufficient for traditional deep learning model training. To overcome this limitation, Generative Adversarial

Network (GAN) was employed to augment the dataset, resulting in a total of one hundred eighty images for training and evaluation purposes.
The results obtained from the ResNet-50 classification model showcased high accuracy rates of 96.03% in training, 91.67% in testing, and 88.89% in validation phases, demonstrating the effectiveness of the proposed deep learning approach for beef quality classification. The study also compared the ResNet-50 model with classical and other deep learning architectures to highlight its efficiency in image classification tasks.

Overall, this research contributes significantly to the field of food quality measurement by introducing a robust deep learning model based on ResNet-50 for accurately classifying beef images into healthy and rancid categories, addressing the challenges associated with limited image resources and enhancing classification accuracy through advanced deep learning techniques.

- “Performance evaluation of ResNet model for classification of tomato plant disease” by S. Kumar, S. Pal, V. P. Singh, and P. Jaiswal [6]

The research paper "Performance evaluation of ResNet model for classification of tomato plant disease" by S. Kumar, S. Pal, V. P. Singh, and P. Jaiswal aims to evaluate the performance of a ResNet model in classifying tomato plant diseases. The study focuses on the effectiveness of the ResNet model in accurately identifying different types of tomato diseases, including early blight, late blight, leaf mold, leaf spot, two-spotted spider mite, target spot, yellow leaf curl virus, and mosaic virus.

To evaluate the performance of the ResNet model, the authors used a dataset of tomato leaves containing six different types of diseases and one healthy tomato leaf class. The dataset was created by collecting images of tomato leaves affected by various diseases and healthy leaves from different sources, including the PlantVillage dataset.

The study also compares the performance of the ResNet model with other deep learning architectures, such as Inception V3 and Inception ResNet V2, to highlight the effectiveness of the ResNet model in tomato plant disease classification tasks. The results obtained from the ResNet model showcase high accuracy rates in training, testing, and validation phases, demonstrating the efficiency of the proposed deep learning approach for tomato plant disease classification.

In conclusion, this research contributes significantly to the field of plant disease classification by providing a comprehensive performance evaluation of the ResNet model for tomato plant disease classification. The findings of this study can be useful for researchers and practitioners working in the agricultural domain, as they can help in selecting the most suitable deep learning model for tomato plant disease classification tasks.

- “An Enhanced Transfer Learning Based Classification for Diagnosis of Skin Cancer” by V. Anand, S. Gupta, A. Altameem, S. R. Nayak, R. C. Poonia, and A. K. Jilani Saudagar [7]

The research paper "An Enhanced Transfer Learning Based Classification for Diagnosis of Skin Cancer" by V. Anand, S. Gupta, A. Altameem, S. R. Nayak, R. C. Poonia, and A. K. Jilani Saudagar aims to develop a deep learning-based model for the diagnosis of skin cancer at benign and malignant stages using the concept of transfer learning. The study focuses on improving the VGG16 model by adding layers and fine-tuning it for better performance in skin cancer classification.

The proposed framework consists of four main components: input dataset, data augmentation, feature extraction using VGG16, and fine-tuning of the VGG16 model. The authors performed data augmentation techniques in the pre-processing stage to increase the randomness and size of the dataset. The model's efficacy was analyzed by tuning various hyperparameters, such as batch size, epochs, and optimizer.

The results of the study demonstrate that the proposed model achieved high accuracy rates in training, testing, and validation phases, indicating the effectiveness of the enhanced transfer learning approach in skin cancer classification. The study also compares the proposed model with state-of-the-art methods to highlight its superiority in diagnosing skin cancer.

In conclusion, this research contributes significantly to the field of skin cancer diagnosis by introducing an enhanced transfer learning-based approach using the VGG16 model for accurate classification of skin cancer at benign and malignant stages. The findings of this study can be useful for researchers and practitioners working in the medical domain, as they can help in developing more efficient and accurate automated systems for skin cancer diagnosis.

- “A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope” by A. W. Salehi et al. [8]

The research paper "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope" by A. W. Salehi et al provides a comprehensive review of the benefits, difficulties, and future prospects of using Convolutional Neural Networks (CNN) and transfer learning in medical imaging applications. The paper discusses how CNN and transfer learning techniques have improved accuracy, reduced resource requirements, and enhanced the efficiency of medical image analysis processes.

The study delves into the challenges faced by deep learning-based techniques in medical imaging, such as limited training data, overfitting, and vanishing gradient issues, while also exploring strategies for selecting the appropriate transfer learning technique for specific problems. It provides insights into the major components of CNNs, including convolutional layers, pooling layers, fully connected layers, and important parameters for building CNN models.

Furthermore, the research paper discusses various imaging modalities used in training models for medical imaging analysis and contrasts the impact of computational resources like GPUs, CPUs, and TPUs on deep learning algorithms. By reviewing related research work and discussing cutting-edge architectures of CNNs, the paper aims to enhance the understanding of CNN and transfer learning techniques among researchers and students in the field of medical imaging.

Overall, this study contributes valuable insights into the application of CNNs and transfer learning in medical imaging, highlighting their advantages, challenges, and future directions for improving diagnostic accuracy and efficiency in healthcare settings.

- “Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning - BMC Medical Imaging,” by M. M. Srikantamurthy, V. P. Subramanyam Rallabandi, D. B. Dudekula, S. Natarajan, and J. Park [9]

The research paper "Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning" by M. M. Srikantamurthy, V. P. Subramanyam Rallabandi, D. B. Dudekula, S. Natarajan, and J. Park aims to develop a deep learning model for the classification of benign and malignant subtypes of breast cancer histopathology images. The study leverages a hybrid CNN-LSTM (Convolutional Neural Network-Long Short-Term Memory) model to classify and predict four subtypes of benign and malignant cancer images.

The proposed model utilizes a transfer learning approach, where the authors pre-train the model on the ImageNet dataset and fine-tune it on the BreakHis dataset, which comprises 2480 benign and 5429 malignant cancer images acquired at 100×, 200×, and 400× magnifications. The model was evaluated using various optimizers, including adaptive moment estimator (Adam), root mean square propagation (RMSProp), and stochastic gradient descent (SGD), with varying numbers of epochs.

The results of the study demonstrate that the hybrid CNN-LSTM model achieved high accuracy rates for both binary and multi-class classification tasks. The model achieved an overall accuracy of 99% for binary classification of benign and malignant cancer subtypes and 92.5% for multi-class classification of benign and malignant cancer subtypes.

The authors also compared the proposed model with existing CNN models, such as ResNet50 and Inception, and observed that the hybrid CNN-LSTM model outperformed the state-of-the-art machine and deep learning models in classifying breast cancer histopathology images.

In conclusion, this research contributes to the field of medical imaging by proposing a hybrid CNN-LSTM model for the classification of benign and malignant subtypes of breast cancer histopathology images, which can potentially improve diagnostic accuracy and efficiency in clinical settings.

- "Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images" Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images – ScienceDirect [10]

The research paper "Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images" by M. M. Srikantamurthy, V. P. Subramanyam Rallabandi, D. B. Dudekula, S. Natarajan, and J. Park focuses on developing a hybrid model that combines U-Net and Convolutional Neural Network (CNN) for the segmentation and classification of skin lesions from dermoscopy images. The study addresses the challenges in accurately segmenting skin lesions due to fuzzy borders, irregular boundaries, and inter-class variances, which make nodule segmentation a complex task.

The proposed model enhances the U-Net architecture by modifying the feature map dimensions to improve segmentation accuracy and automatic segmentation of skin lesions from dermoscopic images. By adjusting the feature map dimensions and increasing the number of kernels for precise nodule extraction, the model aims to achieve more accurate segmentation results.

The effectiveness of the proposed model was evaluated by considering various hyperparameters such as epochs, batch size, types of optimizers, and implementing augmentation techniques to enhance the dataset available in the PH2 dataset. The best performance was achieved with an Adam optimizer, a batch size of 8, and 75 epochs.

Overall, this research contributes to advancing medical imaging by proposing a modified U-Net architecture for accurate segmentation of skin lesions from dermoscopy images. The study's findings can potentially improve diagnostic accuracy in dermatology by providing a more precise and automated method for segmenting and classifying skin lesions based on dermoscopy images.

- “Product Classification in E-Commerce Sites,” Product Classification in E-Commerce Sites by A. Patra, V. Vivek, B. R. Shambhavi, K. Sindhu, and S. Balaji [11]

The research paper "Product Classification in E-Commerce Sites" by A. Patra, V. Vivek, B. R. Shambhavi, K. Sindhu, and S. Balaji delves into the importance of proper product categorization in e-commerce platforms for enhancing user experience, improving searchability, and driving higher conversions. Effective product classification ensures that online stores are easily navigable, provide logical navigation paths, and facilitate a smoother path to purchase for customers.

Key benefits highlighted in the study include:

- Positive Customer Experience: Well-classified catalogs lead to a significantly better shopping experience, increasing browsing and sales while setting up e-commerce websites for successful selling.
- Improved Search-Ability: Properly mapped and cataloged websites enhance search-ability for shoppers, providing a better experience and driving higher sales
- Accurate Site Navigation: Accurate site navigation enables shoppers to easily find and access products through dropdown catalogs and lists on the site.
- Easier Path to Purchase: Accurate data classification streamlines the path to purchase, making it faster, easier, and smoother for shoppers, ultimately boosting e-commerce sales.
- Increased Conversion: Data classification leads to increased conversions through accurate search results, logical site navigation, and an optimal shopping experience.

Furthermore, the study emphasizes that product categorization is not only crucial for enhancing customer experience but also plays a vital role in improving reporting accuracy, decision-making processes, productivity, and communication between teams within an organization. Properly structured product pages allow businesses to analyze sales performance effectively and make informed decisions based on the performance of different products.

In conclusion, the research underscores the significance of product classification in e-commerce sites for creating a seamless shopping experience, optimizing customer engagement, improving operational efficiency, and ultimately driving higher sales conversions.

- “Machine Learning-Based Autonomous Framework for Product Classification Over Cloud,” by A. Motwani, G. Bajaj, M. Arya, S. K. Sar, and S. O. Manoj [12]

The research paper "Machine Learning-Based Autonomous Framework for Product Classification Over Cloud" by A. Motwani, G. Bajaj, M. Arya, S. K. Sar, and S. O. Manoj focuses on developing an autonomous framework for product classification using machine learning techniques deployed over the cloud. The study aims to automate the process of product categorization in e-commerce platforms to enhance efficiency, accuracy, and scalability in handling large volumes of products.

Key points from the research paper include:

- Classification Framework: The methodology involves implementing a classification framework based on multiclass logistic regression for product categorization, enabling automated identification and categorization of products
- Cloud-Based Approach: The framework leverages cloud computing infrastructure to enable scalable and efficient processing of product data, allowing for seamless integration with e-commerce platforms and enhancing the overall classification process
- Autonomous Product Identification: The study emphasizes the development of an autonomous system that can identify and categorize products without the need for expert or domain-specific knowledge, streamlining the product classification process in e-commerce settings.
- Scalability and Efficiency: By utilizing machine learning algorithms and cloud-based resources, the framework aims to improve scalability, efficiency, and accuracy in product classification tasks, catering to the dynamic nature of e-commerce platforms with a large variety of products.

Overall, this research paper highlights the significance of deploying machine learning algorithms over cloud infrastructure for automating product classification processes in e-commerce platforms. The proposed autonomous framework offers a promising solution to efficiently handle product categorization tasks at scale, contributing to improved user experience, streamlined operations, and enhanced decision-making processes in e-commerce businesses.

- “Deep Learning Approach to Recyclable Products Classification: Towards Sustainable Waste Management” by M. I. Basheer Ahmed et al [13]

M. I. Basheer Ahmed et al.'s research article "Deep Learning Approach to Recyclable Products Classification: Towards Sustainable Waste Management" focuses on using deep learning techniques to categorize recyclable products, with the goal of improving waste management practices to promote sustainability. The paper covers the growing issues of waste management, specifically the rise in food and plastic waste caused by factors such as the COVID-19 epidemic, emphasizing the significance of effective trash classification systems.

The research paper overview highlights the following key points: The study intends to create a deep learning model to classify recyclable products, improve waste management methods, and promote sustainability in waste handling operations. The suggested method comprises testing deep learning models on a large public dataset to determine their accuracy in categorizing recyclable items. The research assesses the performance of the constructed model using well-known metrics and compares it with the most recent studies in the literature, revealing the efficacy of deep learning models in trash classification tasks.

The study investigates a variety of deep learning architectures, including Convolutional Neural Networks (CNN), MobileNetV2, ResNet50V2, and DenseNet169, in order to create an effective model for recyclable product classification. The study's findings shed light on the potential of deep learning methodologies for improving waste management procedures, providing a promising alternative for automating and enhancing recyclable product classification processes.

Overall, this research paper contributes to advancing sustainable waste management practices by leveraging deep learning techniques for accurate and efficient classification of recyclable products. The study's outcomes have implications for promoting environmental sustainability, optimizing waste handling processes, and fostering more effective recycling initiatives in support of a greener future.

- “Convolutional Neural Network (CNN) Based Identification of Crop Diseases Using ResNet-50” - IEEE Conference Publication, IEEE Xplore [14]

The research paper "Convolutional Neural Network (CNN) Based Identification of Crop Diseases Using ResNet-50" is a study that explores the use of deep learning techniques, specifically Convolutional Neural Networks (CNNs), for the automatic identification of crop diseases from images. The study was funded by various organizations, including the Centre de Géomatique du Québec (CGQ), Mitacs, the Ministère de l'Économie et de l'Innovation (MEI) of the province of Québec, and Microdrones' company.

The paper surveyed 19 studies that used CNNs to automatically identify crop diseases and analyzed their profiles, implementation aspects, and performance. The study identified major issues and shortcomings of the works in this research area, such as a lack of conformity, which may lead to poor generalization capabilities for unfamiliar data samples and imaging conditions, limiting the practical use of the trained models.

The authors also provided guidelines and procedures to improve the use of CNNs in operational contexts and suggested some directions for future research. They emphasized the potential of deep learning techniques for crop diseases identification and their promising implications for the development of new agricultural tools that could contribute to more sustainable agricultural practices and greater food production security.

- "Breast lesion classification using features fusion and selection of ensemble ResNet method" by Kılıçarslan G, Koç C, Özyurt F, Gül Y. [15]

The research paper "Breast lesion classification using features fusion and selection of ensemble ResNet method" by Kılıçarslan G, Koç C, Özyurt F, and Gül Y focuses on breast lesion classification using an ensemble ResNet method. The study achieved a best accuracy of 84.9% for normal, malignant, and benign classification with the ALL-ResNet NCA model. Experimental studies highlighted the effectiveness of MR-MR, NCA, and Relieff in this context.

The paper addresses the importance of accurate breast cancer diagnosis and monitoring despite technological advancements. It proposes an ensemble classification approach that evaluates various machine learning algorithms like logistic regression, SVM, multilayer perceptron network, decision trees, support vector machines, K-nearest neighbor classifiers, and Naïve Bayes algorithms. The study then selects the top three classifiers based on their F3 score and combines them using a majority-based voting mechanism for improved performance.

Furthermore, the research emphasizes the significance of false negatives in breast cancer classification by utilizing the F3 score. The proposed ensemble method, particularly the majority-based voting mechanism, demonstrated superior performance compared to existing algorithms for Wisconsin Breast Cancer Dataset (WBCD), achieving a notable accuracy of 99.42%.

In summary, this research paper introduces an innovative approach to breast lesion classification by leveraging ensemble methods and ResNet models to enhance accuracy in distinguishing between normal, malignant, and benign cases.

- "Fabric defect detection and classification using modified VGG network" by R. S. Sabeenian, Eldho Paul & C. Prakash [16]

The research paper "Fabric defect detection and classification using modified VGG network" by R. S. Sabeenian, Eldho Paul, and C. Prakash focuses on developing a deep learning framework for detecting various fabric types and classifying defects using artificial intelligence. The study utilizes a modified VGG network to enhance the accuracy of fabric defect detection and classification.

The dataset used in the research consists of five different fabric defects, including broken pick defects, pattern defects, weft yarn deformity, soiled fabrics, and others. By leveraging the capabilities of the modified VGG network, the study aims to improve the identification and classification of these fabric defects, contributing to more efficient quality control processes in textile manufacturing.

The authors, R. S. Sabeenian, Eldho Paul, and C. Prakash, have been actively involved in this research endeavor. Eldho Paul, an Assistant Professor at Christ University, has contributed significantly to this study on fabric defect detection and classification using deep learning techniques.

In summary, this research paper presents a novel approach to fabric defect detection and classification by employing a modified VGG network, showcasing the potential of deep learning in enhancing quality control processes within the textile industry.

2.1.1 THEMES DISCOVERED

The literature on image classification prominently features the utilization of deep neural network models such as CNN, ResNet-50, and InceptionV3, showcasing their effectiveness in various image classification tasks across different domains.

Many of the researches have involved the strategy of Transfer Learning. This approach aims to enhance model performance and generalization capabilities. Transfer learning is a machine learning technique where a model trained on one task is repurposed or fine-tuned for another related task. Instead of starting the learning process from scratch, transfer learning leverages knowledge gained from solving one problem and applies it to a different, but related, problem.

In the context of deep learning and neural networks, transfer learning involves using pre-trained models that have been trained on large-scale datasets, such as ImageNet, which contains millions of labeled images across thousands of categories. These pre-trained models have already learned to extract meaningful features from images and can capture general patterns and representations that are useful for a wide range of image classification tasks.

Overall, transfer learning offers a powerful approach for leveraging existing knowledge and expertise captured in pre-trained models to address new and diverse image classification tasks, making it a popular choice in many research papers across various domains.

Ensemble methods, where multiple models are combined to improve classification accuracy, are also gaining attention in the literature. By aggregating predictions from diverse models, ensemble techniques aim to mitigate individual model biases and enhance overall classification performance.

2.1.2 GAPS IDENTIFIED

A significant gap exists in the literature regarding grocery product classification, particularly within the context of Indian products. While several studies have focused on image classification tasks in various domains, there is a notable absence of research specifically targeting Indian supermarket products.

The lack of specialized datasets for Indian product classification underscores a critical gap in understanding the unique characteristics of Indian market products. Existing datasets predominantly comprise products from foreign markets, which may not accurately represent the diverse range of Indian grocery items.

Furthermore, the absence of studies focusing on the challenges and nuances associated with Indian grocery merchandise classification highlights the need for tailored solutions that consider cultural factors, regional variations, and specific product categories prevalent in the Indian market. Addressing these gaps is essential for developing effective product classifiers tailored to the Indian retail landscape and improving overall classification accuracy and efficiency.

2.2 PROBLEM STATEMENT

Following the trend of automated systems, the spread of modern supermarkets with self-checkouts and automated huge warehouses has transformed the shopping experience, providing unrivaled convenience and access to a vast range of products. Despite these technical advances, the lack of intelligent solutions for autonomous product categorization in retail contexts remains a key barrier, particularly in sectors such as automated retail warehouse storage and self-checkout terminals.

This problem is aggravated by a scarcity of specialist product classifiers capable of accurately identifying Indian food items. The absence of such classifiers not only reduces operational efficiency, but also causes inefficiencies in both customer shopping experiences and warehousing operations. In a country like India, with its enormous population and predominantly manual-operated checkout systems, this problem is exacerbated, resulting in long lines and potential product mismanagement owing to human errors.

The existing literature on picture categorization has primarily focused on generic datasets that include products from foreign markets such as the United States and the United Kingdom. However, there is a considerable gap in the research about the classification of Indian grocery products. This gap emphasizes the need for bespoke solutions that may accurately identify Indian grocery items while taking into consideration their distinct characteristics and cultural subtleties.

To close this gap, this research article presents a thorough study titled "Deep Supermarket: Transfer Learning Approach to Classify Indian Supermarket Products."

The fundamental goal of this project is to create robust deep learning models that can reliably classify Indian supermarket goods utilizing transfer learning techniques. The building of a bespoke picture dataset with 15 separate classes and a total of 37,310 photos serves as the foundation for this research. This unique dataset is required to aid model training and evaluation for Indian market applications because no current datasets containing Indian grocery/supermarket product photos are available.

The study investigates the efficacy of various deep learning models in categorizing Indian retail goods. Initially, customized Convolutional Neural Network (CNN) and VGG19 models are created, however their performance is unsatisfactory. The paper then looks into transfer learning, using pre-trained models such as EfficientNetB7, InceptionResNetV2, DenseNet169, and DenseNet201. Transfer learning models outperform non-transfer learning models, demonstrating their capacity to use prior information to improve classification accuracy.

Moving on, this study project will delve deeper into the dataset, examine its properties, and address the issue of class imbalance. Furthermore, the study will evaluate model architectures, compare their performance using empirical results and graphical analysis, and give suggestions on the best models for the specific Indian supermarket products dataset.

This study project not only addresses a vital vacuum in the literature, but it also has important implications for optimizing retail operations and consumer experiences in the Indian market. This study intends to contribute to the growth of intelligent retail systems by providing tailored solutions and using deep learning techniques, in line with the emerging trend of reducing human involvement and increasing operational efficiency.

2.3 CHALLENGES ADDRESSED

Absence of Specialized Dataset:
The research focuses on the issue of the lack of a specialized dataset specifically designed to meet the specific needs of categorizing Indian retail merchandise. The current databases mostly concentrate on products originating from foreign markets such as the US and the UK, neglecting to encompass the wide variety of articles available in Indian superstores. This work addresses the lack of datasets specifically created for classifying Indian grocery products by constructing a new unique image dataset. The collection consists of 37000+ images, divided into 15 distinct classes.

Lack of Indian Market-Specific Image Classifier:
Another challenge tackled by this research is the lack of Indian market-specific image classifiers capable of accurately categorizing the diverse range of products found in

Indian superstores. Current image classification literature focuses on products available in their own countries (specifically the US and UK) and do not take into account Indian grocery items, which may result in suboptimal performance when these same models are used to classify Indian supermarket products. By exploring various deep learning architectures and leveraging transfer learning techniques, this study aims to develop robust classifiers tailored to the Indian supermarket landscape, addressing the need for accurate and culturally relevant product classification models.

Need for Automated Product Categorization:
With the advent of future retail stores aiming for minimal human interaction, there is a pressing need for intelligent systems capable of autonomously categorizing products. Traditional retail operations heavily rely on manual processes for product categorization, leading to inefficiencies and potential errors. By developing advanced deep learning models for Indian supermarket product classification, this research addresses the need for intelligent technologies that can streamline retail operations, enhance customer experiences, and pave the way for the future of autonomous retail environments in India market places.

2.4 RESEARCH OBJECTIVES

Custom Dataset Creation:
The primary objective of this study is to construct a comprehensive dataset comprising the prominent products found in Indian supermarkets, along with their respective categories. This dataset will be utilized to train a Deep Neural Network (DNN) model for accurately classifying images of these products. This research aims to fill the gap caused by the lack of specialized datasets for classifying Indian supermarket products by creating a unique dataset that accurately represents the specific characteristics and variations of Indian products.

Model Development and Comparison:
Another key objective is to construct multiple DNN image classification models, starting with smaller models like Convolutional Neural Networks (CNN) and VGG19, and subsequently incorporating advanced Transfer Learning models such as InceptionResNetV2, DenseNet, and EfficientNet. Through this iterative process, the research seeks to explore the effectiveness of different deep learning architectures in accurately categorizing Indian supermarket products. By comparing the performance of these models, the study aims to identify the most suitable approaches and architectures for the research problem identified.

Fine-Tuning for Indian Market Context:
A crucial aspect of the research involves fine-tuning the developed models to effectively handle the unique characteristics and challenges presented by Indian market

products. This includes optimizing the models' architectures, hyperparameters, and training strategies to ensure robust performance in classifying Indian supermarket items. By fine-tuning the models for the Indian market context, the study aims to enhance their accuracy and adaptability to the specific requirements of Indian retail environments and products.

Performance Evaluation and Comparative Study:
The research objectives also include the evaluation and comparison of each model's performance in accurately categorizing Indian supermarket products. Through training and test scores evaluation and analysis, the study seeks to assess the efficacy of different deep learning architectures and transfer learning techniques in addressing the classification challenges inherent in Indian retail settings. By quantitatively and qualitatively evaluating the models' performance, the research aims to provide valuable insights into the strengths and limitations of each approach.

Chapter 3

System Architecture and Design

3.1 OVERVIEW OF DEEP LEARNING

Deep learning is a kind of machine learning that use neural networks to simulate the structure and function of the human brain in order to model and solve intricate issues. Its popularity has increased substantially as a result of advancements in processing capability and the accessibility of extensive datasets. Deep learning methods, which utilize artificial neural networks (ANNs), are specifically designed to acquire knowledge from large datasets without the requirement of manual feature engineering.

Important aspects of deep learning are:

• Neural Networks: Deep learning is based on neural networks, which consist of interconnected nodes (neurons) organized in layers. These networks have the ability to acquire intricate patterns and connections in data by uncovering hierarchical properties.

• Applications: Deep learning is extensively utilized in several domains, including image identification, natural language processing, speech recognition, and recommendation systems. Commonly used architectures encompass Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Deep Belief Networks (DBNs).

• Training: The process of training deep neural networks usually necessitates substantial quantities of data and computational resources. The combination of cloud computing and specialized hardware such as Graphics Processing Units (GPUs) has increased the availability of deep network training.

• Advantages: Deep learning has advantages such as efficient processing of unstructured data, identification of hidden relationships, the ability to train without supervision, and the ability to handle volatile data.

Deep learning presents several obstacles, including substantial processing demands, the necessity for extensive labeled datasets, interpretability concerns arising from black-box models, and the risk of overfitting. Deep learning is a potent technique in machine learning that utilizes neural networks to address intricate tasks by autonomously acquiring knowledge from data. The applications of this technology encompass a wide range of fields, facilitating progress in the automation and decision-making procedures.

3.1.1 INTRODUCTION TO DEEP NEURAL NETWORKS

Neural networks are machine learning models that draw inspiration from the structure and functioning of the human brain. They are composed of interconnected nodes that collaborate to tackle intricate issues. These networks consist of multiple layers of nodes, which include input, hidden, and output layers. Each node in these levels processes data and transfers it to the next layer using weights and thresholds.

Deep Neural Networks (DNNs) are a very effective category of machine learning models that draw inspiration from the structure and functioning of the human brain. In recent years, Deep Neural Networks (DNNs) have experienced a significant surge in popularity and achievements in diverse fields such as image and audio recognition, natural language processing, and medical diagnosis. Comprehending the basic principles of Deep Neural Networks (DNNs) is crucial for understanding why these were chosen to for the purpose of this research.

BUILDING DESIGN AND CONSTRUCTION

Deep neural networks are composed of layers of artificial neurons, also known as nodes or units, structured into different layers: input layer, hidden layers, and output layer. Every layer is made up of numerous neurons, with connections between neurons in neighboring layers being defined by adjustable parameters called weights and biases. Neural networks can vary in depth, with some having more hidden layers than others, setting them apart from shallow neural networks.

FUNCTIONING AND LEARNING

DNN can be thought of as a system that processes information through forward and backward propagation, guided by a mathematical concept known as gradient descent optimization. Input data is passed through the network during the forward pass. Computation flows through layers, where each neuron calculates a weighted sum of inputs and applies an activation function to introduce non-linearity.

The activation function allows the network to grasp intricate patterns and connections in the data, thereby improving its ability to represent information. Activation functions commonly used are the rectified linear unit (ReLU), sigmoid, and hyperbolic tangent (tanh).

After the forward pass, the network's output is evaluated against the ground truth labels using a predefined loss function to measure the difference between predicted and actual outputs. The main goal of training a DNN involves reducing the loss function by modifying the model's weights and biases through backpropagation in an iterative manner.

BACKPROPAGATION AND TRAINING

Backpropagation plays a crucial role in updating the network's parameters by utilizing the calculated gradients of the loss function in relation to each parameter. Through a process of continuously fine-tuning the weights and biases to reduce the loss, the network improves its predictive accuracy gradually.

When training a DNN, the usual process includes inputting batches of data into the network, calculating the loss, conducting backpropagation to adjust the parameters, and iterating through this cycle for numerous epochs. Training efficiency and effectiveness are influenced by several factors such as the optimization algorithm chosen, learning rate, batch size, and regularization techniques.

APPLICATIONS AND ADVANCEMENTS

Deep neural networks have shown impressive results in various fields such as image classification, object detection, speech recognition, language translation, and drug discovery. Advancements in DNN architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer models, have significantly enhanced their capabilities and scalability. This has led to significant progress in computer vision, natural language understanding, and autonomous systems.

Hence Deep neural networks are a versatile and powerful class of machine learning models that mimic the complex processing mechanisms of the human brain. It is essential to grasp the fundamental principles of DNNs, such as their structure, operation, and training methods, to fully utilize their capabilities in solving complex real-world issues and fostering innovation in various fields.

3.1.2 IMAGE CLASSIFICATION USING DNN

Deep Neural Networks (DNNs) are widely recognized as a leading method for image classification tasks because of numerous key benefits:

HIERARCHICAL FEATURE LEARNING

Deep Neural Networks are proficient at automatically acquiring hierarchical representations of input data. When it comes to image classification, the lower layers of the network focus on capturing basic features like edges and textures, while the higher layers specialize in learning more intricate and abstract features related to particular object categories. Through hierarchical feature learning, DNNs can effectively distinguish between various object classes using their visual characteristics.

NON-LINEARITY AND FLEXIBILITY

Discussing the non-linear nature and flexibility of activation functions in DNNs, they enable the model to effectively capture intricate relationships and patterns in image data. This adaptability allows DNNs to grasp complex visual concepts and variations by learning highly nonlinear decision boundaries, making them ideal for such tasks.

SCALABILITY AND CAPACITY

When it comes to scalability and capacity, DNN architectures, especially convolutional neural networks (CNNs), are optimized for efficiently processing vast amounts of image data. CNNs use shared weights and local connectivity patterns to decrease the number of parameters, allowing them to be applied to high-resolution images and extensive datasets. In addition, the deep neural networks have the ability to capture intricate, high-dimensional features, which improves their capability to address challenging image classification tasks.

TRANSFER LEARNING AND PRE-TRAINED MODELS

Deep Neural Networks (DNNs) can be fine-tuned for specific image classification tasks using pre-trained models from large-scale image datasets like ImageNet, requiring only small amounts of task-specific data. This approach utilizes the features learned by a pre-trained model, allowing researchers to achieve top performance on new datasets while conserving computational resources and training time.

APPLICATION IN AUTOMATED STORES AND RETAIL WAREHOUSES

Automated stores and retail warehouses utilize DNN technologies in different ways to boost operational efficiency, optimize inventory management, and enhance customer experiences:

Identifying and categorizing products using DNNs helps automate the process based on visual attributes, which is beneficial for inventory management in retail settings. Utilizing DNN-based image classification systems enables automated stores to precisely identify products on shelves and streamline inventory management procedures without human involvement.

Self-Checkout Systems: Deep neural networks drive self-checkout systems in retail stores through automated product recognition and payment processing. Utilizing computer vision algorithms, these systems can identify items placed by customers on checkout counters, eliminating the necessity for traditional barcode scanning and cashier-assisted transactions. Deep neural networks guarantee rapid and precise product recognition, which helps streamline the checkout process and minimize customer wait times.

Utilizing DNNs to analyze customer behavior and preferences from historical data and real-time interactions results in personalized product recommendations and dynamic pricing strategies. Utilizing sophisticated machine learning algorithms, automated stores can customize promotional offers and pricing changes for each customer, ultimately increasing sales and customer satisfaction.

Utilizing deep neural networks (DNNs), historical sales data, market trends, and external factors are analyzed to predict product demand and enhance inventory levels in retail warehouses. Automated stores can maintain optimal stock levels by accurately forecasting future demand, which helps prevent stockouts and lower excess inventory costs.

Overall, DNNs are essential for driving automated stores and retail warehouses through capabilities such as automated product recognition, self-checkout systems, personalized shopping experiences, and efficient inventory management. With their capacity to grasp intricate visual patterns and apply knowledge across various image datasets, they have become essential assets for reshaping the retail industry and improving operational effectiveness in the digital era.

3.2 OVERVIEW OF TRANSFER LEARNING

Transfer Learning (TL) involves utilizing knowledge acquired from solving a particular problem and applying it to a related but different problem. Transfer learning has become a valuable method in deep neural networks (DNNs) for enhancing model performance, decreasing training time, and handling data scarcity problems. Delving into the complexities of transfer learning is crucial for understanding why this approach was chosen to train the models for this research.

3.2.1 INTRODUCTION TO PROCESS OF TRANSFER LEARNING

The process of Transfer Learning involves the following key elements:

UTILIZATION OF PRE-TRAINED MODELS

Utilizing pre-trained models is the initial step in transfer learning. These models have been trained on extensive datasets for general tasks like image classification or language understanding. Pre-trained models act as an initial step, offering a base of knowledge acquired through thorough training on various datasets.

FEATURE EXTRACTION

When using transfer learning, the first layers of pre-trained models are usually kept unchanged because they have already acquired basic features like edges, textures, and shapes that are important for various tasks. Extracting features entails retrieving the learned features from the pre-trained model, which act as informative representations of the input data.

FINE-TUNING

After extracting features, the following layers of the pre-trained model are adjusted or fine-tuned to better suit the specific characteristics of the target task or dataset. Adjusting the weights and biases of the network through additional training on the target data enables the model to learn task-specific features and relationships.

3.2.2 ADVANTAGES OF TRANSFER LEARNING

By utilizing knowledge from pre-trained models, transfer learning greatly decreases the amount of training time needed to excel in a new task. Utilizing pre-trained weights and fine-tuning a subset of the model parameters speeds up the training process convergence in transfer learning.

ENHANCED GENERALIZATION

Transfer learning helps improve generalization to new, unseen data by utilizing knowledge acquired from extensive and varied datasets during pre-training. Initiating the model with representations learned from related tasks allows transfer learning to assist the model in capturing underlying patterns and relationships that are relevant to the target task.

ADDRESSING DATA SCARCITY

When faced with a shortage of labeled training data, transfer learning offers a practical solution by enabling the model to leverage knowledge gained from plentiful data sources. Transfer learning helps reduce overfitting and enhances model performance by applying knowledge from tasks with abundant data to tasks with scarce data.

3.2.3 TRANSFER LEARNING OVER TRADITIONAL APPROACH

Transfer learning surpasses traditional training methods in various aspects:

FEATURE REUSE

Utilizing pre-trained models allows for the reuse of features, enabling the model to concentrate on learning task-specific features. This results in enhanced performance using less data.

UTILIZING PRE-TRAINED MODELS

Pre-trained models are valuable starting points for target tasks, as they already contain pertinent information from related tasks. This setup helps achieve quicker convergence and improved utilization of the available training data.

ADAPTABILITY TO TASK COMPLEXITY

Transfer learning smoothly adjusts to tasks of different complexity levels through the ability to fine-tune model parameters. By fine-tuning models, it is possible to capture detailed patterns and nuances that are unique to the specific task at hand, leading to better performance when compared to models trained from the beginning.

IN CONCLUSION

Transfer learning provides a practical and effective method for training deep neural networks across various tasks. Utilizing insights gained from pre-trained models, transfer learning speeds up training, enhances generalization, and tackles data scarcity problems, establishing itself as a widely used approach in various fields of study and real-world scenarios.

3.3 OVERVIEW OF THE MODELS USED IN RESEARCH

3.3.1 TRADITIONAL CNN MODEL

A Convolutional Neural Network (CNN) is an advanced deep learning model tailored for analyzing structured grid data, like images. CNNs are extensively utilized in image classification, object detection, and various computer vision tasks because of their capacity to efficiently capture spatial hierarchies and patterns within images.

The CNN is composed of multiple layers, each with a distinct role in feature extraction and classification as mentioned below.

Input Layer: At the input layer, raw image data is received, usually in the form of a pixel value grid.
Convolutional Layers: The application of convolution operations to the input image entails sliding a small filter (also known as a kernel) across the image to extract local features. Various filters detect distinct features in the input, like edges, textures, or shapes. The result of the convolution operation is a feature map that emphasizes specific features in various parts of the image.

Activation Function: Usually, a ReLU (Rectified Linear Unit) activation function is applied following each convolutional operation. ReLU introduces non-linearity into the network, enabling it to capture intricate patterns and relationships in the data.

Pooling layers: reduce the spatial dimensions of feature maps generated by convolutional layers while preserving essential information. Pooling operations commonly involve max pooling and average pooling, which preserve either the highest or mean value within each pooling region.

Flattening layers: Following multiple convolutional and pooling layers, the feature maps are transformed into a one-dimensional vector, which is then used as the input for the fully connected layers.

Fully Connected Layers: Referred to as dense layers, these layers establish connections between each neuron in one layer with every neuron in the subsequent layer. High-level feature extraction and classification are carried out by fully connected layers, building upon the features learned by the convolutional layers.

The final predictions or classifications are generated by the output layer. In a classification task, the number of neurons in the output layer aligns with the number of classes, and the choice of activation function is determined by the task's requirements (such as using softmax for multi-class classification).

In conclusion, CNNs use convolutional operations, activation functions, and pooling operations to extract hierarchical representations of input images, gradually learning more abstract features as information flows through the network. Through hierarchical feature learning, CNNs can efficiently classify images by capturing low-level visual patterns and high-level semantic information.

3.3.2 VGG19 ARCHITECTURE

Developed by the Visual Geometry Group (VGG) at the University of Oxford, VGG19 is a convolutional neural network (CNN) architecture. It is well-known for its simplicity and effectiveness in tasks related to classifying images. VGG19 expands upon the original VGG16 model by incorporating extra layers to enhance its capabilities.

The VGG19 model is composed of 19 layers (Figure 3.1), which explains its name. The layers are structured with a sequence of convolutional layers, max-pooling layers, and fully connected layers at the end for classification. Here is a general break down of this architecture:

Input Layer: This is the stage where the input images are introduced into the network. The standard input size is usually 224×224×3, with 3 denoting the RGB channels of the image.

Convolutional Blocks: The convolutional layers are structured into blocks, each consisting of several convolutional layers followed by a max-pooling layer. VGG19 consists of a total of 16 convolutional layers structured into 5 blocks.

Convolutional Layers: These layers involve filters that move across the input image, identifying features like edges, shapes, and textures. Every convolutional layer is succeeded by a Rectified Linear Unit (ReLU) activation function, which adds non-linearity to the network, enabling it to grasp intricate patterns.

Max-Pooling Layers: Following each series of convolutional layers, max-pooling layers are employed to decrease the spatial dimensions of the feature maps, thereby cutting down on the computational complexity of the network while preserving the key features.

Fully Connected Layers: The last layers of the network are fully connected layers, also referred to as dense layers. These layers utilize the high-level features extracted by the convolutional layers to conduct classification. In a multi-class classification task, the output layer usually contains softmax activation, which assigns probabilities to each class.

The functionality of VGG19 entails feeding an input image through the network. The image goes through a sequence of convolutions and pooling operations to extract features at various levels of abstraction. These characteristics are subsequently flattened and fed through fully connected layers for classification.

Throughout the training process, the network's parameters, such as the weights of the convolutional filters and the parameters of the fully connected layers, are fine-tuned using methods like stochastic gradient descent (SGD) or Adam optimization. The process of optimization involves fine-tuning these parameters to reduce the discrepancy between the predicted output and the actual labels in the training dataset.

After being trained, the VGG19 model is capable of classifying new images by running them through the network and collecting predictions from the output layer. Applying the softmax function to the output layer transforms the raw scores into probabilities, representing the likelihood of each class. The predicted label for the input image is determined by selecting the class with the highest probability.

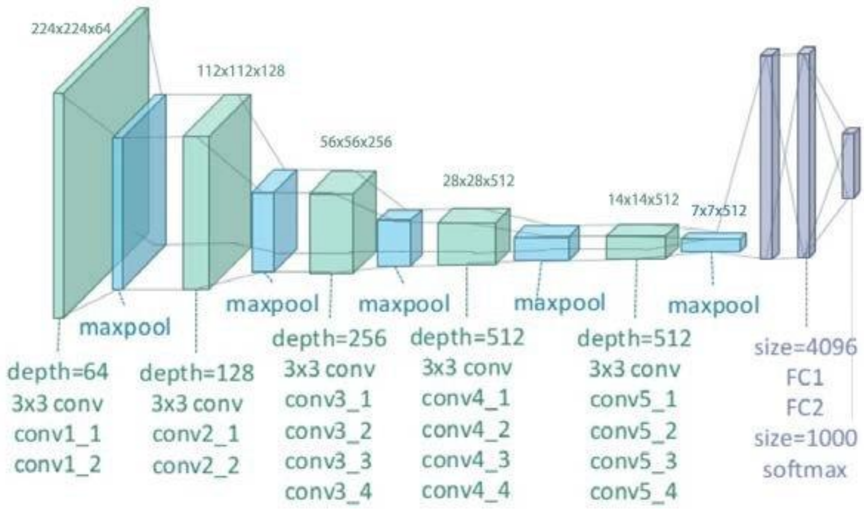

Figure 3.1: VGG19 Architecture

3.3.3 EFFICIENTNETB7 MODEL

EfficientNet, created by Google Brain, consists of a series of convolutional neural network designs that excel in performance and efficiency. EfficientNetB7 stands out as one of the largest models in the EfficientNet lineup, recognized for its exceptional results in image classification assignments.

EfficientNetB7 is a member of the EfficientNet model family, known for its deep neural network architecture. The EfficientNet architecture was introduced by Mingxing Tan and Quoc V. Le in their publication titled "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." This paper presents a new method for scaling models that outperforms traditional approaches by using fewer parameters while maintaining top performance.

The designation "B7" in EfficientNetB7 indicates the particular variant of the EfficientNet architecture. The EfficientNet family includes multiple variants ranging from B0 to B7, with B7 being the largest and most computationally intensive variant. These variations represent scaled versions of the foundational structure, incorporating greater depth, width, and resolution.

An important innovation of EfficientNet lies in its compound scaling method, which systematically scales the network depth, width, and resolution. This method guarantees that the model's capacity is enhanced effectively, leading to better performance without a substantial rise in computational expenses.

EfficientNetB7, as the largest variant in the family, is ideal for tasks that demand high accuracy and have the necessary computational resources for training and deployment of such a large model. Trained on extensive image datasets like ImageNet, it has acquired versatile and robust features that can be customized for different computer vision assignments.

EfficientNetB7 is commonly utilized in various applications like image classification, object detection, and semantic segmentation, where achieving top-notch performance is essential. EfficientNetB7 is known for its computational intensity, yet it provides remarkable accuracy and efficiency, which has led to its widespread adoption among researchers and practitioners tackling complex computer vision tasks.

3.3.4 INCEPTIONRESNETV2 MODEL

InceptionResNetV2 is a DNN model that integrates the Inception architecture and residual connections from ResNet. The Inception-v4 model was presented by Google researchers in their work titled "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning" as an expansion of the Inception family of models.

InceptionResNetV2 is constructed using the Inception architecture, which is distinguished by its frequent utilization of inception modules. The modules comprise many concurrent convolutional layers with varying filter sizes, followed by pooling procedures and concatenation of their outputs. This approach enables the model to efficiently capture features at various spatial scales and resolutions.

InceptionResNetV2 integrates residual connections, which were introduced in ResNet models, alongside the Inception modules. These connections facilitate the model in acquiring residual mappings, hence simplifying the training of deeper networks by alleviating the issue of vanishing gradients. InceptionResNetV2 delivers enhanced performance and scalability by leveraging the advantages of both architectures it combines, surpassing its predecessors.

InceptionResNetV2's design has numerous blocks of inception modules, which are subsequently followed by reduction blocks that decrease the spatial dimensions of the feature maps. An inception module consists of multiple concurrent convolutional layers with varying kernel sizes, such as 1x1, 3x3, and 5x5 convolutions, along with max pooling procedures. The presence of these parallel routes enables the model to effectively capture properties at various levels of abstraction.

In InceptionResNetV2, residual connections are incorporated within each block. This involves adding the output of one block to the output of another block, prior to going through a non-linear activation function. This approach enhances the movement of gradients during the training process and allows the model to acquire more efficient representations of the input data.

InceptionResNetV2 is commonly employed for tasks such as image classification, object identification, and picture segmentation, where attaining high accuracy and resilience is crucial. The model has undergone pre-training using extensive picture datasets such as ImageNet in order to acquire general features that can be further adjusted for specific tasks. InceptionResNetV2 is extensively utilized in both academic and industry applications due to its exceptional performance and efficiency.

3.3.5 DENSENET169 MODEL

DenseNet169 is a member of the DenseNet family of architectures and is classified as a deep neural network (DNN) model. Researchers at Facebook AI Research (FAIR) introduced it as an expansion of the original DenseNet model. DenseNet, also known as Densely Connected Convolutional Networks, is specifically developed to tackle the issue of vanishing gradient and enhance the utilization of features in deep networks. This is achieved by creating dense connections between layers.

DenseNet169 is distinguished by its dense connection topology, wherein each layer is connected to every other layer in a feed-forward manner. This connectivity approach allows for the transmission of features throughout the network and promotes the flow of gradients during training, resulting in enhanced learning dynamics and the reuse of features.

The structure of DenseNet169 of several dense blocks, each composed of a sequence of densely connected convolutional layers. Within each compact cluster, the output feature maps of all preceding layers are merged together and fed as input to following layers. The model benefits from this intricate network structure by effectively utilizing the reuse of features and acquiring more distinct representations of the input data.

DenseNet169 incorporates transition layers alongside dense blocks to regulate the expansion of feature maps and diminish the spatial dimensions of the feature maps between dense blocks. Transition layers commonly comprise a batch normalization layer, succeeded by a 1x1 convolutional layer and a 2x2 average pooling layer. These methods aid in compressing the feature maps and enhancing computing efficiency while maintaining crucial information.

The primary benefit of DenseNet169 is its capacity to promote the reuse of features and enhance the flow of gradients by establishing dense connections between layers. DenseNet169 achieves enhanced performance on several tasks, such as image classification, object detection, and picture segmentation, by establishing dense connections between each layer and all other layers. This lets the model to acquire more resilient representations of the input data.

Typically, during the training process, DenseNet169 is trained using gradient-based optimization methods such stochastic gradient descent (SGD) or Adam. The model is initialized with randomly assigned weights or pre-trained using extensive picture datasets such as ImageNet to acquire general features. One can apply fine-tuning to customize the model for certain tasks or domains by modifying the weights of individual layers or fine-tuning the entire network.

In summary, DenseNet169 is a very effective and efficient deep learning framework that has gained significant popularity in both academic research and real-world applications because of its exceptional performance and resilience.

3.3.6 DENSENET201 MODEL

DenseNet201 is a member of the DenseNet family of architectures and is a deep neural network (DNN) model. The extension of the original DenseNet model was introduced by researchers at Facebook AI Research (FAIR). DenseNet, also known as Densely Connected Convolutional Networks, is specifically created to tackle the issue of vanishing gradients and enhance the utilization of features in deep networks. This is achieved by constructing dense connections between layers.

The architecture of DenseNet201 closely resembles that of other DenseNet models, but it exhibits greater depth and complexity. The structure is composed of several dense blocks, with each block including a sequence of densely connected convolutional layers. Within each compact cluster, the resultant feature maps from all previous levels are combined and fed as input to the following layers. The model benefits from this intricate network structure as it can effectively utilize features several times and acquire more distinctive representations of the input data.

DenseNet201 has transition layers with dense blocks. These transition layers serve the purpose of managing the expansion of feature maps and diminishing the spatial dimensions of the feature maps that exist between dense blocks. Transition layers often comprise a batch normalization layer, succeeded by a 1x1 convolutional layer and a 2x2 average pooling layer. These methods aid in compressing the feature maps and enhancing computing efficiency while maintaining crucial information.

The primary benefit of DenseNet201 is its capacity to promote feature reuse and enhance the flow of gradients by establishing dense connections between layers. DenseNet201 achieves enhanced performance on several tasks, such as image classification, object detection, and image segmentation, by establishing dense connections between each layer and all other layers. This lets the model to acquire more resilient representations of the input data.

DenseNet201 is commonly trained using gradient-based optimization methods like stochastic gradient descent (SGD) or Adam throughout the training process. The model is initialized with randomly assigned weights or pre-trained on extensive picture datasets such as ImageNet to acquire general features. One can apply fine-tuning to customize the model for certain tasks or domains by modifying the weights of individual layers or fine-tuning the entire network.

In general, DenseNet201 is a potent and effective deep learning framework that has gained significant popularity in both academic research and real-world applications because of its exceptional performance and resilience. Its heightened depth and intricacy render it well-suited for demanding tasks that necessitate elevated levels of feature representation and discrimination.

3.4 METHODOLOGY

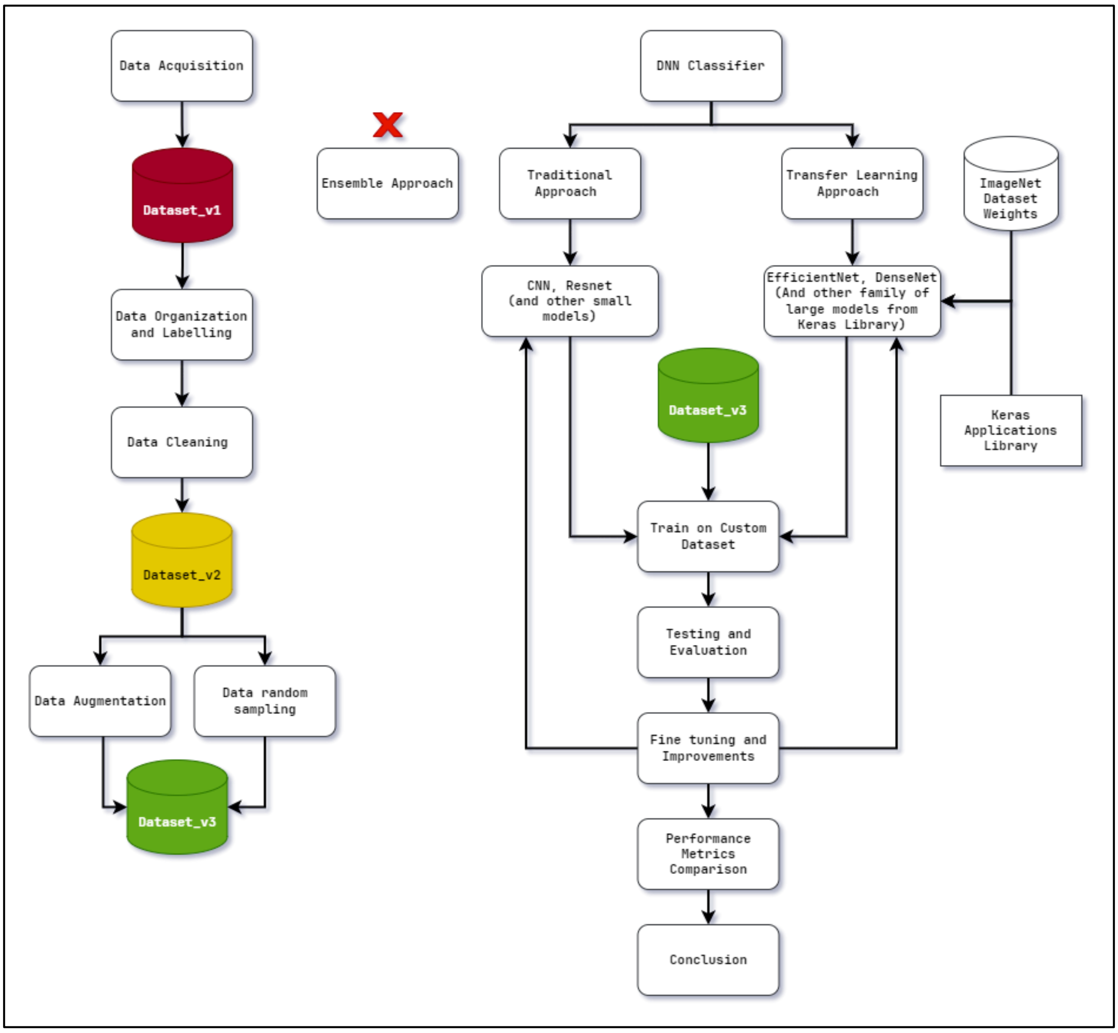

Figure 3.2 : Methodology of Study

3.4.1 DATASET CREATION METHODOLOGY

Data Acquisition: At the onset, images were gathered from online search engines to collect a wide range of visual data.

Dataset_v1: The creation of dataset_v1 (Figure 3.2) signified the compilation of raw images obtained directly from search engine results. The dataset's original structure, which included nested folders within class directories, needed to be adjusted to meet classification needs.

Data Organization and Labelling: After organizing and labelling the data, the images were sorted into their appropriate class folders to ensure correct labelling and improve the hierarchical structure for better classification preparation.

Data Cleaning: During this crucial stage, redundant and irrelevant images were carefully removed from the dataset to maintain data integrity and reduce the risk of contamination.

Dataset_v2: After completing the steps mentioned above, dataset_v2 (Figure 3.2) was obtained. It is a refined version of the dataset, free of unusable images but facing a notable class imbalance issue.

Data Augmentation: Efforts were made to tackle class imbalance by utilizing data augmentation techniques to increase the number of samples for classes with limited images, aiming to create a more balanced dataset.

Random Sampling: On the other hand, classes that had too many images used random sampling to reduce the number of images in order to create a more balanced distribution among classes.

Dataset_v3: Through the combination of data augmentation and random sampling, dataset_v3 (Figure 3.2) has been developed to improve model training by providing a more comprehensive and fair approach.

3.4.2 MODEL CREATION AND COMPARISION

DNN classifier: When formulating the model creation methodology, the main aim was to design an optimal Deep Neural Network (DNN) classifier customized for the classification of Indian supermarket products. Based on an in-depth review of the literature, three main methods for developing DNN classifiers were identified: Ensemble, Traditional, and Transfer Learning (TL) (Figure 3.2).

Traditional Approach: Ensemble methods, despite their strong theoretical foundation, were considered unfeasible because of their intricacy and the specific knowledge needed for their execution. In the past, creating architectures for smaller Convolutional Neural Network (CNN) and VGG models was done manually. Although these models provided a simple way to create and implement, their effectiveness in tackling the specific challenges that might be present in the study, like low-quality data, limited image quantities, class imbalances, and dataset cleanliness, was still unclear. However, for a thorough assessment of model performance, the conventional DNN models of CNN and VGG architectures are constructed also.

Transfer Learning Approach: Transfer Learning (TL) has emerged as a highly promising approach that closely aligns with the requirements of the research problem. Utilizing pre-trained weights from the ImageNet dataset, TL models architecture could be imported from the Keras Application library which accelerated model development, leading to decreased training time and the possibility of improved performance in comparison to conventional approaches. As a result, TL models were selected as the foundation for model creation approach, focusing on utilizing model architectures from the EfficientNet, DenseNet, and InceptionResNet families (Figure 3.2).

Train, Test, Evaluation on Custom Dataset: The chosen models underwent training on the custom dataset (dataset_v3) and were then assessed according to their performance metrics in testing and training situations.

Fine-tuning and Hyperparameter adjustments: Utilizing the findings from this assessment, adjustments were made to fine-tune the model and optimize its performance through hyperparameter adjustments and other approaches.

Performance Comparison and Conclusion: the methodology would conclude with a thorough evaluation of model performances, considering both testing and training scores, along with visual representations. Based on the findings, a conclusion was reached that addressed the research problem and offered insights into identifying the best model for classifying Indian supermarket products.

Chapter 4

Proposed Solution

4.1 IMPLEMENTATION

The research journey began with the creation of dataset_v2, serving as a vital base for future investigations. The investigation started by the development of traditional models, that included CNN and VGG19 architectures, using traditional training approaches. Despite best attempts to enhance these models, their performance was disappointing when trained on the unbalanced dataset_v2. The dataset's significant imbalance greatly hindered the effectiveness of the models, emphasizing the urgent requirement for balanced data. After dataset_v3 was introduced, carefully adjusted to address the issues of the previous version, the training of the conventional models was redone. Although there was a slight improvement in performance, the results were still subpar, resulting in ratings of poor to average quality.

The study's exploration into Transfer Learning started with implementing the EfficientNetB7 model with dataset_v2. At first, both versions of the Transfer Learning model displayed underwhelming performance, reflecting the difficulties seen with traditional models. With the introduction of dataset_v3, there was a significant improvement in the model's performance, highlighting the importance of balanced data for better classification results.

As the study delved deeper into more promising models of Transfer Learning, the study focused on large DNN architectures such as InceptionResNetV2, DenseNet169, and DenseNet201. These models demonstrated exceptional classification potential, easily handling the research's image classification of product images with their natural abilities. Although they achieved early success, a significant issue arose with overfitting, demonstrated by the large difference in scores between training and testing. Further Improvements were applied to the models to tackle the problem of overfitting which helped in improving these models' performance even more.

4.2 PROPOSED SYSTEM

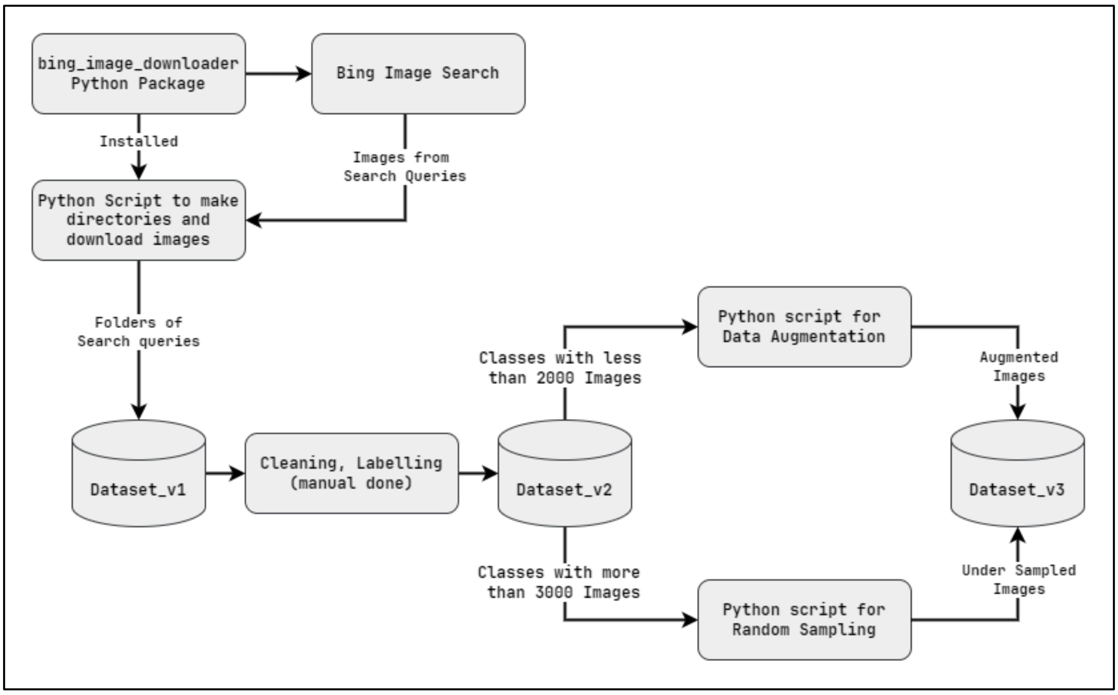

Figure 4.1: Proposed System for Dataset Creation

4.2.1 DATASET CREATION

In the pursuit of creating a custom dataset tailored to the classification of Indian supermarket products, a series of steps were undertaken to acquire, clean, and organize the dataset for subsequent model training. The dataset creation process can be delineated as follows:

4.2.2 INITIAL DATA ACQUISITION

Given the time constraints and logistical challenges, the decision was made to scrape images from online image search engines such as Google Images and Microsoft Bing. Python scripts were employed to automate the image scraping process, initially utilizing the google_images_download package. However, due to structural updates to the Google Images website, the bing_image_downloader package was later adopted as an alternative (Figure 4.1).

Online grocery websites like BigBasket and JioMart were browsered to compile a list of Indian supermarket products for image acquisition. Images were scraped and downloaded using the Python scripts, with each product class organized into respective folders based. The images were then downloaded into another folder (named the same as the search query / product name) inside these class folders.

The proposed system's (Figure 4.1) initiation began with conducting data acquisition tasks. A Python script was created to streamline the process of gathering images from internet search engines. The script made use of the bing_image_downloader package and utilized the downloader.download() method to trigger search queries to the Bing image search engine. Despite the specified limit of 80 images per product (search query), the number of images retrieved varied from 30 to 50, depending on the product's popularity. The images that were downloaded were structured into directories, where each category had folders for specific search terms and the related product images. The directory structure was as follows (this was incorrect directory structure):

Images / Class1 / searchQuery1 / product1 images
Images / Class1 / searchQuery2 / product2 images
Images / Class1 / searchQuery3 / product3 images
…
Images / Class n / searchQuery1 / product1 images
Images / Class n / searchQuery2 / product2 images
Images / Class n / searchQuery3 / product3 images

After completing the acquisition phase, the "dataset_v1" (Figure 4.1) was obtained from Google Drive, requiring a switch to Windows Explorer for additional data cleaning and organization.

4.2.3 DATASET CLEANING AND ORGANIZATION

Following data acquisition, the dataset underwent cleaning to remove irrelevant or unusable images manually. The folder structure of the dataset was reorganized to facilitate ease of training, with images categorized by product class. However, limitations in the Python package necessitated manual adjustments to achieve the desired folder structure.

The transition to Windows Explorer was initiated due to the constraints of Google Drive in managing files with identical names, which raised worries about possible loss of data from file replacement. The cleaning and organization process was carried out manually for about a week. The data was structured into the correct folder hierarchy, with each class containing all relevant product images. The directory structure now was as follows (this is optimal for creation of Image Generators which is used for image handling when training the models):
Images / Class1 / all product images of class1
Images / Class2 / all product images of class2
…
Images / Class n / all product images of class n

The middle directory of query-named folders was removed completed and all the images inside those folders were put into the parent class folder directly. This created the correct folder structure required for ease of training the Deep Neural Network models.

The final dataset version 2, although cleaned and well-structured, Analysis of the dataset revealed significant class imbalance (Figure 4.2). This resulted in less-than-ideal performance of initial models such as CNN, VGG, and EfficientNetB7. After careful examination, it was found that this disparity was especially noticeable in categories like snacks_and_sweets, as 4 of the 15 classes were making up more than 60% of the dataset (Figure 4.3). In order to tackle this problem, dataset_v2 was split into two sections according to the number of images (Figure 4.1).

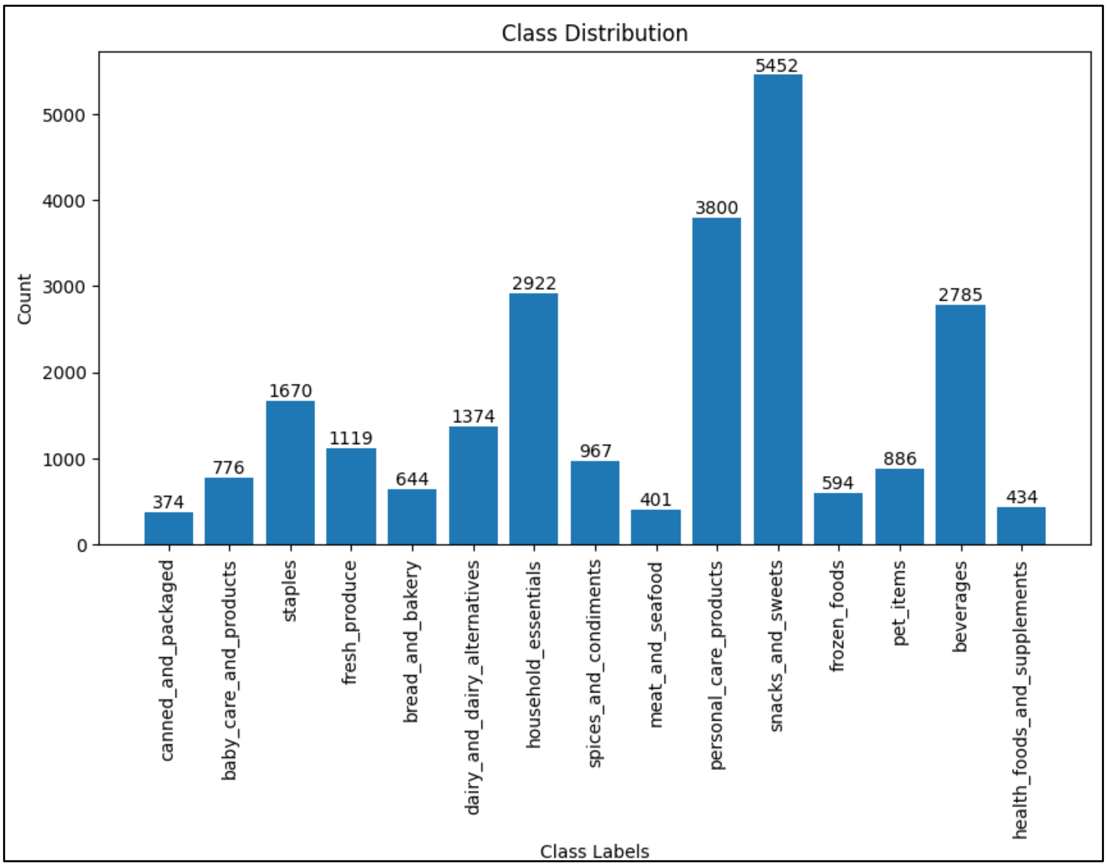

Figure 4.2: Class Distribution of dataset_v2

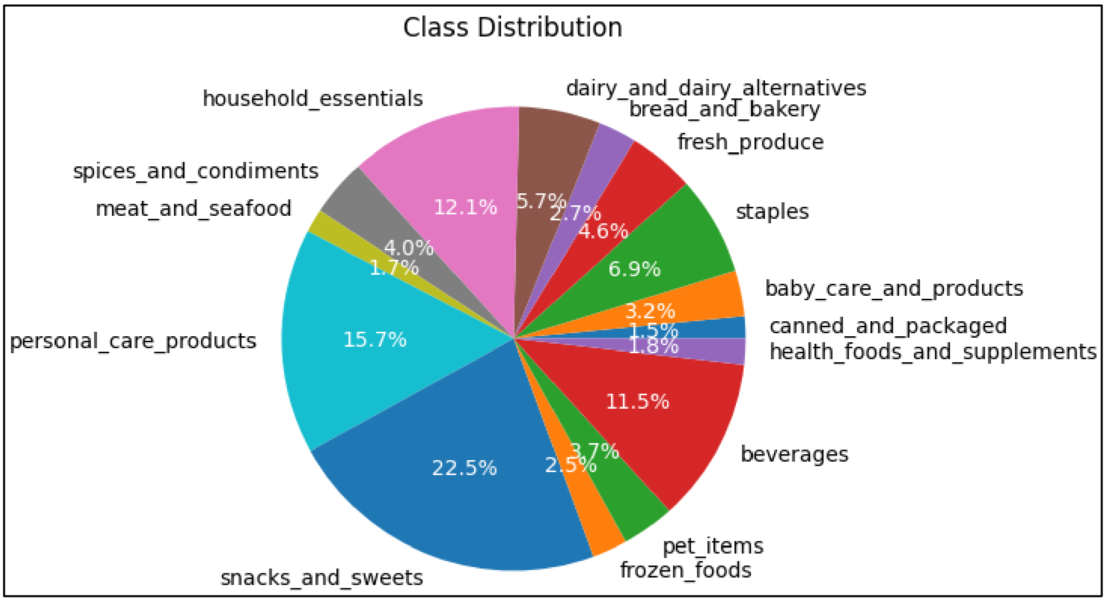

Figure 4.3 : Percentage Class Distribution of dataset_v2

4.2.4 CLASS BALANCING

To address this imbalance, dataset_v3 was created by balancing the number of images in each class using data augmentation techniques for classes with fewer images and random sampling for classes with excessive images (Figure 4.1).

A Python script was developed for data augmentation in classes with fewer than 2000 images. The script employed an image generator that included specific transformations like rotation, shear, and zoom to generate extra images and equalize the class distributions.

On the other hand, a distinct script was used for classes containing over 3000 images to randomly select 2700 images, reducing the impact of imbalance.

The culmination of these efforts resulted in the creation of dataset_v3 (Figure 4.1), characterized by improved balance across most classes, with image quantities ranging between 2000 and 3000 for the majority of classes (Figure 4.4). Notably, exceptions to these included classes such as fresh_produce, staples, dairy_and_dairy_alternatives, and meat_and_seafood, which exhibited slightly higher or lower image counts, thus providing a more equitable foundation (Figure 4.5). This balanced dataset was utilized for subsequent model training and comparison throughout the research.

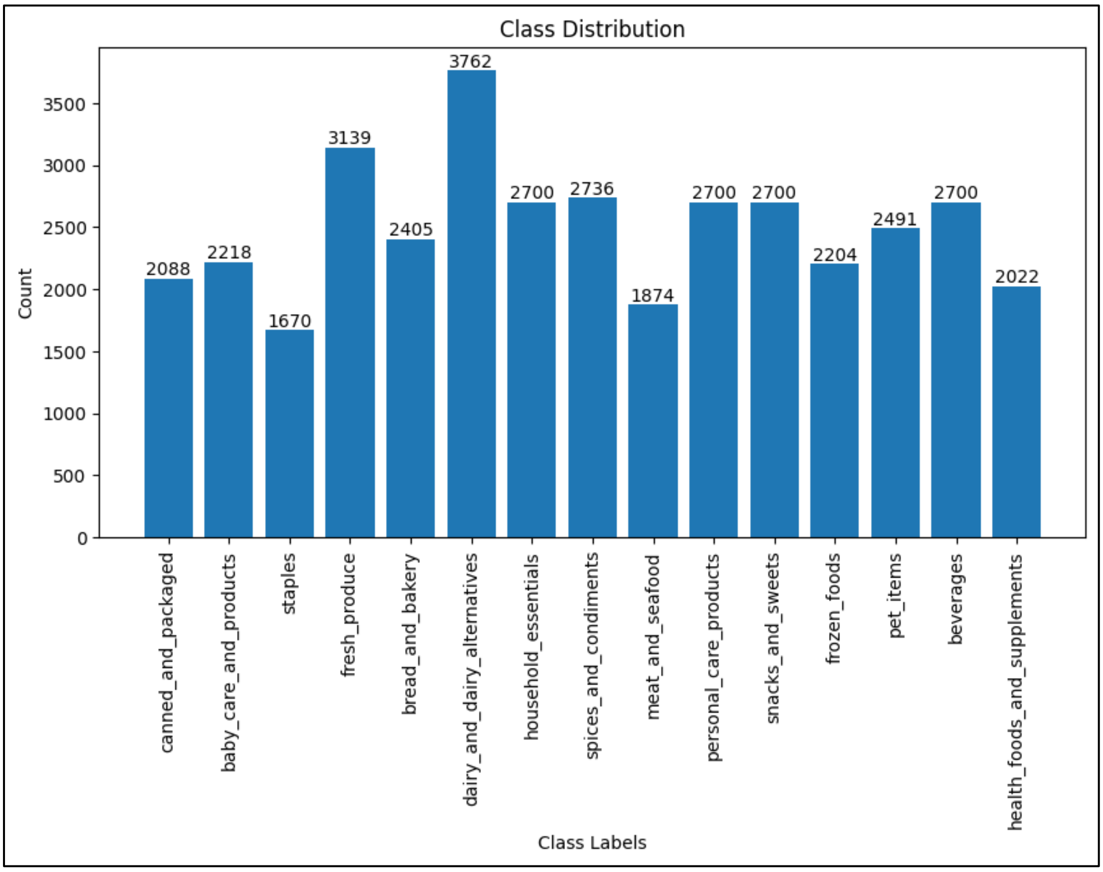

Figure 4.4 : Class Distribution of dataset_v3

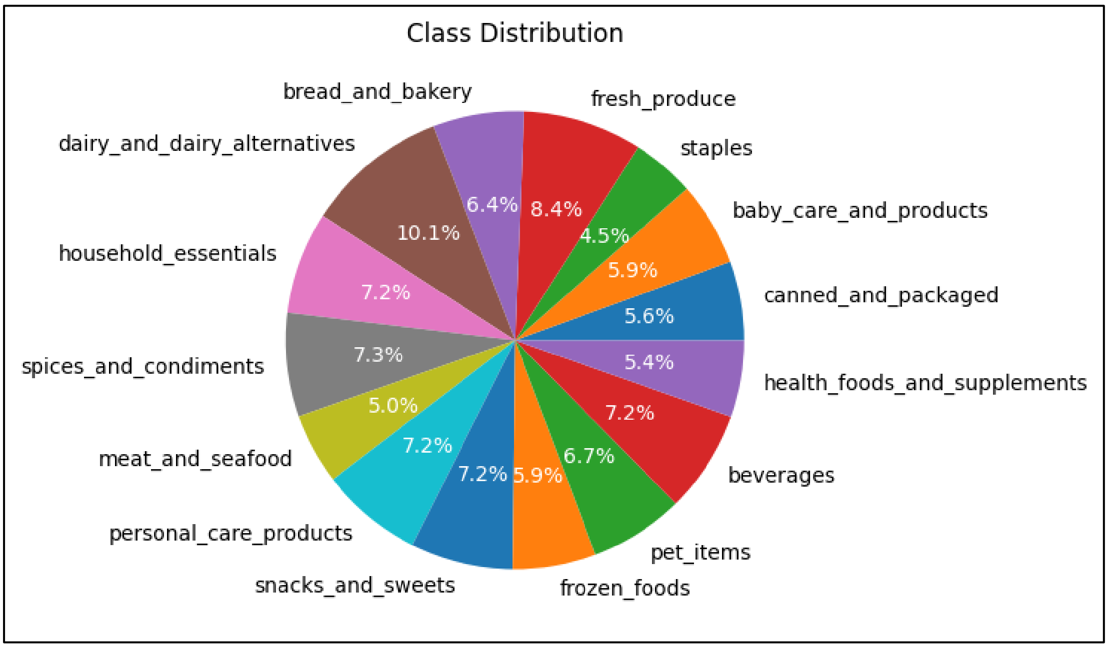

Figure 4.5 : Percentage Class Distribution of dataset_v3

4.2.5 PROPOSED SYSTEM FOR MODEL COMPARISION

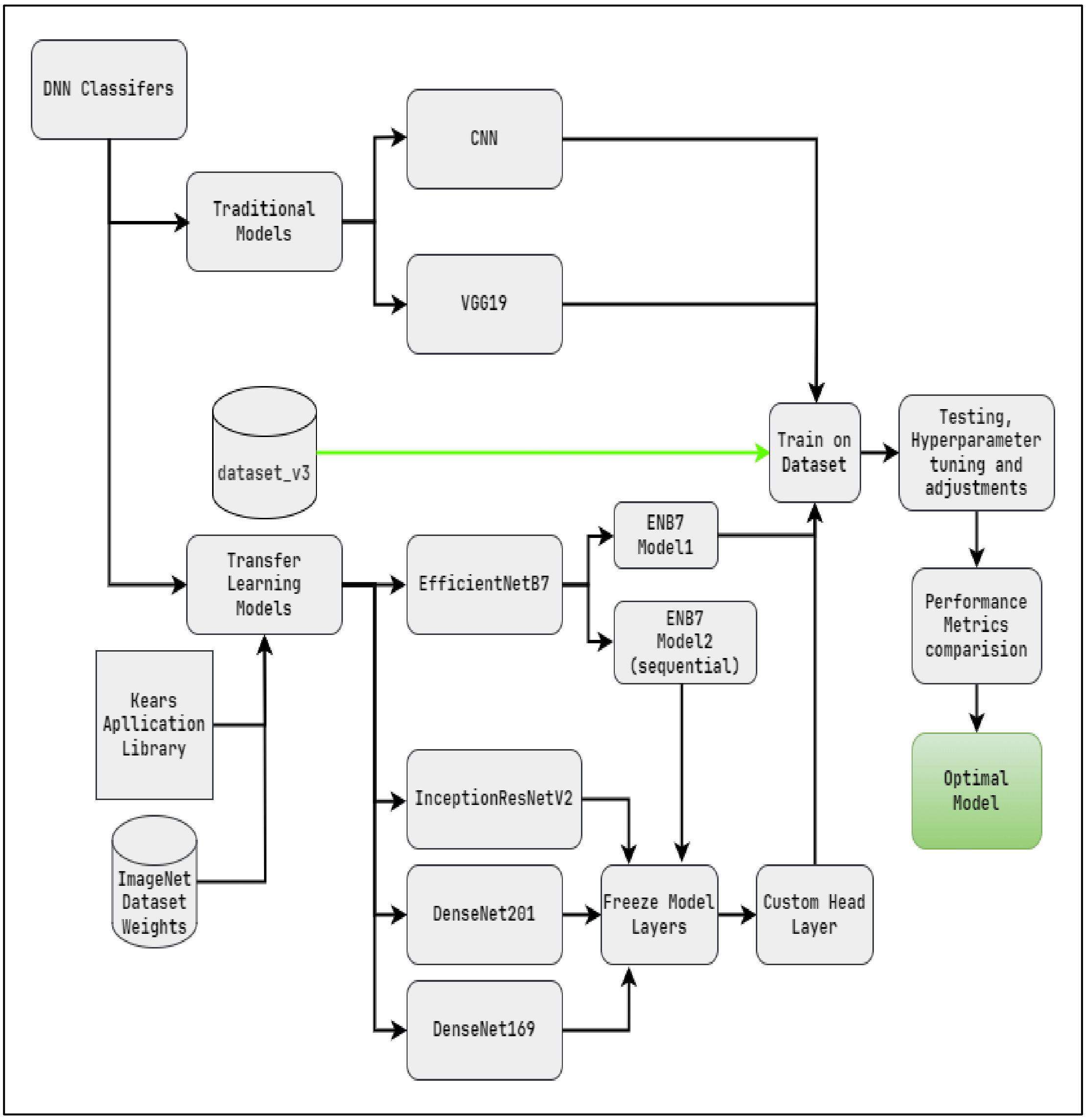

Figure 4.6: Proposed System for Model Creation and Comparison

In the proposed system for model comparison (Figure 4.6), the study adopted two distinct approaches: Traditional and Transfer Learning. First, CNN (Table 4.1) and VGG19 (Table 4.2) models were developed, layer by layer, using the Sequential method (models. Sequential ()). Subsequently, both models were compiled with the Adam optimizer, categorical cross-entropy loss function, and accuracy metric for training evaluation. Despite initial training on dataset_v2, the presence of class imbalances severely hindered their performance. To rectify this, the models were retrained on dataset_v3, incorporating hyperparameter tuning and additional features like learning rate scheduler and early stopping callbacks. Although dataset_v3 yielded improved performance, it remained suboptimal for the classification of product images which was required for the study.

CNN ARCHITECTURE

Table 4.1: Architecture of traditional CNN

BLOCK	LAYER	DESCRIPTION
Initial Block	Input Layer	The input layer specifies the shape of the input data, which in this case is a 3D tensor representing RGB images with dimensions 256x256 pixels.
Convolutional Blocks	1st Block	Consists of a Conv2D layer with 64 filters (kernels) of size 3x3 and ReLU activation function. Followed by a MaxPooling2D layer with a pool size of 2x2, which reduces the spatial dimensions of the feature maps by taking the maximum value within each pool.
	2nd Block	Similar to the first block but with an increased number of filters (128).
	3rd Block	Again, similar structure with 256 filters.
	4th Block	Similar structure with 512 filters.
Flatten Block	Flatten Layer	Flattens the output from the convolutional layers into a 1D vector to feed into the fully connected layers.
Fully Connected Block	1st Dense Layer	Consists of 1024 neurons with ReLU activation function, enabling complex feature extraction and representation
	Dropout Layer	Helps prevent overfitting by randomly dropping a fraction (50%) of neurons during training
	2nd Dense Layer (Output)	Has 15 neurons corresponding to the number of classes in the classification task (softmax activation function), producing probability distributions over the classes

VGG19 ARCHITECTURE

Table 4.2: Architecture of Traditional VGG19

BLOCK	LAYER	DESCRIPTION
1st block	conv1_1	First block consists of two convolutional layers with 64 filters each, and is followed by a max-pooling layer.
	conv1_2	
	max-pooling	
2nd block	conv2_1	Block 2 consists of two convolutional layers with 128 filters each, and is then followed by a max-pooling layer.
	conv2_2	
	max-pooling	

3rd block	conv3_1	Block 3 consists of four convolutional layers containing 256 filters each, with a subsequent max-pooling layer.
	conv3_2	
	conv3_3	
	conv3_4	
	max-pooling	
4th block	conv4_1	Block 4 consists of four convolutional layers containing 512 filters each, with a subsequent max-pooling layer
	conv4_2	
	conv4_3	
	conv4_4	
	max-pooling	
5th block	conv5_1	Block 5 consists of four convolutional layers containing 512 filters each, with a subsequent max-pooling layer
	conv5_2	
	conv5_3	
	conv5_4	
	max-pooling	
Fattening and Fully Connected Block	FC_1	Two layers are fully connected with 4096 neurons each and are then followed by ReLU activation functions. The dense layers function as effective feature extractors, capturing advanced representations of the input data
	FC_2	
	Dropout	Dropout layers are added after each fully connected layer with a dropout rate of 0.5 to prevent overfitting. This involves randomly dropping a fraction of the neurons during training.
Final Block	Final Layer	The final output layer is composed of a dense layer containing 15 neurons, which correspond to the number of classes in the classification task (15 classes for the research requirements) When the softmax activation function is used on the output layer, it generates a probability distribution across the classes. This feature enables the model to generate probabilities for each class, facilitating multi-class classification.

In the Transfer Learning approach, models were sourced from the Keras applications library, leveraging their pre-trained weights from the ImageNet dataset.

The EfficientNetB7, renowned for its extensive parameterization, was employed in two variants:

Model1 - a straightforward implementation without customization (addition of only and extra input and output layer)

Model2 - a sequential model with added custom head, Flatten, and dropout layers. (Keras model + sequential layers)

Despite their potential, the EfficientNetB7 models exhibited modest improvements over traditional models on both dataset_v2 and dataset_v3. Further modifications, including hyperparameter optimization and regularization techniques, yielded only marginal enhancements in performance.

ENB7 MODEL1 ARCHITECTURE

ENB7 Model1 uses the EfficientNetB7 architecture directly (without much changes) from the TensorFlow Keras applications library to create a custom model tailored for the particular image classification task of the study. Given below is the architectural break down of this model (Table 4.3)

Table 4.3: Architecture of ENB7 Model 1

MODEL PARTS	**DESCRIPTION**
Base Model Initialization	The code initializes the EfficientNetB7 model from the Keras applications library with the following parameters • Excludes the top (classification) layers of the pre-trained EfficientNetB7 model, allowing for customization. • Loads pre-trained weights trained on the ImageNet dataset, providing a strong foundation for feature extraction. • Specifies the input shape of the images to be processed by the model (256x256 pixels with 3 color channels). • Applies global average pooling to the output of the final convolutional layer, reducing the spatial dimensions to a single vector.
Custom Model Construction	A sequential model is constructed using the Sequential method from Keras library, which allows for the sequential stacking of layers. • The EfficientNetB7 base model is added as the first layer. This layer serves as a feature extractor and is set to non-trainable layers to preserve the pre-trained weights. • A flatten layer is added to convert the output of the base model into a one-dimensional vector.

	• A dense layer with 15 neurons and softmax activation is added as the output layer, representing the number of classes in the classification task.
Model Compilation	The model is compiled with the following parameters: • Adam optimizer for gradient descent optimization. • categorical cross-entropy loss as the loss function for multi-class classification tasks. • Accuracy as the evaluation metric to monitor model performance during training.

ENB7 MODEL2 ARCHITECTURE

ENB7 Model2 is a customized architecture derived from the EfficientNetB7 convolutional neural network (CNN) model, designed for a particular image classification task. EfficientNetB7 Version 2 functions in a manner akin to the original EfficientNetB7 model, but includes extra dense layers and dropout regularization to enhance performance and generalization.

The model utilizes pre-trained EfficientNetB7 base layers to extract features from input images, which are then processed through custom dense layers and dropout layers for transformation and regularization. Adjusting the final 20 layers (unfreezing some of the layers) enables the model to adapt to the specific classification task while leveraging the pre-trained knowledge from ImageNet. (Table 4.4)

Table 4.4: Architecture of ENB7 Model2

MODEL PART	**DESCRIPTION**
Pre-trained EfficientNetB7 base model	The code loads the pre-trained EfficientNetB7 model from the Keras applications library. This model serves as the base architecture for feature extraction, initialized with weights trained on the ImageNet dataset. *Parameters:* • Excludes the top (classification) layers of the pre-trained EfficientNetB7 model, allowing for customization. • Loads pre-trained weights trained on the ImageNet dataset for feature extraction. • Specifies the input shape of the images (256x256 pixels with 3 colour channels).

	• Applies global average pooling to the output of the final convolutional layer, reducing the spatial dimensions to a single vector.
Custom Sequential Model	A sequential model using the Keras Library's Sequential method, is constructed to incorporate additional layers on top of the pre-trained EfficientNetB7 base model. *Added layers are as follows:* • A Keras Layer is added to integrate the EfficientNetB7 base model as the first layer. This layer is set to non-trainable to preserve the pre-trained weights. • A flatten layer is added to convert the output of the base model into a one-dimensional vector. • Two dense layers with 512 and 256 neurons, respectively, are added for feature transformation. These layers utilize ReLU activation functions and apply L2 regularization with a regularization strength of 0.001. • Dropout layers are inserted after each dense layer to prevent overfitting by randomly dropping 50% of the neurons during training. • Finally, a dense output layer with softmax activation is added to produce class probabilities for the classification task.
Model Compilation	*The model is compiled with the following parameters:* Adam optimizer with a custom learning rate of 0.0002. Categorical cross-entropy loss, suitable for multi-class classification tasks. Accuracy is chosen as the evaluation metric to monitor model performance during training.
Fine-Tuning	The code also unfreezes the last 20 layers of the model to allow for fine-tuning. Batch normalization layers are excluded from fine-tuning to prevent destabilizing the model.

Conversely, the InceptionResNetV2 (Table 4.5), DenseNet169 (Table 4.6), and

DenseNet201 (Table4.7) models, renowned for their complex architectures, showcased remarkable performance from the outset. These models were constructed with by first freezing all of their layers and then making custom sequential models for each to make then optimized for the classification task of this research. After this, the custom head being attached to the base Keras models and compiled with the adam optimizer, categorical cross-entropy loss function, and accuracy metric for training evaluation. Following initial training on dataset_v3, these models displayed superior performance.

INCEPTIONRESNETV2 ARCHITECTURE

Table 4.5: Architecture of InceptionresNetV2

MODEL PART	**DESCRIPTION**
Base Model Initialization	The model code loads the InceptionResNetV2 model with pre-trained weights from the ImageNet dataset from the Keras Applications library. *Parameters:* • Excludes the fully connected layers at the top of the network, allowing for customization. • Initializes the model with pre-trained weights learned on the ImageNet dataset. • Specifies the input shape of the images (256x256 pixels with 3 colour channels).
Freezing Base Model Layers	All layers of the base model are set to non-trainable to prevent their weights from being updated during training. This step ensures that only the custom head layers are trained.
Custom Head Creation	The custom head layers are defined to be added on top of the base InceptionResNetV2 model. The output of the base model is passed through a global average pooling layer to reduce the spatial dimensions of the feature maps. A dense layer with 1024 neurons and ReLU activation is added to perform feature transformation. Finally, a dense output layer with softmax activation is added to produce class probabilities for the classification task.
Connecting Base Model	The base model's input and the custom head's output are connected to create the final model.

with Custom Head	This step creates a new model architecture where the input passes through the base model's layers and then through the custom head layers.
Model Compilation	*The model is compiled with the following parameters:* Adam optimizer with a learning rate of 0.001. Categorical cross-entropy loss function, suitable for multi-class classification tasks. Accuracy is chosen as the evaluation metric to monitor model performance during training.

DENSENET169 ARCHITECTURE

The breakdown of the architecture for the DenseNet169 implemented for the purposes of this research is as follows (Table 4.6):

Table 4.6: Architecture of DenseNet169

MODEL PART	**DESCRIPTION**
Base Model Initialization	The model code initializes the DenseNet169 model with pre-trained weights from the ImageNet dataset. By setting not importing the top layer, only the convolutional base of the DenseNet169 model is loaded, excluding the fully connected layers at the top. The input_shape parameter specifies the dimensions of the input images expected by the model.
Frozen Base Model Layers	After loading the pre-trained DenseNet169 model, the code freezes its layers. This prevents the weights of the base model from being updated during training, preserving the learned features from the ImageNet dataset.
Custom Head	A custom head is added to the base model to adapt it to the specific classification task for the research. The output of the base model serves as the input to the custom head. This output is stored in a variable. Global average pooling is applied to aggregate spatial information across the feature maps generated by the base

	model. This reduces the spatial dimensions to a single vector while retaining important feature information. Subsequently, a fully connected dense layer with 1024 units and ReLU activation is added to introduce non-linearity and further abstract feature representations. Finally, a dense output layer with 15 units and softmax activation is added to produce the predicted class probabilities. The number of units in the output layer corresponds to the number of classes in the classification task.
Model Compilation	The model is compiled using the Adam optimizer with a learning rate of 0.001. Categorical cross-entropy is chosen as the loss function to measure the difference between the predicted probabilities and the true labels. The accuracy metric is specified to monitor the performance of the model during training.

DENSENET201 ARCHITECTURE

The model used for this research paper is based on DenseNet201 model from the Kears Library with pre-trained ImageNet weights, incorporates a custom classification head, and compiles the model for training.

Table 4.7: Architecture of DenseNet201

MODEL PART	**DESCRIPTION**
Base Model Initialization	The code initializes the DenseNet201 model with pre-trained weights from the ImageNet dataset by importing the model's architecture from Keras Applications library's. By not importing the top layer, only the convolutional base of the DenseNet201 model is loaded, excluding the fully connected layers at the top. The input shape parameter specifies the dimensions of the input images expected by the model (256x256 pixels with 3 channels for RGB images).
Frozen Base Model Layers	After loading the pre-trained DenseNet201 model, the code freezes its layers. This ensures that the weights of the base model are not updated during training, preserving the learned features

	from the ImageNet dataset.
Custom Head	A custom head is added to the base model to tailor it for the specific research classification task. The output of the base model serves as the input to the custom head, stored in a variable. Global average pooling is applied to collapse the spatial dimensions of the feature maps generated by the base model into a single vector while retaining essential feature information. Subsequently, a fully connected dense layer with 1024 units and ReLU activation is added to introduce non-linearity and further abstract feature representations. Finally, a dense output layer with 15 units and softmax activation is appended to produce the predicted class probabilities.
Model Compilation	The model is compiled using the Adam optimizer with a learning rate of 0.001. Categorical cross-entropy is chosen as the loss function to quantify the discrepancy between the predicted probabilities and the true labels. The accuracy metric is specified to evaluate the performance of the model during training and validation.

Subsequent fine-tuning efforts (Figure 4.6), such as unfreezing lower layers (lower most 20) and integrating adaptive learning rate and regularization layers, and implementing additional callbacks (EarlyStopping() and ReduceLROnPlateau()) during the training significantly mitigated overfitting while enhancing overall performance. Among these models, the DenseNet family emerged as the most promising, with DenseNet201 exhibiting the most optimal performance for the research problem.

Finally, the performance data from all models were aggregated and compared, with visualization techniques employed to identify the most optimal model for the research problem. Through systematic evaluation and iterative refinement, the study aimed to identify the model best suited for the classification of Indian supermarket products which was the primary objective of this research.

Chapter 5

Results and Discussions

5.1 RESULTS

This research shed light on the efficacy of various models for the classification of Indian supermarket products. Through a meticulous methodology encompassing dataset creation, model construction, and performance evaluation, the study gained valuable insights into the strengths and limitations of different approaches.

The traditional model approach, where CNN and VGG19 architectures were developed, encountered significant challenges stemming from class imbalances within the dataset. Despite efforts to address these imbalances and fine-tune model parameters, the performance of traditional models remained suboptimal, indicating their limited suitability for the classification task for the study.

5.1.1 PERFORMANCE OF CNN

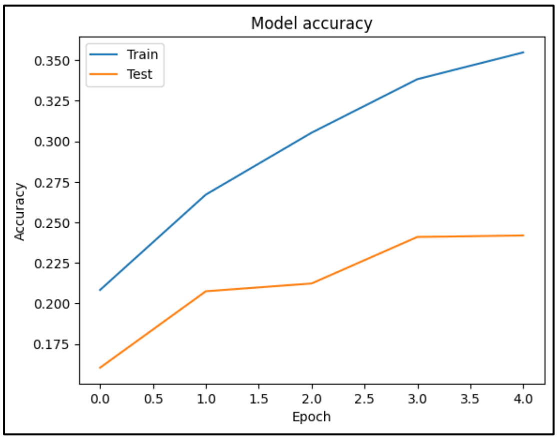

Figure 5.1: Accuracy of Traditional CNN

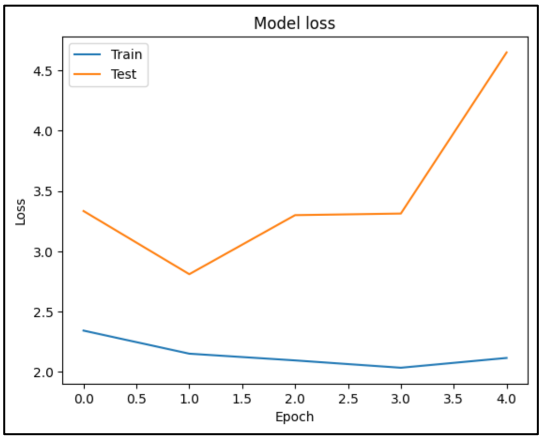

Figure 5.2: Loss of Traditional CNN

Table 5.1: Performance Scores of Traditional CNN

S NO.	PARAMETERS	SCORES
1	Training Accuracy	35.48 %
2	Testing Accuracy	24.19 %
1	Training Loss	2.1162
2	Testing Loss	4.3453

5.1.2 PERFORMANCE OF VGG19

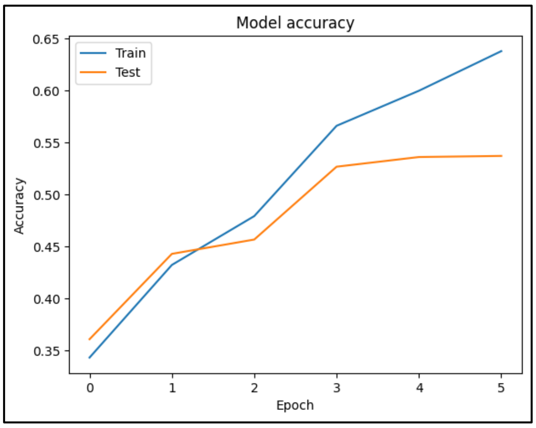

Figure 5.3: Accuracy of Traditional VGG19

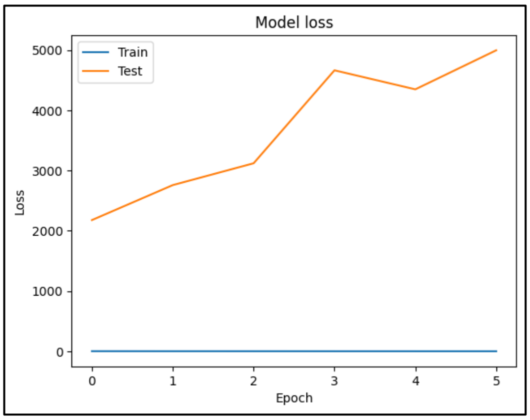

Figure 5.4: Loss of Traditional VGG19

Table 5.2: Performance Scores of Traditional VGG19

S NO.	PARAMETER	SCORES
1	Training Accuracy	63.81 %
2	Testing Accuracy	53.73 %
1	Training Loss	1.0536
2	Testing Loss	2541.41

On the other hand, Transfer Learning emerged as a promising strategy, leveraging pre-trained models of EfficientNetB7, InceptionResNetV2, DenseNet169, and DenseNet201. These models, with their complex architectures and pre-trained weights from the ImageNet dataset, showcased superior performance compared to traditional models. However, initial training revealed issues of overfitting, prompting further refinement through fine-tuning and regularization techniques.

5.1.3 PERFORMANCE OF EFFICIENTNETB7

ENB7 - MODEL1

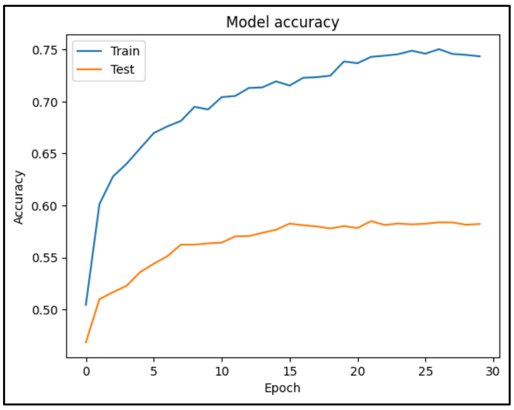

Figure 5.5: Accuracy of ENB7 - Model1

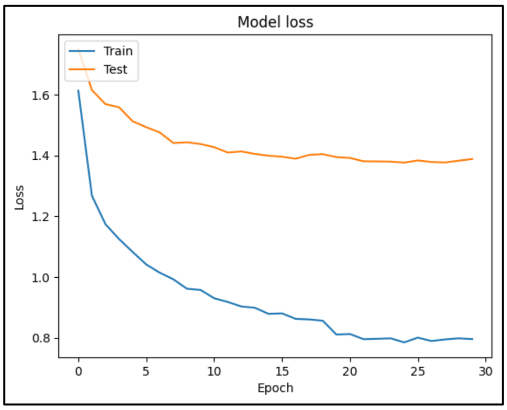

Figure 5.6: Loss of ENB7 - Model1

Table 5.3: *Performance Scores of ENB7 - Model1*

S NO.	PARAMETER	SCORES
1	Training Accuracy	74.33 %
2	Testing Accuracy	58.22 %
1	Training Loss	0.7959
2	Testing Loss	1.3884

ENB7 - MODEL2

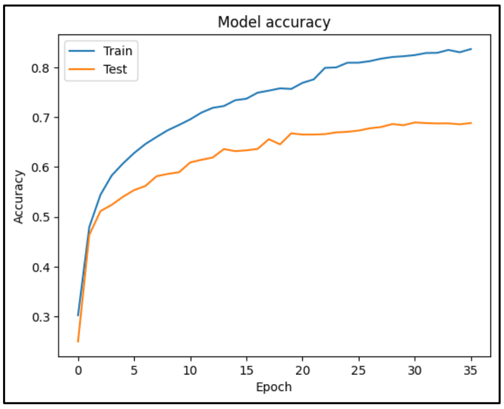

Figure 5.7: Accuracy of ENB7 - Model2

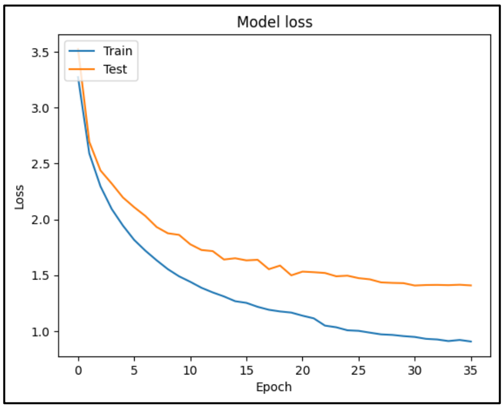

Figure 5.8: Loss of ENB7 - Model2

Table 5.4: *Performance Scores of ENB7 - Model2*

S NO.	PARAMETERS	SCORES
1	Training Accuracy	83.69 %
2	Testing Accuracy	68.95 %
1	Training Loss	0.9056
2	Testing Loss	1.4071

5.1.4 PERFORMANCE OF INCEPTIONRESNETV2

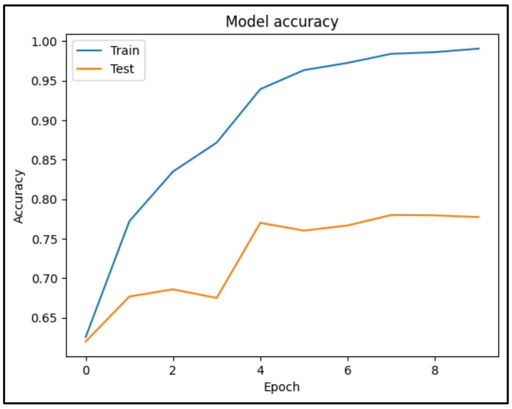

Figure 5.9: Accuracy of InceptionResNetV2

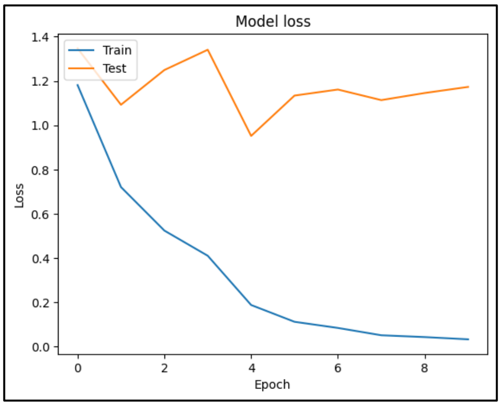

Figure 5.10: Loss of InceptionResNetV2

Table 5.5: *Performance Scores of InceptionResNetV2*

S NO.	**PARAMETER**	**SCORE**
1	Training Accuracy	99.04 %
2	Testing Accuracy	77.74 %
1	Training Loss	0.0326
2	Testing Loss	0.9767

5.1.5 PERFORMANCE OF DENSENET169

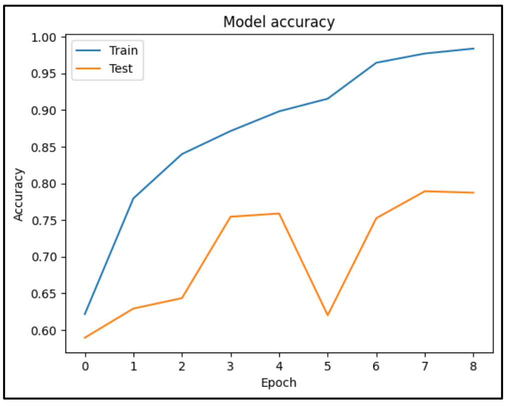

Figure 5.11: Accuracy of DenseNet169

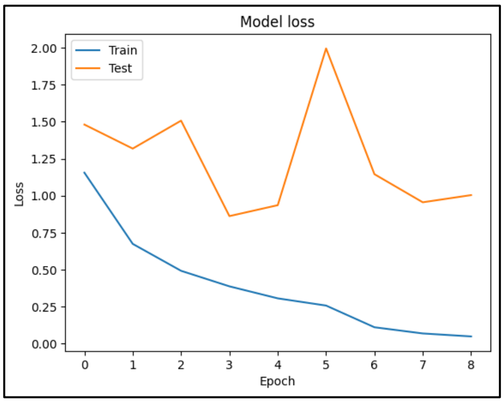

Figure 5.12: Loss of DenseNet169

Table 5.6: *Performance Scores of DenseNet169*

S NO.	**PARAMETER**	**SCORE**
1	Training Accuracy	98.31 %
2	Testing Accuracy	78.73 %
1	Training Loss	0.0502
2	Testing Loss	0.8535

5.1.6 PERFORMANCE OF DENSENET201

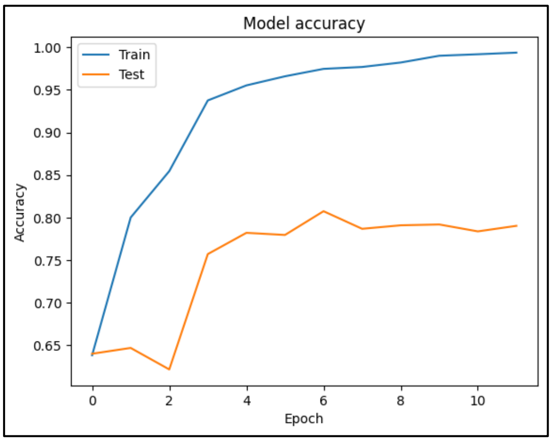

Figure 5.13: Accuracy of DenseNet201

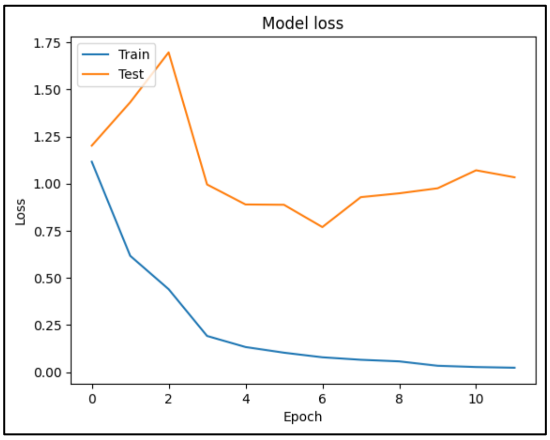

Figure 5.14: Loss of DenseNet201

Table 5.7: Performance Scores of DenseNet201

S NO.	PARAMETER	SCORE
1	Training Accuracy	99.36 %
2	Testing Accuracy	80.47 %
3	Training Loss	0.0229
4	Testing Loss	0.7684

Among the Transfer Learning models, the DenseNet family, particularly DenseNet201, stood out as the most effective for the research task. Despite initial challenges, including overfitting, the DenseNet models demonstrated remarkable adaptability and performance improvements following fine-tuning and regularization.

5.1.7 COMPARITIVE RESULTS

This section has an overview at the comparing the scores between the model and visualize their performances on the dataset_v3.

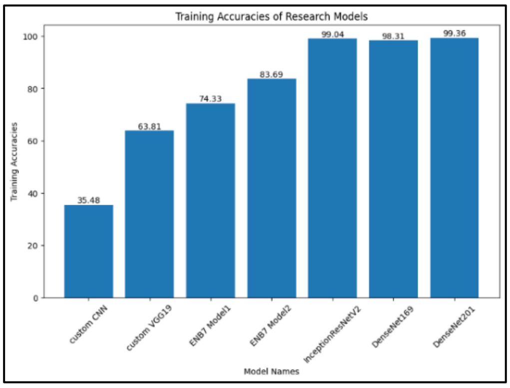

Figure 5.15:Training Accuracies of all Models

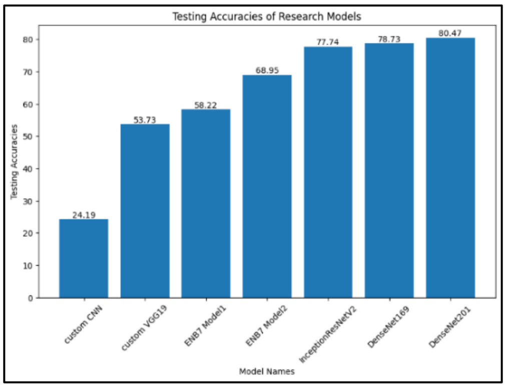

Figure 5.16: Testing Accuracies of all Models

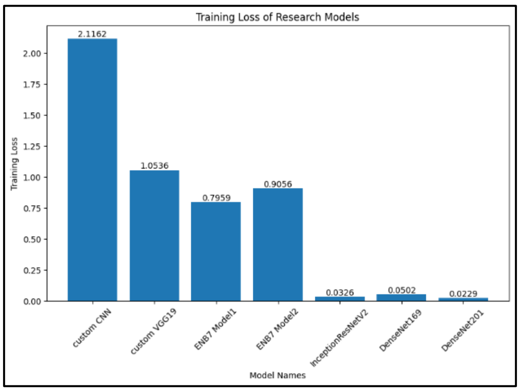

Figure 5.17:Training Loss for all Models

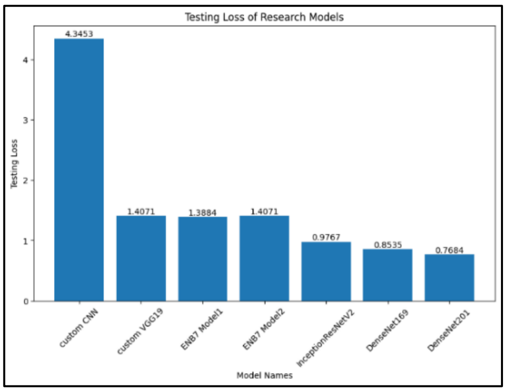

Figure 5.18: Testing Loss for all Models

Table 5.8: Table of Comparison between all Models

MODEL NAME	TRAINING ACCURACY (%)	TESTING ACCURACY (%)	TRAINING LOSS	TESTING LOSS
Custom CNN	35.48	24.19	2.1162	4.3453
VGG19	63.80	53.73	1.0536	1.4071
ENB7 Model1	74.33	58.22	0.7959	1.3884
ENB7 Model2	83.69	68.95	0.9056	1.4071
InceptionResNetV2	99.04	77.74	0.0326	0.9767
DenseNet169	98.31	78.73	0.0502	0.8535
DenseNet201	99.36	80.47	0.0229	0.7684

5.2 DISCUSSION

The results shows that approach taken considerably affected categorization performance. Each model was tested and compared for categorizing Indian supermarket products using accuracy and loss measures. The Transfer Learning approach to train DNN is a game changing technique which not only gives better performance for the classification tasks but also reduces the training time for the models significantly as majority of the weights of these models do not need to be retrained.

The custom-built CNN and VGG19 models, while the easiest to implement, performed poorly compared to the TL models. The low training and testing accuracies of these models reflect a problem generalizing to unknown data. Elevated testing loss values demonstrate these models' poor performance.

In contrast, EfficientNetB7 (ENB7) models outperformed CNN and VGG19 architectures. In both training and testing accuracy, ENB7 Model2 outperformed Model1. Compared to the other models, ENB7 Model2 had the lowest percentage point difference between Training and Testing accuracies and had lower testing loss values.

Among all models, InceptionResNetV2, DenseNet169, and DenseNet201 were most accurate. These models had minimal overfitting and high training and testing accuracies. Low testing loss values indicate that the models can generalize to new data, making them ideal for the research's classification goal. DenseNet201 performed best among various topologies, making it the best for classifying Indian supermarket items.

Chapter 6

Conclusion And Future Work

6.1 CONCLUSION

The results show that the chosen approach greatly affected Indian retail product classification. A game-changer, Transfer Learning (TL) techniques improved classification performance while reducing model training time. Although easy to execute, the traditional architectures did not perform well even after many attempts to improve their performances as compared. The more advanced Transfer Learning models had the maximum accuracy, minimum overfitting, and robust generalization, with DenseNet201 being the best model for categorizing Indian supermarket goods.

6.2 FUTURE WORKS

The findings of this study offer useful insights into the effectiveness of various deep learning architectures in categorizing Indian supermarket products. However, there are numerous opportunities for further investigation and enhancement in the future.

Further investigation into data augmentation techniques and strategies for balancing the dataset has the potential to enhance the performance of all models, especially the traditional architectures.

Due to time constraints and a lack of expertise, only limited number of models could be studied. Experimenting with other architectural designs may lead to finding an even better architecture optimal for the research problem. Additionally, a wide range of models may be found not only in the Keras application library but also in Google's TensorFlow Hub, Model Zoo, and similar online platforms. These libraries provide access to numerous pre-trained models trained on different datasets. Transfer learning can be utilized to efficiently apply these combinations of models and weights to the research problem.

Performing extensive hyperparameter tuning trials could further optimize the performance of the models. Optimizing parameters such as the learning rate, batch size, and dropout rates might potentially enhance accuracy and generalization.

Ensemble Methods: Exploring ensemble learning techniques, such as model averaging or stacking, may be advantageous in harnessing the capabilities of numerous models and enhancing overall classification performance.

Investigating alternative transfer learning techniques and pre-trained structures, not considered in this study, could reveal additional architectures that are very suitable for the classification job.

Designing domain-specific architectures that are customized to the unique characteristics of Indian supermarket product photos has the potential to enhance classification accuracy and robustness.

To further advance the findings of this study and improve the efficiency and accuracy of automated retail systems, future research should focus on these specific areas. This will enable the development of more effective and accurate models for identifying Indian supermarket products.

REFERENCES

[1] Wiryana, M., Harmanto, S., Fauzi, A., Qisthi, I., & Utami, Z. (2023). Store product classification using convolutional neural network. IAES International Journal of Artificial Intelligence (IJ-AI), 12, 1439. https://doi.org/10.11591/ijai.v12.i3.pp1439-1447

[2] Muhathir, M. F., Dwi Ryandra, R. B. Y., Syah, N., Khairina, N., & Muliono, R. (2023). Convolutional Neural Network (CNN) of Resnet-50 with Inceptionv3 Architecture in Classification on X-Ray Image. Convolutional Neural Network (CNN) of Resnet-50 with Inceptionv3 Architecture in Classification on X-Ray Image | SpringerLink. https://link.springer.com/chapter/10.1007/978-3-031-35314-7_20

[3] Anonymous. (2022). An effective CNN and Transformer complementary network for medical image segmentation. ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0031320322007075

[4] Anonymous. (2023). Comparison of CNN-based deep learning architectures for rice diseases classification. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2589721723000235

[5] Abdallah, S. E., Elmessery, W. M., Shams, M. Y., & Al-Sattary, N. S. A. (2022). Deep Learning Model Based on ResNet-50 for Beef Quality Classification. Arab Journals Platform. https://digitalcommons.aaru.edu.jo/isl/vol12/iss1/24

[6] Kumar, S., Pal, S., Singh, V. P., & Jaiswal, P. (2023). Performance evaluation of ResNet model for classification of tomato plant disease. De Gruyter. https://www.degruyter.com/document/doi/10.1515/em-2021-0044/html

[7] Anand, V., Gupta, S., Altameem, A., Nayak, S. R., Poonia, R. C., & Saudagar, A. K. J. (2022). An Enhanced Transfer Learning Based Classification for Diagnosis of Skin Cancer. MDPI. https://www.mdpi.com/2075-4418/12/7/1628

[8] Salehi, A. W., et al. (2023). A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. MDPI. https://www.mdpi.com/2071-1050/15/7/5930

[9] Srikantamurthy, M. M., Rallabandi, V. P. S., Dudekula, D. B., Natarajan, S., & Park, J. (2023). Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. SpringerLink. https://link.springer.com/article/10.1186/s12880-023-00964-0

[10] Anonymous. (2022). Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images. ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0957417422022485

[11] Patra, A., Vivek, V., Shambhavi, B. R., Sindhu, K., & Balaji, S. (2021). Product Classification in E-Commerce Sites. SpringerLink. https://link.springer.com/chapter/10.1007/978-981-33-4299-6_40

[12] Motwani, A., Bajaj, G., Arya, M., Sar, S. K., & Manoj, S. O. (2023). Machine Learning-Based Autonomous Framework for Product Classification Over Cloud. SpringerLink. https://link.springer.com/chapter/10.1007/978-981-19-9638-2_6

[13] Basheer Ahmed, M. I., et al. (2023). Deep Learning Approach to Recyclable Products Classification: Towards Sustainable Waste Management. MDPI. https://www.mdpi.com/2071-1050/15/14/11138

[14] Anonymous. (Year of Publication). Convolutional Neural Network (CNN) Based Identification of Crop Diseases Using ResNet-50. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10370122

[15] Kılıçarslan, G., Koç, C., Özyurt, F., & Gül, Y. (2023). Breast lesion classification using features fusion and selection of ensemble ResNet method. International Journal of Imaging Systems and Technology, 33(5), 1779-1795. https://doi.org/10.1002/ima.22894

[16] Sabeenian, R. S., Paul, E., & Prakash, C. (2022). Fabric defect detection and classification using modified VGG network. The Journal of The Textile Institute, 114(7), 1032-1040. https://doi.org/10.1080/00405000.2022.2105112

[17] Zheng, Y., Yang, C., & Merkulov, A. (2018). Breast cancer screening using convolutional neural network and follow-up digital mammography. IEEE Conference Publication. https://doi.org/10.1117/12.2304564

[18] Khalil, M., Tehsin, S., Humayun, M., Jhanjhi, N., & AlZain, M. (2022). Multi-Scale Network for Thoracic Organs Segmentation. Computers, Materials and Continua, 70, 3251-3265. https://doi.org/10.32604/cmc.2022.020561

[19] Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning. https://www.jmlr.org/papers/volume21/20-074/20-074.pdf

[19] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. Proceedings of the AAAI conference on artificial intelligence, 31(1). https://doi.org/10.1609/aaai.v31i1.2438