2026-04

Machine Vision–Based Classification of Rare Fruits in Bangladesh Using Transfer Learning and Custom CNN Models

Hasan, Md. Hasib

IUB

Machine Vision–Based Classification of Rare Fruits in Bangladesh Using Transfer Learning and Custom CNN Models

April 2026

Prepared by:

Md. Hasib Hasan
ID: 2220166

Afsana Islam
ID: 2222077

Rabeya Bosri
ID: 2230150

Department of Computer Science and Engineering
Independent University, Bangladesh

Supervised by

Dr. Md. Tarek Habib
Associate Professor
Department of Computer Science and Engineering
Independent University, Bangladesh

Attestation

We are aware of the fact that plagiarism is strictly prohibited and it conflicts with our university's rules and regulations. We guarantee the authenticity of our work. We have used some copyrighted material and models of others in our project which has been properly cited following the international standards proper guidelines.

Author Name: Md. Hasib Hasan

. . . .

Signature:

. . . .

Author Name: Afsana Islam

. . . .

Signature:

. . . .

Author Name: Rabeya Bosri

. . . .

Signature:

. . . .

Letter of Transmittal

April 2026

To
Md. Tarek Habib, PhD
Associate Professor, Department of Computer Science and Engineering
Independent University, Bangladesh

Subject: Submission of Undergraduate Thesis

Dear Sir,
We present our undergraduate thesis entitled "Machine Vision–Based Classification of Rare Fruits in Bangladesh Using Transfer Learning and Custom CNN Models" as part of the requirement towards the Bachelor of Science in Computer Science and Engineering degree.

Our study introduces a deep learning–based framework for classifying six rare and indigenous fruits of Bangladesh using both transfer learning models and a proposed Custom CNN architecture. The model achieves a highest classification accuracy of 97.40%, demonstrating strong performance and efficiency in real-world conditions.

It is our utmost pleasure to express our gratitude for your continuous mentoring, supervision, and support during the course of our study. Your invaluable guidance has been essential to the successful completion of this research.

Thank you very much for your attention.

Sincerely,

Md. Hasib Hasan
Afsana Islam
Rabeya Bosri

Evaluation Committee

Supervisor

Name: ______________________ Signature: ______________________

Internal Examiner 1

Name: ______________________ Signature: ______________________

Internal Examiner 2

Name: ______________________ Signature: ______________________

External Examiner

Name: ______________________ Signature: ______________________

Acknowledgement

We would like to express sincere gratitude to our respected supervisor Dr. Md. Tarek Habib, Associate Professor, Independent University, Bangladesh. His dedicated guidance, insightful suggestions, remarkable patience, and the generous investment of his valuable time played a pivotal role in the successful completion of our project. Our supervisor's expertise and mentor ship greatly enhanced the quality and depth of our work, and we are truly appreciative of his unwavering support throughout our academic journey.

Additionally, our heartfelt gratitude goes to Fab Lab IUB for their extraordinary help and consistent support. The collaborative and resourceful environment provided by Fab Lab IUB significantly contributed to the development and execution of our project. Moreover, the encouragement and assistance have been instrumental in shaping our academic and research field. We are grateful for the opportunities and resources extended to us during our research, which undoubtedly enriched our learning experience.

We acknowledge and appreciate the contributions of all individuals who supported us, both directly and indirectly, during this endeavor.

Abstract

Bangladeshi rare fruits are indigenous and culturally significant agricultural resources that remain under-documented and are gradually becoming less common due to limited commercial cultivation and declining public awareness. Automated recognition of these fruits can play a crucial role in supporting biodiversity conservation, consumer awareness, and sustainable agriculture. This study proposes a deep learning–based image classification framework for recognizing six rare fruit species commonly found in Bangladesh, namely Bilimbi, Custard Apple, Roselle, Star Fruit, Tropical Almond, and Water Chestnut. A custom convolutional neural network is designed and evaluated alongside four widely used transfer learning models: MobileNetV2, InceptionV3, ResNet-50, and DenseNet-121. A dataset comprising 1,800 real-world fruit images was constructed, with 300 images per class captured under varying illumination, background, and orientation conditions. To ensure reliable generalization assessment, the dataset was stratified into 70% training, 15% validation, and 15% testing subsets. Experimental results demonstrate that the proposed Custom CNN achieves the highest classification accuracy of 97.40%, outperforming all baseline models. In addition, the Custom CNN attains superior ROC–AUC (0.9990) and PR–AUC (0.9955) scores, indicating strong discriminative capability and robustness under class imbalance. Comparative analysis further reveals that the proposed lightweight CNN consistently outperforms deeper pre-trained architectures while maintaining lower computational complexity. These findings confirm the effectiveness of domain-specific model design for fine-grained rare fruit recognition. The proposed system has strong potential for real-world applications in digital agriculture, educational tools, and biodiversity preservation, contributing to the sustainable conservation of Bangladesh's rare fruit heritage.

Contents

Letter of Transmittal **2**

1 Introduction **11**
1.1 Background . . . 11
1.1.1 Definition of Rare Fruits . . . 12
1.1.2 Characteristics of Rare Fruits . . . 12
1.1.3 Biodiversity and Rare Fruits . . . 12
1.1.4 Rare Fruits in Bangladesh . . . 12
1.2 Importance of Rare Fruits . . . 13
1.3 Nutritional Value of Rare Fruits . . . 13
1.4 Geographical Distribution of Rare Fruits Across Climatic Regions . . . 14
1.5 Local Fruit Diversity and Choice of Rare Fruits in Bangladesh . . . 15
1.6 Health Benefits of Rare Fruits . . . 16
1.7 Economic and Cultural Importance . . . 16
1.8 Challenges and Conservation of Rare Fruits . . . 17
1.9 Nutritional and Socioeconomic Importance . . . 17
1.10 Global Perspective of Rare Fruits . . . 19
1.11 Climate Change and Rare Fruits . . . 19
1.11.1 Scope of the Study . . . 19
1.11.2 Limitations of the Study . . . 20
1.11.3 Underutilization of Rare Fruits . . . 20
1.11.4 Lack of Awareness Among Consumers . . . 20
1.11.5 Loss of Biodiversity and Habitat . . . 21
1.11.6 Limited Scientific Research and Documentation . . . 21
1.11.7 Need for Comprehensive Research . . . 21
1.11.8 Limited Technological Integration . . . 21
1.11.9 Need for Integrated Research . . . 22
1.12 Problem Statement . . . 22
1.12.1 Research Contributions . . . 22
1.13 Purpose of the Study . . . 22

2 Systematic Literature Review **23**
2.1 Objective of the Review . . . 24
2.2 Review Questions . . . 24
2.3 Review Methodology . . . 26
2.3.1 Data Sources and Digital Libraries . . . 26
2.3.2 Search Strategy . . . 27
2.3.3 Inclusion Criteria . . . 28
2.3.4 Exclusion Criteria . . . 29
2.3.5 Quality Assessment of Selected Studies . . . 29

2.3.6 Data Extraction . . . 31
2.4 Outcomes of the Literature Search . . . 32
2.4.1 Distribution by Publication Year . . . 32
2.4.2 Deep Learning Architectures Used . . . 32
2.4.3 Distribution by Publication Type . . . 33
2.4.4 Common Research Topics . . . 33
2.4.5 Dataset Size Distribution . . . 33
2.4.6 Geographic Distribution of Studies . . . 34
2.5 Findings from the Review . . . 35
2.5.1 Application Areas of Fruit Classification . . . 35
2.5.2 Datasets Used in Prior Studies . . . 36
2.5.3 Evaluation Metrics and Reported Performance . . . 37
2.5.4 Limitations in Existing Literature . . . 37
2.6 Comparative Analysis of Selected Studies . . . 37
2.7 Research Gaps Identified from the SLR . . . 38

3 Theoretical Foundations **40**
3.1 Image Processing Fundamentals . . . 40
3.1.1 Digital Image Representation . . . 41
3.2 Convolutional Neural Networks (CNNs) . . . 42
3.2.1 Basic Structure of CNNs . . . 43
3.2.2 Working Mechanism of CNNs . . . 43
3.3 Transfer Learning Theory . . . 44
3.3.1 Mathematical Formulation of Transfer Learning . . . 45
3.4 Optimization Algorithms . . . 46
3.5 Evaluation Metrics Formulation . . . 48

4 Dataset Preparation and Preprocessing **51**
4.1 Overview of Dataset . . . 51
4.2 Data Collection Process . . . 51
4.2.1 Source of Data and Field Collection . . . 52
4.2.2 Tools and Image Acquisition Setup . . . 52
4.2.3 Image Variations and Dataset Diversity . . . 53
4.2.4 Data Collection Time . . . 53
4.2.5 Challenges and Dataset Balancing . . . 53
4.3 Data Cleaning and Annotation . . . 54
4.3.1 Data Cleaning and Preprocessing . . . 54
4.3.2 Label Verification and Quality Assurance . . . 55
4.4 Image Preprocessing . . . 55
4.4.1 Image Resizing . . . 56
4.4.2 Color Space Standardization . . . 56
4.4.3 Pixel Normalization . . . 56
4.4.4 Noise Reduction and Image Quality Enhancement . . . 56
4.5 Data Augmentation . . . 57
4.5.1 Integration with Preprocessing Pipeline . . . 57
4.5.2 Impact on Model Generalization . . . 57

5 Proposed Methodology **58**
5.1 Research Framework 58
5.2 Transfer Learning Models 60
5.2.1 MobileNetV2 60
5.2.2 InceptionV3 62
5.2.3 ResNet-50 64
5.2.4 DenseNet-121 65
5.3 Proposed Custom CNN Architecture 67
5.4 Model Training Strategy 68
5.5 Hyperparameter Settings 69

6 Experimental Setup **72**
6.1 Hardware Software Environment 72
6.2 Training Protocol 73
6.3 Complexity Analysis 74
6.3.1 Time Complexity 74
6.3.2 Space Complexity 75
6.3.3 Model Efficiency and Implications 75
6.4 Training and Analysis of Validation Performance of Custom CNN 76

7 Results and Performance Analysis **78**
7.1 Model Comparison 78
7.2 Confusion Matrix Analysis 79
7.2.1 Confusion Matrix Evaluation Metrics 80
7.2.2 Confusion Matrix Analysis Modelwise 80
7.2.3 Comparative Analysis 81
7.2.4 Observations and Discussion 81
7.3 Precision–Recall (PR-AUC) vs ROC-AUC Comparison 82
7.4 Error Analysis 84
7.5 Computational Efficiency 86

8 Discussion **88**
8.1 Why Custom CNN Outperformed 88
8.2 Transfer Learning vs Custom Model 89
8.3 Practical Implications for Bangladesh 90
8.4 Contribution to Agricultural AI 90
8.5 Model Deployment Possibilities 91

9 Conclusion and Future Works **92**
9.1 Summary of Findings 92
9.2 Research Contributions 92
9.3 Limitations 93
9.4 Future Research Directions 94

Bibliography **94**

List of Figures

1 Geographical Distribution of Rare Fruits 15
2 Pictures of Rare Fruits : (a)Bilimbi , (b)Roselle, (c) Custard Apple, (d) Star Fruit, (e) Tropical Almond, (f) Water chestnut 16
3 Health Benefits Of Rare Fruit 17
4 Economic and Cultural Importance 18
5 Challenges and Conservation of Rare Fruits 18

6 Key Research Questions on Fruit Classification 26
7 PRISMA flowchart for the literature search process 30
8 Distribution of Reviewed Studies by Publication Year 32
9 Distribution of deep learning architectures used in fruit classification research 33
10 Distribution of studies by publication type 34
11 Geographic distribution of fruit classification studies 35

12 Image Processing Pipeline 40
13 Working Mechanism of CNNs 44

14 Structure of the rare fruit dataset showing class distribution and data split 52
15 Workflow of the data collection process for Fruit Image Classification Dataset 53
16 Overview of key challenges encountered during data collection 54
17 Manual annotation and quality assurance pipeline 55

18 Proposed Research Framework 58
19 Architecture of MobileNetv2 [41] 61
20 Architecture of InceptionV3 [47] 63
21 Architecture of ResNet-50[52] 65
22 Architecture of DenseNet-121 [49] 66
23 Architecture of Custom CNN 68

24 Training and Validation Accuracy Curve 76
25 Training and Validation Loss Curve 77

26 Confusion Matrix of Proposed Custom CNN 79
27 Precision–Recall (PR-AUC) Curve 83
28 Receiver Operating Characteristic (ROC) Curve 84
29 Correct and Incorrect Extinct Rare Fruit Image Classification 85

30 Comparative Analysis of Custom CNN outperforms transfer learning . . . 88

31 Key Research Contributions 93

List of Tables

1.1 Importance of Rare Fruits 13
1.2 Nutritional Components and Health Benefits of Rare Fruits 14

2.1 Review Questions and Their Rationale 25
2.2 Digital Libraries Used for Literature Search 27
2.3 Keyword Groups Used in the Search Strategy 28
2.4 Quality Assessment Criteria 30
2.5 Data Extraction Fields Used in the Review 31
2.6 Dataset Size Distribution in Selected Studies 34
2.7 Comparative Summary of Fruit Classification Studies 38

4.1 Summary of the Data Collection Process 54
4.2 Summary of Data Augmentation Parameters 57

5.1 Custom CNN Architecture Layer Configuration 67
5.2 Hyperparameter Settings of the Proposed Model 69

6.1 Hardware Specifications to be used to train the model. 72

7.1 Evaluation Metrics of Extinct Rare Fruit Image Classification 79
7.2 Models Comparative Performance Analysis of Confusion Matrix. 81
7.3 Computational Comparison of Models 87

8.1 Comparison Between Transfer Learning Models and Custom CNN 89

1 Introduction

1.1 Background

Fruits are central to our diets, farming systems and the environment. They are an excellent source of the nutrients that we need to maintain good health: vitamins, minerals, antioxidants, fiber and sugars, which in turn help us combat chronic diseases. And in terms of their role in the global food economy and food system, fruits play a critical role. Fruits are important in developing nations that have agriculture as the main source of livelihoods. That we see in the large scale production of common fruits such as mango, banana, apple, orange, pineapple, papaya and grape which are very successful in the market due to their popularity, storage potential, and supply chains. But there is a countless number of rare or underutilized fruits. These are less well-known, cultivated and traded fruits. These are fruits that are generally found in a specific ecological niche (forests, rural areas) that they have evolved in. While they are very nutritious and beneficial for the environment we don't see them as often due to lack of knowledge, commercialisation and research. In the past many of these rare fruits formed an integral part of diets but as modern methods of farming have developed with an emphasis on producing a high crop yield and marketable product we have seen the use of these fruits diminish. Another key factor which makes rare fruits so unique is their regionalism. They may only grow well in very specific climates which may be tropical or sub tropical and which has the right temperature, precipitation and soil conditions for the fruit to grow. Unlike common fruits which are well stored and shipped, the rare fruits don't store well, don't have regular shapes and are harder to transport to market which then makes bulk production and marketing a less viable option. Also they are not as well studied as the major crops and common fruits so we don't know as much about their health and farming potential. The life blood of rare fruits is biodiversity. It is the diversity of life in an environment that is crucial for ecosystem health and stability. We see this in tropical forests where unique plant species (such as rare fruits) are to be found. These are the ideal habitats for vegetation. However, we are facing environmental problems such as deforestation, urbanisation and climate change which are impacting these areas, so we are seeing reduction in the biodiversity and as a consequence we may face the loss of many rare fruit species. In Bangladesh where we have a tropical climate, fertile land and receives much rainfall we have a wide range of native fruits. These fruits include bilimbi, roselle, star fruit, tropical almond, water chestnut and custard apple and are found in home gardens and village gardens. Although highly useful nutritionally and culturally they are not widely grown commercial crops and are becoming obscure. Younger groups. Awareness and consumption is a great threat to their extinction.

1.1.1 Definition of Rare Fruits

Rare fruits are plant species that yield eatable fruit but are less known, used, or available than commercially grown fruit crops. These fruits generally are not integrated into conventional agricultural production or international trade networks due to demand and awareness, absence of standardised production and processing methods, and constraints in storage, transportation or marketability. Rare fruits are typically confined to specific environments such as forests, swamplands, home gardens, and are often cultivated in small-scale traditional agricultural practices. They are often seasonal and their production is highly sensitive to factors like soil quality, rainfall and temperature. Although rare fruits are not widely commercially available, they are nutritionally important, containing high levels of vitamins, minerals, antioxidants and other bioactive compounds that promote health and prevent disease. They also contribute to biodiversity conservation and ecosystem sustainability. But because of limited research effort, lack of integration into markets, and dominance of globally traded fruit species, these species remain overlooked and underutilised in terms of both scientific and economic potential.

1.1.2 Characteristics of Rare Fruits

Rare fruit varieties differ in that they put out and reach markets in different ways from what we see with common commercial fruits. Also it is the case that farmers do not put as much of their effort into the former because they are focused on which crops have the greatest demand from the consumer, which the best shelf life, and which have the best transport infrastructure. Also it is true that rare fruits tend to have odd shapes, short shelf lives, and don't do as great a job in the market. Also of note is that they are very much products of their geography. Many will only grow in very specific climate regions which mostly are tropical and sub tropical which in turn plays a role in their status as under utilized species. Also rare fruits do not get the same level of scientific study as do main crop staples like rice, wheat and also popular commercial fruits. This lack of research in turn does not give us the info we need on their health benefits and nutritional value.

1.1.3 Biodiversity and Rare Fruits

Biodiversity is the variety of living things within an ecosystem which also plays a key role in maintaining ecological balance. Rare fruits are a key element of biodiversity as they present unique genetic resources which in turn add to plant diversity. In tropical areas which have high biodiversity we see that which also support the growth of rare fruits we are talking about ample rainfall, rich soil and right temperatures. These conditions in turn support a large variety of plant species which also are yet to be studied. Also biodiversity is put under threat by deforestation, urbanization and climate change. These do in fact cause loss of habitat and in that also reduce the available natural environment for rare fruit growth thus at the same time we see an increased risk of their extinction

1.1.4 Rare Fruits in Bangladesh

Bangladesh's a tropical setting which has fertile soil, ample rainfall, and very good for agriculture which in turn presents a great diversity of native fruits. Many of these fruits are found to grow in rural homesteads and village gardens as opposed to large scale farms. We see in Bilimbi, Roselle, Star Fruit, Tropical Almond, Water Chestnut, and Custard Apple which are also known for their health benefits and cultural value. But also we are seeing that many of these fruits are becoming hard to find as farmers are more into growing what is more commercial like mango, banana and jackfruit. Also, Bangladesh is a bio-diverse country which has that tropical climate, rich soil, and great rain fall which in turn supports a large variety of fruit species that includes many indigenous and rare kinds. Also in the rural areas these are traditional grown and used in local food and medicine. Although they play a large role in our culture and health still many of these fruits are becoming rare due to which our youth are not as aware of them and we are seeing a shift to more commercial fruits.

1.2 Importance of Rare Fruits

Rare fruits play a role in many aspects which include biodiversity conservation, agricultural sustainability, nutrition, and socio economic development. We see that they mainly in their role of preserving genetic diversity which is of great importance. Genetic diversity is what we need for crop improvement which in turn gives us a wide range of traits like disease resistance, drought tolerance, and also which make plants do well in diverse environments. In terms of agriculture rare fruits bring in diversity which in turn reduces our dependence on a few crops. This in term increases resilience against pests, diseases and climate change. From a nutrition point of view rare fruits are very rich in vitamins, minerals and anti oxidants at times out doing common fruits. Also from an economic stand point rare fruits support rural livelihoods by which they provide extra income through local markets and small scale processing industries. Also they play a role in culture which includes traditional food, medicine and community heritage. Although they bring in a lot of benefits their decline is brought about by issues like habitat loss, climate change and also lack of awareness. [1]

Table 1.1: Importance of Rare Fruits

Aspect	**Description**	**Significance**
Biodiversity Conservation	Rare fruit species contribute to preserving genetic diversity and unique plant traits.	Helps maintain ecological balance and supports environmental sustainability.
Agricultural Sustainability	Promotes diversification of crops, reducing reliance on a limited number of large-scale species.	Enhances resilience against pests, diseases, and climate change.
Nutritional Value	Rich in vitamins, minerals, antioxidants, and dietary fiber.	Improves human health and helps prevent nutrient deficiencies.
Economic and Cultural Value	Supports local farmers and strengthens connections to traditional practices and heritage.	Boosts livelihoods while preserving cultural identity and indigenous knowledge.

1.3 Nutritional Value of Rare Fruits

Rare which are few in number are full of essential nutrients that better our health and well being. They have a great variety of vitamins in them like vitamin C, vitamin A and B complex which in turn support the immune system, vision and energy metabolism.

Also they supply key minerals like calcium, potassium, magnesium and iron which are very important for strong bones, heart function and oxygen delivery into the blood. Also present in these rare fruits is dietary fiber which in turn aids in digestion and also decreases the risk of chronic diseases like diabetes and heart issues. Also they are a great source of anti oxidants which include polyphenols and flavonoids which in turn protect the body from oxidative stress and cell damage[2]. Also they have natural sugars which at the same time give a healthy energy source as opposed to that which is in processed foods. Also it is the combination of these nutrients which brings out the high value of rare fruits in terms of nutrition which in turn makes them a very important part of a balanced diet. We should see to it that we eat them more which in the long run will improve public health greatly in particular in areas which do not have access to a large choice of food [2] [3].

Table 1.2: Nutritional Components and Health Benefits of Rare Fruits

Rare Fruit Name	**Scientific Name**	**Vitamins & Nutrients**	**Health Benefits**	**Additional Information**
Bilimbi	*Averrhoa bilimbi*	Vitamin C, antioxidants, potassium	Improves immune system, aids digestion, reduces inflammation, protects cells	Very sour in taste; commonly used in cooking rather than eaten raw
Roselle	*Hibiscus sabdariffa*	Vitamin C, Vitamin A, antioxidants, fiber	Improves heart health, aids digestion, helps reduce blood pressure	Widely used in herbal tea, traditional medicine, and home remedies
Star Fruit	*Averrhoa carambola*	Vitamin C, Vitamin A, potassium, fiber	Boosts immunity, supports eye health, helps regulate blood pressure	Distinctive star-shaped slices; popular in pickles
Tropical Almond	*Terminalia catappa*	Vitamin E, healthy fats, magnesium, calcium	Supports heart health, strengthens bones, improves brain function	Edible seeds similar to nuts; relatively rare and location-specific
Water Chestnut	*Trapa natans*	Vitamin B6, potassium, fiber, antioxidants	Aids digestion, supports metabolism, helps maintain blood pressure	Grows in freshwater environments like ponds and lakes
Custard Apple	*Annona squamosa*	Vitamin C, Vitamin B6, magnesium, natural sugars	Provides energy, supports nervous system, boosts immunity	Soft, sweet, creamy pulp; commonly found in tropical regions

1.4 Geographical Distribution of Rare Fruits Across Climatic Regions

Rare fruits are found all over the world which mostly grows as a result of climate, soil, rainfall, and altitude. In the tropical regions which include countries like India, Brazil and Indonesia we see the greatest diversity because of warm temperatures and rich biodiversity which in turn sees ecosystems like the Amazon Rainforest as key in that. Also in the subtropical regions of China, India, the southern U.S., Australia, and Europe we see a great variety of rare fruits which do well in moderate climates. But in the temperate regions although there are few species what we do see are unique fruits like wild berries and traditional orchard varieties. Also these rare fruits have adapted to very specific eco niches which has seen them develop traits like drought resistance and adaptation to poor soils. But at the same time their distribution is put at risk by climate change, deforestation and urbanization. Also that they are not widely cultivated and not very much in the commercial eye means many of these species are contained in very small local areas. We must study this issue out if we are to do conservation, promote sustainability in use, and protect what is at out best value.

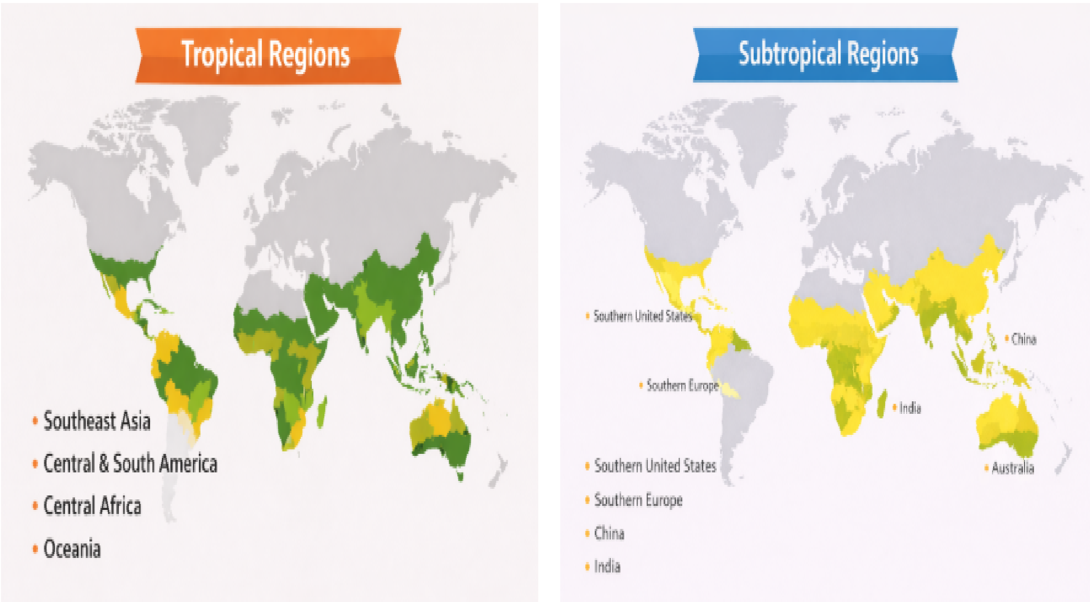

Figure 1: Geographical Distribution of Rare Fruits

1.5 Local Fruit Diversity and Choice of Rare Fruits in Bangladesh

Bangladesh has a great diversity of local fruits which grow naturally in our rural areas, homesteads and forests although we do not fully utilize them and they are slowly becoming a thing of the past for younger generations. Of these the Bilimbi is a sour tropical fruit which is very rich in vitamin C and is used in cooking, also we have the Roselle which though not very well known is valued for its red calyces used in drinks and for traditional medicine which also has high levels of antioxidants and health benefits. Also we have the Star fruit which is a hit for its unique shape and balance of sweet and sour taste which also is a good source of vitamin C and fiber, also the Tropical almond which produces very nutritious edible seeds which are great for heart health. We also have the Water chestnut which grows in fresh water and is a good source of fiber and minerals, also the Custard apple which is a favorite for its sweet pulp and energy boosting nutrients. Although these fruits are very healthy and play a large role in our culture they are not grown on a large scale and are at risk of being forgotten which is why there is a need for more awareness and conservation [5] .

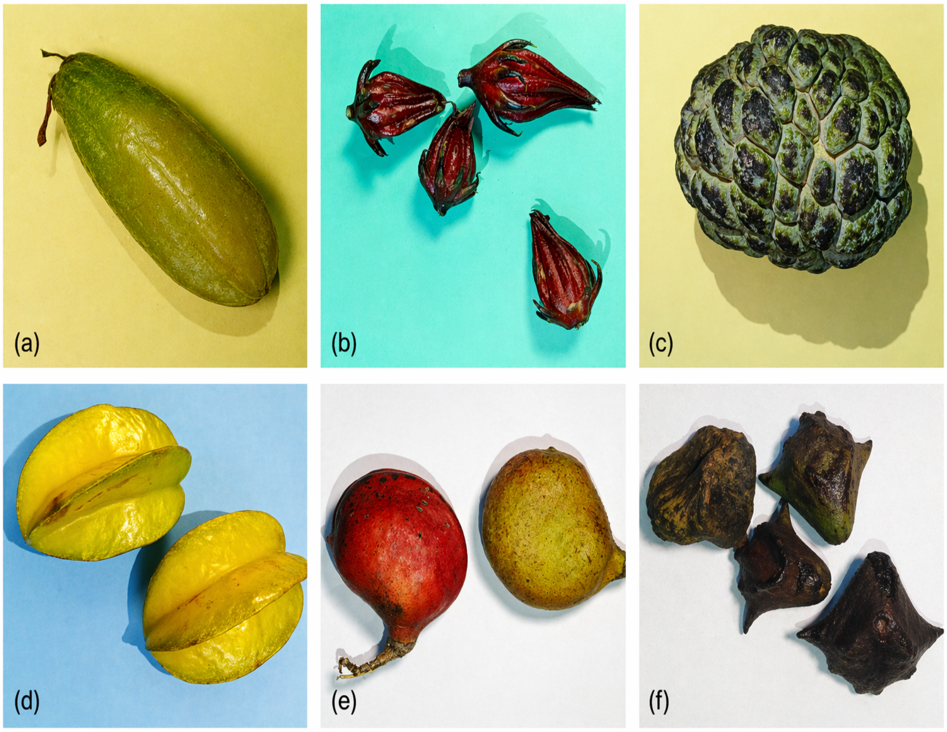

Figure 2: Pictures of Rare Fruits : (a)Bilimbi , (b)Roselle, (c) Custard Apple, (d) Star Fruit, (e) Tropical Almond, (f) Water chestnut

1.6 Health Benefits of Rare Fruits

The intake of rare fruits is very beneficial for health. They have a high content of antioxidants that in turn reduces free radical damage and which also put chronic disease like cancer, diabetes and cardiovascular issues at lower risk. Also these fruits' high in dietary fiber improve digestive health and support healthy gut bacteria. Rare fruits also play a role in improving the immune system because of their high vitamin C content. We see minerals like potassium and magnesium present in these fruits which play a role in health by improving blood pressure and circulation. Also some rare fruits have anti-inflammatory and anti-microbial properties which also play a role in traditional medicine.

1.7 Economic and Cultural Importance

Fruits and their rare variety are of economic importance which is more so in rural communities where they serve as a livelihood. What we find in these communities is that the farmers are the growers of these fruits which they then bring to the local markets or they add value to these fruits by turning them into jams, cordials and pickles. In terms of culture rare varieties of fruits are certainly part of our culture which includes their

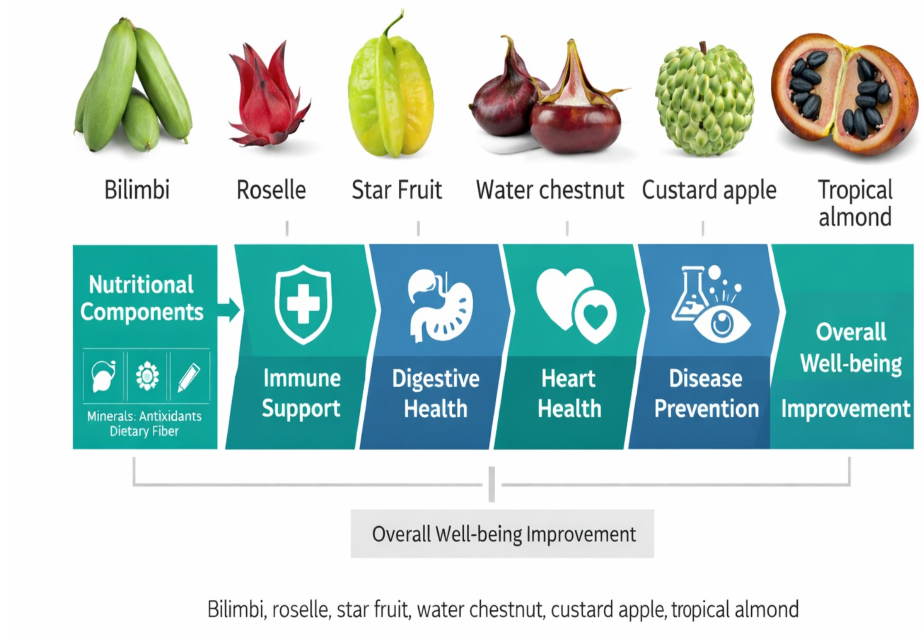

Figure 3: Health Benefits Of Rare Fruit

use in food, festivals and medicine. They are part of the local knowledge systems which are passed on from generation to generation. Therefore we should conserve them for our economic future and for the conservation of our culture.

1.8 Challenges and Conservation of Rare Fruits

Despite their value, rare fruits put forth a fight which they often lose. Deforestation and urbanization which cause habitat loss are the main players in their decline. Also to this goes Climate change which0 transforms environmental conditions that are key to their growth. Also we see that there is a lack of consumer awareness and little scientific research which plays a role in the under utilization of these species. Thus conservation strategies which must be put in place include habitat protection, sustainable agricultural practices and more research. Also in to this we have Technology which is a great help via machine learning and computer vision in the identification and documentation of rare fruit species.

1.9 Nutritional and Socioeconomic Importance

In Bangladesh's tropical climate which also presents ample rain and fertile soil the growth of over 60 fruit species is seen which include many indigenous types that play a large role in nutrition and traditional diets. Fruits such as bilimbi, roselle, star fruit, tropical almond, water chestnut, and custard apple are rich in essential nutrients like vitamins, minerals, fiber, and antioxidants which in turn promote health. For example bilimbi and roselle have high levels of vitamin C and antioxidant content, star fruit supports digestive and cardiovascular health [14]. Tropical almond and water chestnut supply healthy fats, fiber, and minerals and the custard apple provides energy also it's a boost for the immune [12][18]. Also these native species contain bioactive compounds with

Figure 4: Economic and Cultural Importance

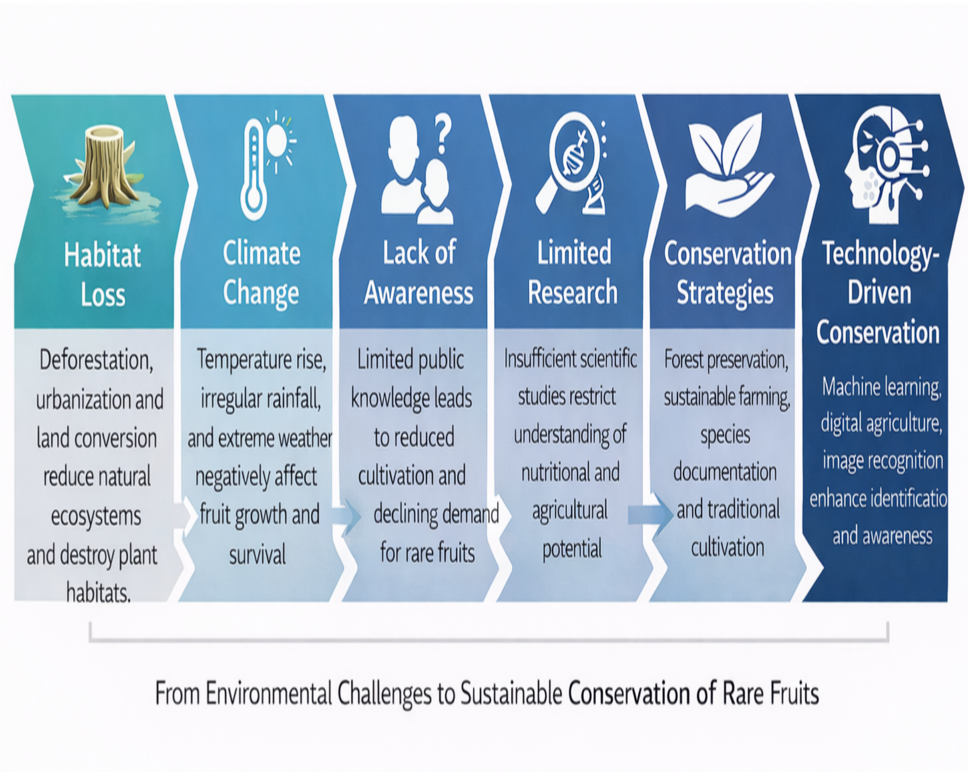

Figure 5: Challenges and Conservation of Rare Fruits

medicinal properties which include anti diabetic and anti inflammatory benefits. Though they have health benefits they are still very much underused due to low awareness and commercial cultivation. By promoting their use we can improve dietary variety, address issues of micronutrient deficiency and support food security in particular in rural areas. Also they play a role in conservation of biodiversity and provide economic input for local communities thus they are very much so a part of sustainable agriculture and public health in Bangladesh [6].

1.10 Global Perspective of Rare Fruits

Rare fruits do not grow only in Bangladesh, we see them in many different parts of the world especially in tropical and subtropical areas which have high biodiversity. In Southeast Asia fruits like durian and mangosteen are very much a part of what is known there but in the global market they are still seen as rare or exotic because of their limited availability and high perishability[23]. Also in South America we find very nutrient rich rare fruits like açaí and camu camu which are noted for their very high levels of antioxidants and vitamin C [9] [10]. In Africa the baobab fruit is a very important native resource which has high fiber, calcium and vitamin content and it plays a large role in local nutrition and traditional medicine. Also in the Mediterranean we have unique fruit varieties of different types of figs which although they have been cultivated for centuries are still very much under utilized in the modern commercial systems. From these examples we see that rare fruits are a global phenomenon which are tied into regional climate, culture and ecosystem. Study of rare fruits on a global scale not only increases the academic value of the research but also brings to light the issue of preserving biodiversity and promoting less known food resources world wide.

1.11 Climate Change and Rare Fruits

Climate change presents as one of the greatest issues for global agriculture which includes the growth and distribution of fruit species. We see that environmental variables like temperature, rain fall, humidity, and soil quality play key roles in fruit production and survival. What we are seeing is that these variables' changes which are a result of global climate change do in fact disrupt flowering patterns, reduce yield and in some cases put certain fruit species at risk of extinction. We note that rare fruits which have adapted to very specific local environments are very much at risk from such changes. But also some indigenous fruit species do show greater resilience to environmental stress because of their natural adaptation to local conditions which includes resistance to drought, poor soil, or irregular rain. This in turn makes them a great resource for the development of climate resistant agricultural systems. Also we see that increased temperatures, unpredictable rain fall patterns and extreme weather events put serious threats to their survival which is also true when coupled with deforestation and habitat loss. Thus it is very important that we study the relationship between climate change and rare fruits in order to put in place sustainable agricultural practices and at the same time preserve biodiversity for the benefit of future generations.

1.11.1 Scope of the Study

This study focuses on six rare indigenous fruits of Bangladesh—bilimbi, roselle, star fruit, tropical almond, water chestnut, and custard apple—analyzing them from both technological and agricultural perspectives. The primary scope is the development of a deep learning–based image classification model using data collected under real-world conditions. The work is limited to classification only and does not include detection or segmentation tasks.

Alongside the technical aspect, the study highlights the nutritional, economic, and ecological importance of these underutilized fruits. It also addresses the impact of commercial farming practices, which are gradually replacing traditional fruit cultivation and increasing dependence on imported fruits. Overall, the study aims to promote awareness of rare fruits while supporting sustainable agriculture and modern technological applications.

1.11.2 Limitations of the Study

Despite our in depth approach this study has issues. We mainly used secondary data which while large in scale do have the issue of not having recent or regional updates. Also we did not do primary data collection which would include field studies or lab analysis which in turn does not allow us to back up some of our results. Also we looked at a chosen set of rare fruits and not at the full range of underutilized fruit species. We had time and resource constraints which did not allow us to include all rare fruits found world wide or in Bangladesh. Also we note a lack of scientific research done on rare fruits. While we have major crop research in great detail we do not have that for many indigenous fruits which in turn leaves gaps in info related to their nutrition, medicine, and agriculture. Also we saw that environmental issues like climate change and habitat loss are very complex and dynamic which made it hard to predict their long term play on rare fruit species. These inexact predictions in turn affected the broad application of our study's results. That said the study does bring to light the value of rare fruits and puts forth the case for more research, conservation and education.

1.11.3 Underutilization of Rare Fruits

Bangladesh like many other tropical countries is a home to great diversity of indigenous fruit species which we attribute to very favorable climate and very fertile soil. Although we have this great natural endowment, what we see is a very narrow focus in terms of agricultural practices and research which is directed mainly at what are the commercial favorites like mango, banana and jackfruit. This narrow focus has caused the neglect of a large array of rare and underutilized fruit species. Many of these fruits are not seen on large scale cultivation nor are they available in the markets. Thus they are not well documented and are in the process of disappearing from day to day diet. Also we see that there isn't enough integration of these rare fruits into main stream agriculture which has in turn reduced their role in the food supply chain. This underutilization does not only1 limit diet diversity but also leads to the loss of what may prove to be very important genetic resources for future agricultural development.

1.11.4 Lack of Awareness Among Consumers

One of the key challenges in Bangladesh for rare fruits is consumer awareness. Several rare and indigenous fruits, though nutritionally and ecologically valuable, are under consumed due to lack of awareness about their nutritional value and availability. People tend to gravitate towards more popular fruits, diminishing the market for rare fruits. This impacts both consumer behaviour and farming. Additionally, a lack of identification methods and online resources makes it challenging to identify these fruits. This research helps to overcome this problem by suggesting an automatic classification system that can assist in raising awareness and education.

1.11.5 Loss of Biodiversity and Habitat

Another major issue is that we are seeing a great loss of biodiversity which is a result of human actions like urbanization, deforestation, and agricultural expansion. We are seeing natural habitats which have been home to rare fruit species for years destroyed or changed which in turn is a cause for these plants' struggle for survival. Also it is reported that we are putting forward high yield crops which in turn is leading to the replacement of the traditional fruit varieties with what are simply more profit oriented options. This shift not only is bringing down the amount of crop diversity we have but also is to blame for a weak ecosystem. The loss of biodiversity also has long term effects which include reduced ability to adapt to climate change and increased susceptibility to pest and disease.

1.11.6 Limited Scientific Research and Documentation

Despite the fact that rare fruits are a valuable resource we should be paying more attention to, in the academic and agricultural community we see that main focus is on the top producing crops which in turn causes many indigenous fruits to be ignored. This gap in research leads to a lack of information on their health benefits, uses in medicine, and role in the ecosystem. Also we see that without this research which is at present very thin which in turn makes it hard to put these fruits forward to the public and to include them in agricultural policy and development. Also we do not have the data we need for these fruits to really take off in a commercial sense and be adopted by a wider audience. What we do need is a large scale research effort that looks at the issues of these fruits' distribution, what they do for health, and what their economic value is.

1.11.7 Need for Comprehensive Research

Considering the issues raised we see that there is a great need for in depth study of rare fruits. Such research should report on what these fruits are like, what they contain in terms of nutrition, and what role they play in terms of biodiversity conservation and sustainable agriculture. By looking at these issues we put forth that research which in turn will increase public awareness, promote the growth of rare fruits' culture and support in the development of policy. In the end this will work towards the preservation of what we have of precious plant species, improving food security and at the same time encouraging more sustainable agricultural practices.

1.11.8 Limited Technological Integration

Recent we've seen growth in technology which includes in machine learning and computer vision and is very present in agricultural research. We see these tech tools for the identification, classification and study of biodiversity. At the same time though which we are at the early stages of seeing that same tech applied to the study of rare fruits. Some research has indeed looked at the use of image classification models for fruit species identification, what we are missing is wide scale implementation and integration of these tools into research and practice. That is a large gap which we must address in order to better document and bring to light rare fruits.

1.11.9 Need for Integrated Research

In this context it is clear that there is a great need for research which looks at many aspects of rare fruits at the same time which includes their distribution, health benefits, role in the ecosystem, and also how they fit into economic structures. This study is put forth to fill in these gaps by taking a more in depth and wide ranging look at rare fruits. We also are to present a multi disciplinary approach which will include review of current literature as well as other methods in an attempt to put forth a better picture of rare fruits and their role in sustainability.

1.12 Problem Statement

Bangladesh is home to a wide array of rare and indigenous fruits, but these are underused and many are endangered because of a lack of awareness and identification techniques. Visual identification can be complex, time-consuming and requires special expertise, particularly when fruits look similar. Machine learning algorithms are generally trained using common fruit images and cannot effectively identify rare local fruit varieties. So, there is a need for a fast and automated system to classify rare fruits using images. This can help in agricultural development, accessibility and biodiversity conservation.

1.12.1 Research Contributions

This research has put forth in many dimensions. To start with the study which is a in depth look at rare fruits and their global distribution. By looking at multiple scientific resources the research brings to light the diversity and value of what we term as underutilized fruit species. Also we see that the research puts forward the health and nutritional benefits of rare fruits. By way of analysis of what is present in the chemistry of these fruits and what they do for health we see the health promoting elements of rare fruits which in turn play a role in better nutrition and public health. Also we see that the research brings to the fore the issue of conservation of rare fruit species as a component of biodiversity conservation efforts. What we learn of the ecological value of these fruits supports sustainable agricultural practices and environmental protection. Also the study does it's part in to increase public awareness of rare fruits and their role in food systems. We put forth information which we hope will encourage the culture, consumption and conservation of these fruits.

1.13 Purpose of the Study

Rare fruits are becoming increasingly scarce in today's world, highlighting the urgent need for enhanced research and proper documentation. This study aims to explore the global distribution of rare fruits across different climatic zones and analyze their nutritional composition, including essential vitamins, minerals, antioxidants, and dietary fiber, along with their health benefits.

Furthermore, the study examines the economic, agricultural, and cultural significance of rare fruits within diverse communities. It also identifies the major threats these fruits face, including environmental changes and human activities. By addressing these aspects, this research emphasizes the importance of conserving rare fruit species, which contributes to biodiversity conservation, strengthens food security, and helps preserve valuable natural and cultural resources for future generations.

2 Systematic Literature Review

The use of artificial intelligence and computer vision has completely changed the landscape in most areas of agriculture, creating automated platforms which can be used to process visual data to accomplish tasks like crop monitoring, plant disease detection, yield estimation, and fruit classification. One such application has been in the field of agricultural image classification, where deep learning has been applied to identify plant species, fruits, and crops in digital photos. Convolutional Neural Networks (CNNs) and transfer learning-based models have proven to be very effective in visual recognition tasks and can, therefore, be used to develop automated classification systems that can enhance modern agricultural practices.

The recognition of fruits is of special concern in agricultural machine vision, as the correct identification of the fruit species can be used to monitor crops, automate agriculture, record biodiversity, and provide agricultural services digitally. Fruit recognition system: Automated systems can help farmers, researchers and agricultural organizations to classify the types of fruit varieties using automated classification systems rather than depending on knowledge only of experts. Most of the current fruit classification research, however, is mostly centered around commercially popular fruits or more commonly available data sets, including apples, bananas, oranges and any other species of fruit grown worldwide. As a result, indigenous or regional fruits are frequently overlooked in machine vision studies despite their ecological and cultural importance. The thesis involves rare and indigenous fruits in Bangladesh, of which most are slowly disappearing because of decreasing production, lack of commercial interest, and little knowledge among people about these species. Some of these fruits constitute a significant portion of the biodiversity of the country and the culture of traditional food that is underrepresented in current machine vision data and literature. An automated image classification system of these fruits could be developed to raise awareness and aid in biodiversity documentation and the acceptance of these species into the agricultural and technological sphere. Thus, to learn the current state of research, a systematic review of previous studies on the topic of fruit image classification, plant recognition based on deep learning, the application of transfer learning methods to agriculture, and the application of specialized CNN models to specific datasets is needed.

In this regard, the chapter will perform a Systematic Literature Review (SLR) to examine the existing studies in the context of the topic of fruit image classification and deep learning-based agricultural recognition systems. This review aims to determine the most popular methodologies, datasets, deep learning structures, and evaluation plans and to outline research gaps that are currently found in the context of rare or underrepresented fruit species. The outcomes of this review form a basis of the proposed study by determining the existing trends, constraints, and opportunities that lead to the creation of the machine vision-based framework of rare fruits classification proposed in this thesis.

2.1 Objective of the Review

This Systematic Literature Review (SLR) aims to review the current studies on fruit image classification and plant recognition based on the machine vision and deep learning methods. Techniques based on computer vision have been increasingly used in agricultural applications over recent years, especially with the introduction of Convolutional Neural Networks (CNNs) and transfer learning methods. Automated identification of fruits and plants using image data and deep learning networks has been studied in multitude of studies. The nature of the studies, methodologies, and datasets, however, differ significantly, and it is thus required to analyze previous research in a systematic manner to be able to comprehend the current research trends and limitations.

The aim of this review is to determine the most frequently applied methods and computation approaches in literature to perform fruit classification tasks. Specifically, the review will analyze the categories of deep learning architecture that have been used by other researchers, such as popular transfer learning models, such as MobileNetV2, InceptionV3, ResNet, and DenseNet, and other designs of CNN created to be used in specific classification problems. Also, the review explores the prevalence with which transfer learning has been utilized in the field of agricultural image classification and its comparison with wholly tailored deep learning models.

The second important goal of this review is to examine the datasets of the previous fruit classification research. This involves the investigation of the fruit species used in the datasets, the size of the dataset, data collection procedures and the use of publicly available datasets as opposed to custom datasets. Knowledge of dataset properties is also significant in this study as most current datasets only concentrate on fruits commonly produced but little has been done on rare or native fruit species.

Additionally, this review will appraise the evaluation metrics and performance indicators used in contemporary studies, and limitations made by researchers. Through these points, the review aims to ascertain the issues that are common in fruit classification studies, including lack of diversity of datasets, imbalanced classes, and generalization problems across varying settings.

Lastly, a major goal of this SLR is to explore whether current literature is focused on the classification of rare, indigenous, or region-specific fruits, especially in the setting of biodiversity-rich countries, like Bangladesh. As most indigenous fruits are still underrepresented in the existing machine vision studies, the discovery of this gap is a compelling incentive to conduct the research. The results of this review are thus useful in determining the research direction of this thesis since it reveals the trends in methodology, limitation of data sets and research gaps that support the rationale of developing a machine vision-based classification system of rare fruits in Bangladesh.

2.2 Review Questions

A set of research questions was created to direct the review process to systematically analyze the existing literature that relates to the fruit image classification and machine vision applications in agriculture. The following review questions will be addressing the following research directions and computational methods, datasets, evaluation practices, and limitations that are reported in previous research. The answers to these questions will help the review offer a systematic insight into the recent progress in fruit classification with deep learning and detect research gaps that drive the current study on rare and indigenous fruits in Bangladesh.

Table 2.1: Review Questions and Their Rationale

ID	Review Question (RQ)	Rationale
RQ1	What are some of the machine vision and image-based methods of fruit classification?	Knowing the variety of machine vision methods applied in fruit recognition helps identify prevailing methods and technological trends in agricultural image classification.
RQ2	What are the most common deep learning architectures used to classify fruit images?	Identifying popular models such as CNN variants and transfer learning approaches helps determine efficient computational techniques for fruit recognition tasks.
RQ3	To what extent is transfer learning used in studies of fruit classification?	Transfer learning is useful when datasets are limited. Analyzing its usage helps understand how researchers handle data scarcity and computational constraints.
RQ4	What types of datasets are used in previous fruit classification research?	Exploring dataset characteristics such as number of classes, size, and image sources provides insight into how data availability influences model performance.
RQ5	Which preprocessing and data augmentation methods are commonly applied in fruit image classification?	Image preprocessing and augmentation techniques improve model generalization and help address dataset limitations, especially in agricultural domains.
RQ6	Which evaluation metrics are most commonly used in assessing fruit classification models?	Evaluating metrics such as accuracy, precision, recall, and F1-score helps understand how model performance is measured across different studies.
RQ7	What are the reported levels of classification performance in past literature?	Comparing reported performance levels helps assess the effectiveness of different models and identify challenges in achieving reliable fruit classification.
RQ8	What limitations and challenges are commonly identified in fruit classification research?	Identifying issues such as limited datasets, class imbalance, environmental variability, and poor generalization highlights unresolved challenges in the field.
RQ9	To what extent have indigenous or region-specific fruits been considered in previous research?	Most studies focus on commercially common fruits, while indigenous fruits remain underrepresented. This helps identify gaps in biodiversity-focused research.
RQ10	What research gaps motivate the development of a machine vision-based classification system for rare fruits in Bangladesh?	Identifying research gaps provides the foundation for the proposed study and justifies the need for specialized datasets and models targeting rare fruits.

These research questions help to organize the systematic review and to organize the analysis discussed in the subsequent sections. The answers to these questions assist in determining the methodological trends, data constraints and research gaps that exist in the current literature, and thus, this information is used to design and generate the proposed framework of classifying rare fruits. These research questions are visualized in

Figure 6 that gives a brief overview of the major areas of focus in this review.

Figure 6: Key Research Questions on Fruit Classification

2.3 Review Methodology

A systematic review procedure was followed so that the process of literature selection was transparent, reproducible and the studies that are pertinent to machine vision applications in fruit classification were considered. The review methodology encompasses a number of steps, such as identification of digital databases, creation of search strategies, application of inclusion and exclusion criteria, process of study selection and extraction of the relevant data of the selected studies. Such measures were chosen to guarantee that the review will have the most pertinent and the best quality of research contributions in the field of fruit classification by using the deep learning and transfer learning methods.

The approach of this review is based on the common practices of systematic reviews in computing science and artificial intelligence research. It started with the identification of the appropriate digital libraries with peer-reviewed research articles. A search strategy based on a set of keywords associated with fruit classification, plant recognition, deep learning, and computer vision was then created. Screening of the retrieved studies was then carried out based on the predefined inclusion and exclusion criteria to eliminate irrelevant or duplicate studies. Following this filtering, the rest of the papers were further processed to obtain the essential information regarding datasets, model architectures, preprocessing methods, metrics of evaluation, and reported performance.

2.3.1 Data Sources and Digital Libraries

The literature search relied on the use of the main databases such as Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink, and ACM Digital Library. These are well known in indexing quality journal articles and conference papers in engineering, computer science and applied artificial intelligence research. Besides these primary sources, Google Scholar was employed as an auxiliary search engine to find other relevant studies that can have been missed in the main databases.

The search was limited mainly to peer-reviewed journal articles and conference papers because these sources generally include comprehensive descriptions of the methods, results of the experiments, and the validated research results. Other forms of publications, like editorials, non-peer-reviewed reports, and non-technical articles, were not taken into consideration to preserve the quality and reliability of the reviewed literature as an academic source.

Table 2.2: Digital Libraries Used for Literature Search

Database	**Research Coverage**	**Reason for Selection**
Scopus	Multidisciplinary scientific database covering computer science, engineering, and AI	Indexes high-quality peer-reviewed journals and conferences extensively.
Web of Science	Multidisciplinary citation database	Ensures inclusion of highly cited and influential research.
IEEE Xplore	Electrical engineering, computer science, and artificial intelligence	Contains major conference and journal articles in computer vision and deep learning.
ScienceDirect	Engineering, data science, and applied AI research	Hosts numerous machine learning and agricultural AI research publications.
SpringerLink	Computer vision, machine learning, and data science	Provides a large collection of conference proceedings and journals on deep learning.
ACM Digital Library	Computing and machine learning research	Important source for computer vision and AI research publications.
Google Scholar	Broad academic search engine	Used as a secondary source to identify additional relevant papers.

2.3.2 Search Strategy

To seek the relevant studies on the topic of fruit classification by the methods of machine vision and deep learning, a structured search strategy was created to select the relevant studies in a systematic manner. The search strategy was constructed to encompass research contributions by several related areas such as computer vision, deep learning, agricultural informatics, and plant recognition systems. To do so, a list of keywords was created relying on the main themes of this study, i.e., fruit classification, image recognition, deep learning models, and farm usage. The keywords were chosen by using a preliminary exploratory search of the literature about the topic and by examining the terms that are shared by the past published works that focus on the topic of fruit classification and plant recognition. The keywords were then grouped into a few concept groups that are reflective of the key elements of the topic being studied. Every group includes terms that are closely related, describing similar concepts, which were joined with the help of Boolean operators to create the final search query. The main keyword categories of this review are the terms associated with the type of fruits, classification tasks, deep learning approaches, and agricultural settings. Keywords were related within each group of concepts with the OR operator, and the various groups of concepts were combined with the AND operator to make sure that only studies that had all main thematic aspects to the research were retrieved.

Table 2.3: Keyword Groups Used in the Search Strategy

Concept Group	**Keywords Used**
Fruit Domain	"fruit", "rare fruit", "indigenous fruit", "fruit image", "crop image", "plant fruit"
Classification Task	"classification", "recognition", "identification", "image classification"
Machine Learning Methods	"deep learning", "convolutional neural network", "CNN", "transfer learning", "computer vision"
Application Domain	"agriculture", "plant science", "agricultural image analysis", "biodiversity"

2.3.3 Inclusion Criteria

Originally, 221 papers were found by searching databases. Since duplicate records were eliminated, the number of articles was reduced to 111. We filtered 20 papers out of the review papers, short papers, data papers, and abstracts. The screening again was done manually by reading the titles and abstracts of these articles to determine their relevance with the inclusion criteria and narrowed the list to 15 articles. Inclusion criteria were: Studies were included when they:

- Targets a problem in fruit classification, fruit recognition, plant image classification, or a similar agricultural image classification problem.
- Image-based classification using applied machine learning, deep learning, convolutional neural networks (CNNs) or transfer learning methods.
- Experiment results and metrics of evaluation allowing performance to be analyzed.
- Made full-text available to examine methodology in detail.
- Were published in peer-reviewed journals or conference proceedings.
- Solved image-based classification tasks in both computer vision and agricultural applications.
- Published since 2014 to keep up with recent developments in deep learning.

2.3.4 Exclusion Criteria

Along with inclusion criteria, various exclusion criteria were used to eliminate the studies which were not directly related to the aim of this study. These criteria were meant to make sure that the chosen studies narrow down to image-based fruit classification, and it has to use machine learning or deep learning models. During the screening process, the studies which failed to satisfy the methodological or topical criteria of this review were eliminated. The following exclusion criteria were used:

- Such articles as reviews, editorials, theses, posters and book chapters were left out because this review only discusses original research articles and conference papers that report experimental outcomes.
- The research that was not based on the fruit classification or the plant recognition on the basis of the image was rejected.
- Articles that were solely about plant disease detection, leaf disease detection or pest detection were dropped unless fruit classification was a main part of the research.
- During the screening, duplicate studies were eliminated as found in numerous digital libraries.
- Only non-English publications were omitted so that there was consistency in interpretation and analysis.
- Articles that lacked adequate methodological description or experimental testing were blocked out because they lack adequate information that would aid in determining their methods or findings.
- Articles with more emphasis on object detection, segmentation, hyperspectral analysis, or other non-classification problems were not considered as the main emphasis of this thesis is on fruit image classification but not detection or segmentation.
- The studies that were published earlier than 2015 were omitted because the review should concentrate on the latest development of deep learning and computer vision methods.

2.3.5 Quality Assessment of Selected Studies

In order to make sure that the studies used in this systematic literature review were reliable and had scientific validity, a quality assessment process was carried out. This assessment was done to determine the methodological clarity, experimental rigor, and reporting quality of the research articles that were chosen. The evaluation of the quality of studies will help to guarantee that the synthesized findings of this review have credible and well-documented research. Each of the chosen studies was rated on a number of predetermined quality criteria of machine learning and computer vision studies. Such criteria cover significant parts of experimental design and reporting, such as the transparency of datasets, description of the model, the possibility to repeat experiments, the methodology of their evaluation, and the discussion of the research limitations. Research with well-presented methodological information and extensive experimental support is regarded as more stable in its content to grasp the current trends of research in the field of fruit classification and agricultural machine vision.

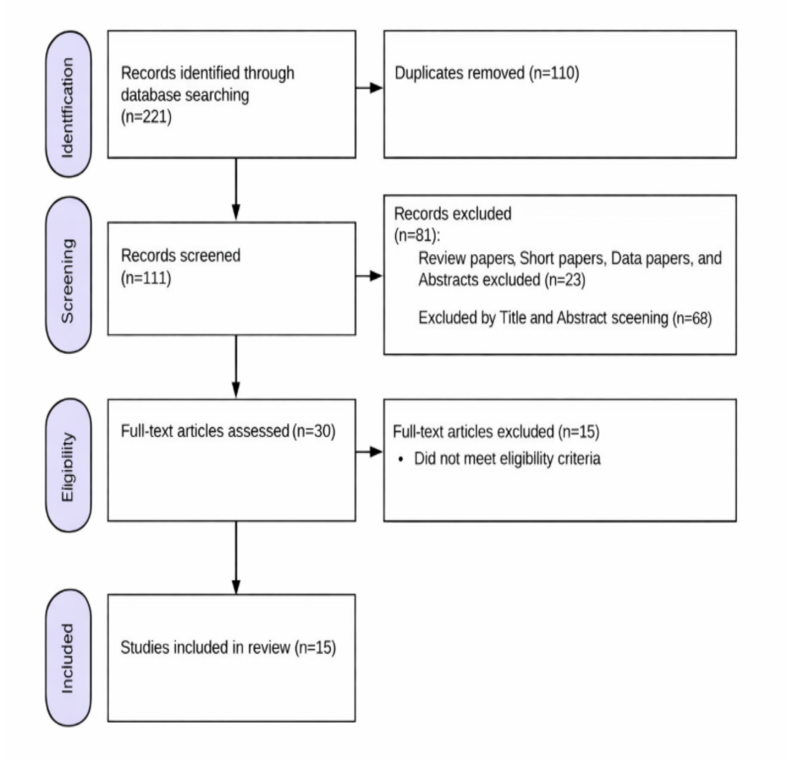

Figure 7: PRISMA flowchart for the literature search process

Table 2.4: Quality Assessment Criteria

Criteria ID	Quality Assessment Criteria	Purpose of Evaluation
QA1	Clear description of the data set that is used in the research	Maintains a good documentation of dataset characteristics, size and class distribution.
QA2	Elaboration of the proposed model/ algorithm	Checks that model architecture and training process are well outlined.
QA3	Reproducible experimental setup	Assures adequate information is given to reproduce the experiment.
QA4	Use of appropriate evaluation metrics	Ensures that the measurement of performance is done by standard measures that include accuracy, precision, recall or F1-score.
QA5	Comparison with baseline or existing methods	Establishes whether or not the study compares its methodology with previous methods.
QA6	Limitations or problems discussed	Assesses whether authors acknowledge potential weaknesses or constraints of their work.

2.3.6 Data Extraction

Following the last selection of articles after the screening and quality assessment steps, a systematic process of data extraction was carried out to extract the appropriate data in each of the chosen articles. This process aimed to systematize the essential information in the literature in a coherent format to enable trends, methodological trends, and research gaps in fruit image categorization to be examined. A data extraction template was developed using standardized items to capture the key characteristics of the research methodology, data features, model structures, and evaluation practices of the studies examined. Structured extraction of this information will assist in creating consistency throughout the analysis and make comparisons of various research works meaningful. The information that has been extracted pays special attention to the factors that are relevant to this thesis, including the kind of fruit datasets, deep learning models, preprocessing methods, evaluation measures, and performance that have been reported. These aspects shed light on the application of machine vision methods in the issue of classifying agricultural images and present shortcomings that drive the creation of a system to classify rare fruit in Bangladesh. This systematic extraction enables the literature to be examined in a variety

Table 2.5: Data Extraction Fields Used in the Review

Field	**Description**
Author and Year	Name of the authors and publication year.
Country / Research Domain	Location or research domain of the study.
Fruit Classes	Number and type of fruit categories used.
Dataset Size	Total number of images included in the dataset.
Preprocessing Method	Image preprocessing techniques applied.
Data Augmentation	Techniques such as rotation, flipping, scaling, and cropping.
Model Used	Deep learning architecture used (e.g., CNN, MobileNet, ResNet).
Transfer Learning	Indicates whether pre-trained models were utilized.
Evaluation Metrics	Performance measures such as accuracy, precision, and recall.
Best Accuracy	Highest classification accuracy reported in the study.
Limitations	Challenges or limitations identified by the authors.
Relevance to Thesis	Explanation of how the study relates to rare fruit classification.

of perspectives, such as characteristics of datasets, applications of models and methods of assessment. In order to further analyze the extracted data, the reviewed studies were sorted into some thematic categories. These themes can be used to determine the research patterns and the methodological approaches common in the literature. The main themes that are taken into account in this review are:

- Dataset Characteristics: Dataset size, fruit class count and image data sources are analyzed.

- Deep Learning Architectures: Recognition of popular CNN models, including MobileNet, ResNet, Inception and custom CNN architecture.
- Learning Strategy: Analysis of application of the transfer learning to the completely tailored deep learning models.
- Image Preprocessing and Augmentation:Comparison of preprocessing strategies and augmentation strategies to enhance model generalization.
- Evaluation Metrics and Performance: Investigation of the performance metrics used and reported classification accuracy.
- Limitations and Challenges: Recognition of typical research limitations like imbalance of dataset, small sample of images, or varying environments.

2.4 Outcomes of the Literature Search

2.4.1 Distribution by Publication Year

The subsection throws light on the time trend of the research in this field. It aids in determining changes in the interest in fruit classification over time. Figure 8 demonstrates the publication rates of the studies per year, which have been steadily rising in the recent years with a significant rise in the last years because of the deep learning progress.

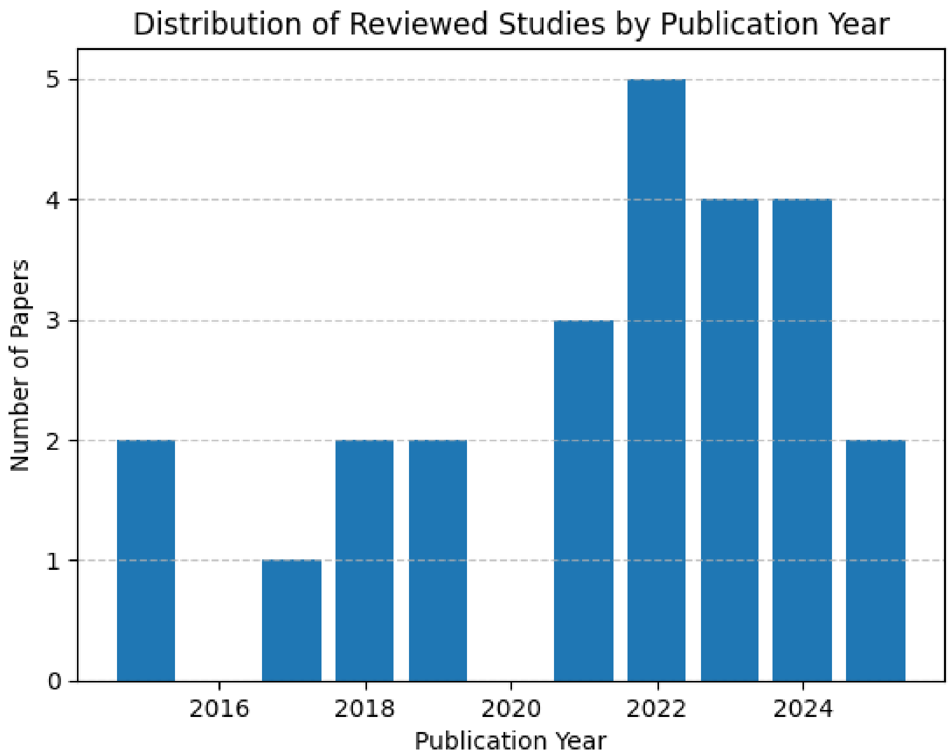

Figure 8: Distribution of Reviewed Studies by Publication Year

2.4.2 Deep Learning Architectures Used

This subsection examines the types of models commonly applied in fruit classification tasks. It is a mirror of the preferences of researchers in technologies. Figure 9 shows the distribution of the architectures including MobileNet, ResNet, Inception, DenseNet, and custom CNNs, with CNN-based models being the most prevalent.

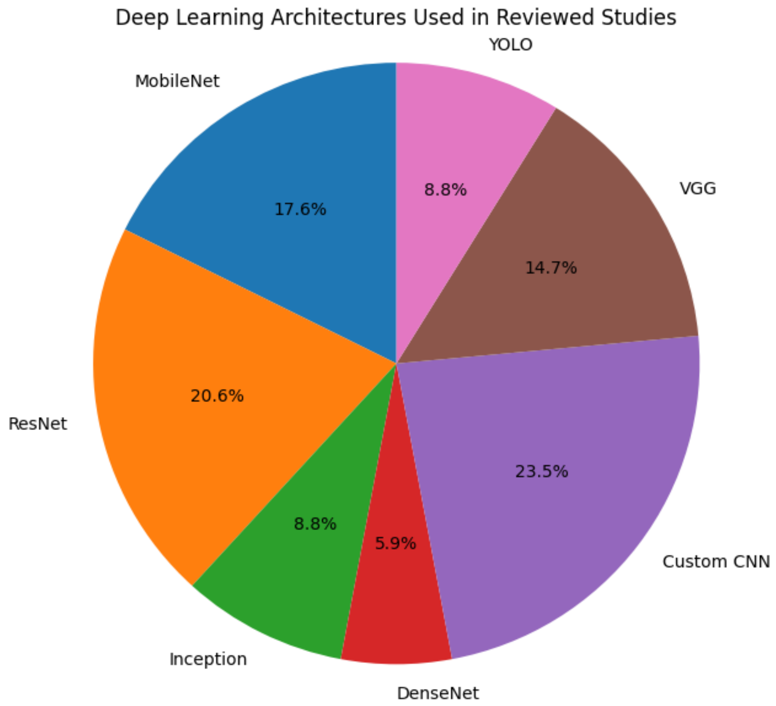

Figure 9: Distribution of deep learning architectures used in fruit classification research

2.4.3 Distribution by Publication Type

This subsection classifies the chosen studies according to the source of publication. It gives an idea on the dissemination of research findings. Figure 10 shows the balance between journal and conference publications with some contributions of both detailed studies and recent innovations.

2.4.4 Common Research Topics

The studies reviewed were divided according to their main research interest so as to get the general themes in the literature. The most recurrent research themes obtained are: Fruit Image Classification using deep learning models Image Dataset based Plant Species Recognition. Computer vision Analysis of Agricultural Image. Transfer Learning in classification of crops and fruits. Among such topics, the classification of fruits with the help of convolutional neural networks and transfer learning models is the most studied one.

2.4.5 Dataset Size Distribution

The size of the datasets used in the experiments of the reviewed studies was also analyzed. The size of the dataset contributes to a large extent to the performance of the deep learning models, with larger datasets generally allowing more learning of features and generalization of models. Table 2.6 groups the chosen studies based on the data size.

The descriptive findings in this section give a representation of the literature environment in terms of image classification of fruits and the deep learning-based agricultural

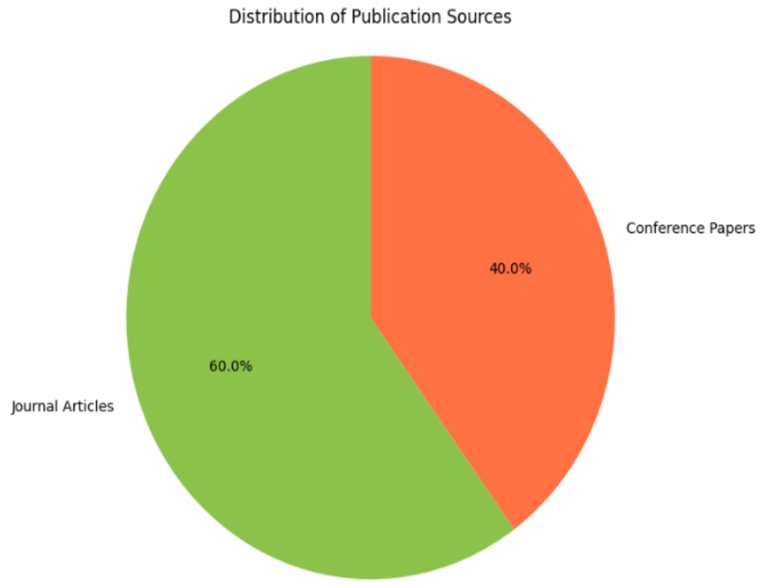

Figure 10: Distribution of studies by publication type

Table 2.6: Dataset Size Distribution in Selected Studies

Dataset Category	**Number of Studies**	**Description**
Small Dataset	1	Fewer than 1,000 images
Medium Dataset	3	Between 1,000 and 10,000 images
Large Dataset	11	More than 5,000 images

vision system. The analysed papers prove the growing research interest in the field as a number of them are devoted to CNN-based models and transfer learning methods. The literature, however, also reveals that the majority of studies concentrate on fruits typically grown and publicly available datasets, and little is done on rare or indigenous fruit species. These studies are further analyzed in the following sections where they consider the methodologies, datasets and experimental approaches employed in research.

2.4.6 Geographic Distribution of Studies

The countries that have active fruit classification research communities, such as China, India, the United States, and several European countries, are the source of many fruit classification studies. In these studies, the species of fruits are commonly locally significant or publicly available data on fruits. Nevertheless, there is a rather limited amount of researches dedicated to the investigation of rare or native fruits in the areas of biodiversity like Bangladesh. This fact shows the low presence of region-specific fruit datasets in existing machine vision studies and the significance of classification systems of underrepresented fruit specie. Figure 11 shows the distribution of the chosen studies according to the country of the first author geographically.

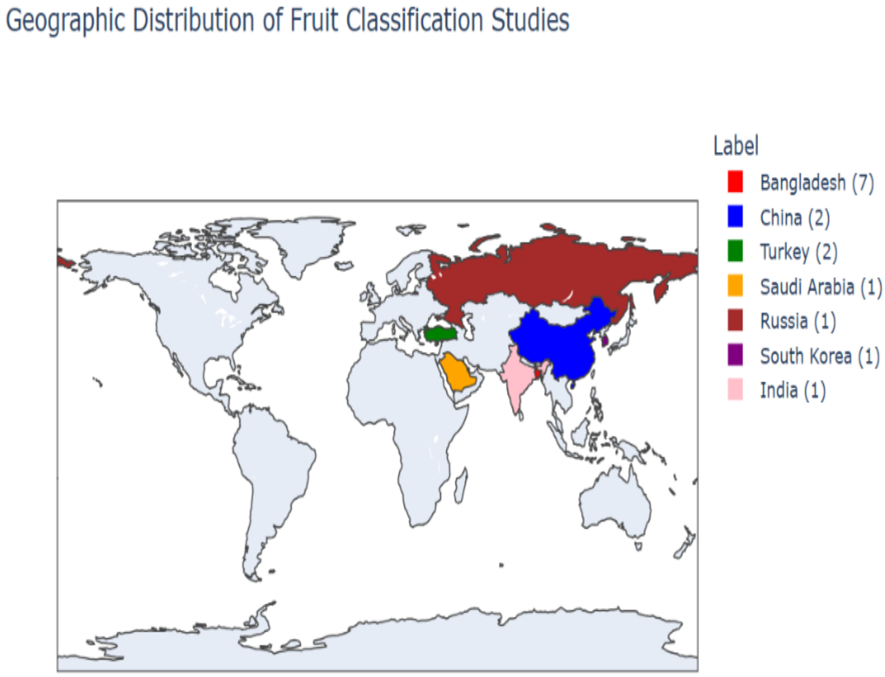

Figure 11: Geographic distribution of fruit classification studies

2.5 Findings from the Review

In this section, the main findings of the chosen researches, which were revealed during the systematic literature review, are synthesized. The results are not discussed paper by paper but are presented in a system of major research themes that are seen throughout the literature. These topics showcase the prevailing lines of research, popular datasets and models, experimentation and shortcomings of research in the field of fruit image classification. Arranging the literature thematically enables a better understanding of methodological patterns and gaps in research in relation to the goals of this thesis.

2.5.1 Application Areas of Fruit Classification

The study of machine vision in classification of fruits has witnessed a significant growth with the development of deep learning technologies. Most of the research in this field is aimed at identifying popular commercial fruits, including apples, bananas, oranges, grapes and mangoes. Such fruits are popular in international data, and are often employed as reference groups to classify agricultural images. Other works expand the classification of fruits to mixed agricultural products recognition, whereby the models are trained to recognize a variety of types of crops or plant products in a single dataset. In other instances, studies are done aiming at the classification of plant leaf type or the detection of a disease, where visual characteristics of the leaves are compared to determine a plant species or the presence of a disease on a crop. Even though these studies add to the agricultural machine vision research, their main interest is not similar to the fruit classification tasks. There are a comparatively small number of studies on rare and indigenous species of fruits and other species of fruits that are region-specific and not in international markets or publicly available datasets. Consequently, fruits of local biodiversity and cultural heritage tend to stay underrepresented in the machine vision studies. This point illustrates the need to create classification systems that target rare or native fruits, e.g., those in Bangladesh.

2.5.2 Datasets Used in Prior Studies

The datasets are significant in the training and evaluation of the fruit classification models. The research papers analyzed operate with various data sets, which may be mostly divided into publicly available data sets and individual data sets. Public datasets are popular due to their easy access and the possibility to compare the findings of the researcher with the literature. These include large datasets on fruit images that have various categories of fruits that are commonly grown. These datasets typically include thousands of labeled images and are typically used to compare deep learning models. Conversely, some studies use ad-hoc datasets gathered by scientists, particularly in the case of researching a particular species of fruits, or farming conditions. Images of orchards, farms, or regulated laboratory settings are usually used to form custom datasets. Although these datasets enable the researcher to target a given fruit, they tend to be smaller than the public datasets.

The sizes of datasets in studies on fruit classification are quite diverse. Other studies involve small datasets consisting of less than one thousand images whereas others involve large datasets consisting of several thousand images and various classes of fruits. There is also a significant variation in the number of categories of fruits across studies with small numbers of classes used to multi-class datasets with dozens of fruit types.

Another important aspect observed in the literature is image acquisition conditions. Photos may be taken in different lighting conditions, backgrounds, and camera angles, which may impact the classification performance. Moreover, there would be a problem of class imbalance in most fruit datasets, i.e. one or two of the fruit classes will have a large number of images, compared to other classes. These imbalances may affect model training and may need more data augmentation or balancing methodologies. subsection-Deep Learning and Transfer Learning Models The sphere of the fruit image classification is dominated by deep learning models, especially Convolutional Neural Networks (CNNs). Most studies use transfer learning methodology, with pre-trained models being fine-tuned on datasets of fruit images. The popularity of transfer learning is due to the fact that models can be trained to utilize features acquired with large-scale datasets, like ImageNet, to achieve higher performance in classification despite limited training data. The most popular architectures include MobileNet V2, Inception V3, ResNet variants and DenseNet. The algorithm MobileNetV2 is commonly applied in agricultural context because of its lightweight architecture and computational efficiency, it is applicable in resource-constrained settings. The inception-based models are also popular due to the capacity to model multi-scale visual features. ResNet architectures are especially well-liked because they have a residual learning architecture, allowing them to train deeper neural networks, and reducing the vanishing gradient problem. Instead, denseNet models do enhance the ability to reuse features by linking every single layer to all of the other layers in a dense block, resulting in an efficient feature propagation. Besides transfer learning models, other research works present custom CNN models directly formulated to handle the problem of fruit classification. When researchers want to create lightweight models or adjust architectures to smaller datasets, or optimize model performance to particular types of fruit, custom CNN models are frequently employed. In general, the literature shows that the transfer learning models are the most popular because of the good performance and the minimization of training.

2.5.3 Evaluation Metrics and Reported Performance

The performance of fruit classification models is evaluated with the help of evaluation metrics. Classification accuracy, the percentage of correctly classified images, is the most common metric in the reviewed studies. Besides accuracy, other metrics are reported by several studies, including precision, recall, and F1-score that give a more comprehensive picture of the model performance in various classes. They are especially helpful when there are various types of fruits in the dataset or the imbalance between classes exists. Confusion matrices are also used to visualize the results of classification between various fruit classes in many studies. Confusion matrices are used to determine which classes are often misclassified and give information on the strength and weakness of the model. In others, researchers also consider Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) to assess the classification performance. The accuracy of classification has been reported to fluctuate based on the size of data set, number of classes, and model structure. The high accuracy values can be found in studies that employ transfer learning models and well-curated datasets, indicating that deep learning methods are effective in recognizing fruit-related tasks.

2.5.4 Limitations in Existing Literature

Despite the significant progress in fruit image classification, several limitations remain within the existing literature. One major limitation is the over-reliance on common fruit datasets, where most studies focus on widely available commercial fruits. This leaves many indigenous or rare fruit species underrepresented in current machine vision research. Another limitation relates to dataset size and diversity. Many custom datasets used in fruit classification studies contain relatively small numbers of images, which may limit model generalization. Additionally, variations in lighting conditions, background environments, and image quality can affect classification performance. Class imbalance is another commonly reported challenge, where some fruit categories contain significantly more training images than others. This imbalance can bias model training and lead to uneven classification performance across different classes. Furthermore, while transfer learning models achieve strong performance, they are often trained using datasets that do not represent local biodiversity. As a result, their ability to classify rare or region-specific fruits remains limited. These limitations highlight the need for specialized datasets and classification frameworks designed specifically for rare and indigenous fruits. The identification of these challenges provides a strong motivation for the present research, which focuses on developing a machine vision–based classification system for rare fruits in Bangladesh using both transfer learning models and a custom CNN architecture.

2.6 Comparative Analysis of Selected Studies

In order to make the methodologies and results of the past studies easier to comprehend, Table 2.7 provides a comparative overview of the chosen studies. This table is a structured way to arrange major features of the literature that was reviewed so that the trends in the research, models widely used, nature of data, and limitations of various studies could be identified.

These types of comparison tables are crucial to systematic literature reviews, since they enable the synthesis of numerous studies in a concise and interpretable way. By comparison of datasets, models and performance measures, one can see methodological patterns recurring and identify gaps that encourage further research.

Table 2.7 contains the variety of small custom datasets such as 480 images in Mia et al., 2019 and huge ones of over 170,000 images[25]. Various methods are also discussed in the reviewed works, including conventional machine learning (SVM), custom CNN architectures, transfer learning models, and hybrid or attention-based deep learning models.

Table 2.7: Comparative Summary of Fruit Classification Studies

Study	**Dataset / Fruit Types**	**Method**	**Transfer Learning Model**	**Reported Accuracy**	**Key Limitation**
Rahman et al. [24]	3,240 images, 8 Bangladeshi local fruits	CNN	MobileNet, ResNet-50, VGG-19, Inception-V3	99.21%	Limited to selected local fruits
Sultana et al. [25]	4 datasets (>170,000 images)	Custom CNN (XAI-FruitNet)	MobileNetV2, ResNet50, VGG16 (compared)	97–99%	High model complexity
Mia et al. [26]	480 images, 6 rare fruits	SVM + feature extraction	None	94.79%	Very small dataset
Gulzar et al. [27]	Multiple datasets (100–120,000 images)	Review (CNN-based studies)	VGG, ResNet, MobileNet, DenseNet	>95% (varies)	Not experimental (review study)
Al Haque et al. [28]	15,000 images, 5 mango varieties	Custom CNN	None	92.80%	Limited fruit variety
Min et al. [29]	3,500–70,000 images, 50–107 fruits	Attention-based CNN (MSANet)	VGG, ResNet, DenseNet (baseline)	∼98–99%	High computational cost
Kutyrev et al. [30]	25,000+ orchard images (apple)	CNN + RNN	None	94.1%	Focused only on apples
Ukwuoma et al. [31]	Fruit-360 (81,226 images)	Review / CNN	ResNet-50	99%	Review-based, not original dataset
Mukhiddinov et al. [32]	12,000 → 43,667 images, 20 classes	YOLOv4-based DL	YOLOv4	73.5% (AP)	Lower classification accuracy
Hussain et al. [33]	10,000 images, 20 categories	DCNN	None	96%	Limited architecture depth
Hamim et al. [34]	12 classes (fruit + vegetable)	CNN, KNN, SVM	None	95%	Mixed dataset (not fruit-only)
Sowrna et al. [35]	2,082 images (banana)	Transfer Learning	NasNetMobile, MobileNetV3, Inception	97.12%	Small dataset
Alhawas & Tüfekci [36]	∼1,600 images, mango varieties	Transfer Learning + ML hybrid	MobileNet, ResNet50	∼100%	Small dataset, overfitting risk
Sukhetha et al. [37]	90,483 images, 131 classes	ML + DL	ResNet (pre-trained)	95.83%	High computational cost
Al Haque et al. [38]	10,000 images, banana varieties	Custom CNN	None	93.4% – 98.3%	Limited to banana dataset

2.7 Research Gaps Identified from the SLR

According to the comparative analysis of Table 2.7, there are certain research gaps, which are determined:

1. Limited Focus on Rare and Indigenous Fruits: The majority of the studies are about commercially common types of fruits or fruit categories (e.g. mango, banana, apple). Although the use of local datasets is performed, the amount of fruit classes is still scarce. There has been an apparent absence of study on the various rare or indigenous fruits, especially those native to Bangladesh [24].

2. Lack of Large Bangladeshi Rare Fruit Datasets: Although there are studies that use Bangladeshi fruits, none of the studies has a specific, well-organized dataset of rare and underrepresented fruit species in Bangladesh. This leaves a big gap in the availability of data sets as well as regional AI research in the agricultural sector

3. Reliance on Large Data: Most successful designs are based on large-scale data (e.g., >100,000 images). But this is not possible with rare fruits. Small datasets are usually limited by the overfitting and decreased generalization [26].

4. Lightweight CNNs Exploration: Despite the prevalence of transfer learning, less research has been done on lightweight and task-specific CNNs that are optimized to:

- small datasets
- low computational environments
- rare fruit classification

5. Lack of Generalization Analysis: High accuracy (as high as -100 percent) has been reported in many studies but,

- is not tested robustness in real-world conditions.
- do not perform in a variety of environments.

This brings about issues of model generalization.

6. Narrow Application Scope A number of studies are confined to:

- single fruit type
- controlled datasets
- laboratory conditions

This limits real-world applicability.

7. Lack of strong connection with Biodiversity Preservation: Most research treats the task of classifying fruit as a technical AI problem, without taking into consideration:

- biodiversity conservation
- cultural significance
- knowledge of rare fruits.

It is evident that there is lacking research on integrating deep learning-fruits classification with the detection and conservation of rare and indigenous fruits, specifically in Bangladesh. The presence of this gap inspires the creation of a machine vision-based classification system particularly targeted at rare fruits both with transfer learning and custom CNN methods.

3 Theoretical Foundations

3.1 Image Processing Fundamentals

Image processing is an essential part of computer vision systems, which makes machines process and interpret visual information. It is a sequence of calculation processes that are used on digital images to improve quality, extract valuable information and prepare data to be subjected to further processing like classification or recognition. Applying machine vision to identify fruits, image processing is the first step that converts raw visual information into structured representations that can be used in deep learning networks [7] .

An average image processing pipeline can be divided into three key steps: image acquisition, preprocessing and feature extraction. The raw input data is initially captured in the form of images with the use of cameras or sensors. Such images usually have noise, lighting changes, background distraction, and distortions that might have a detrimental impact on the performance of the model. Thus, resizing, normalization, filtering, and contrast enhancement are some of the preprocessing techniques that are used to standardize the input and enhance the quality of the data.

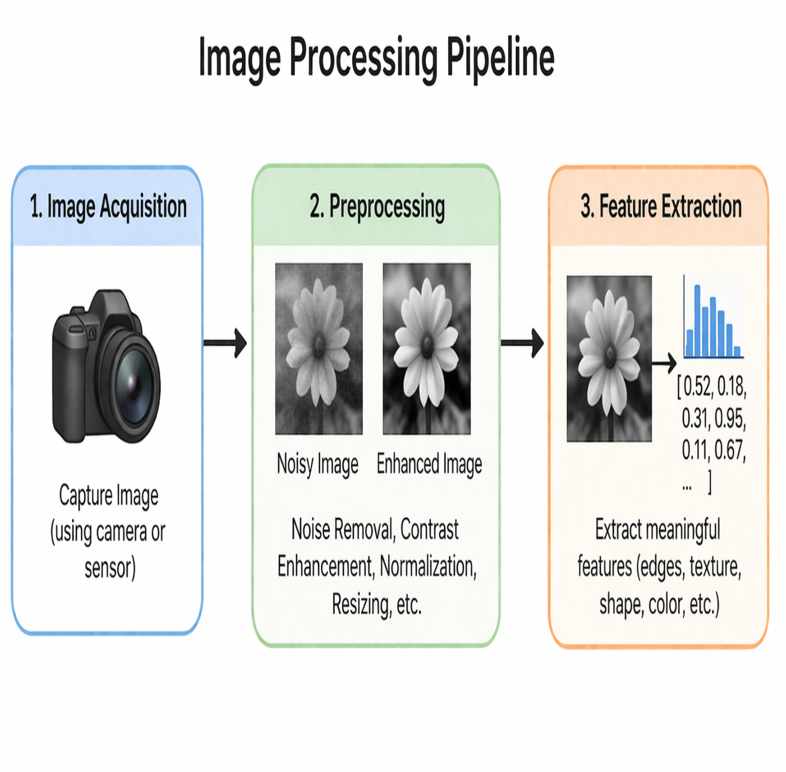

Figure 12: Image Processing Pipeline

Preprocessing is an essential step of maximizing the visual features that are relevant, and minimizing the irrelevant ones. A variety of techniques, like grayscale conversion, noise reduction, histogram equalization, and geometric transformations, are most popular in enhancing the clarity and uniformity of images. As an example, contrast-limited adaptive histogram equalization (CLAHE) and noise reduction preprocessing techniques have been shown to greatly enhance the visibility of features and model accuracy in medical image classification problems. On the same note, during tasks that involve classification of fruits, pre-processing can enable the identification of color, texture, and shape characteristics, which are critical during proper classification.

After preprocessing, feature extraction is carried out to extract significant patterns in the image. Conventional methods were based on manual features including color histograms, texture descriptors and edge detectors. Nonetheless, such approaches have a tendency to be poorly generalized and demand domain knowledge. In modern methods, deep learning models are employed especially Convolutional Neural Networks (CNNs), that learn hierarchical representations of features directly through pixel data. In the early layers, CNNs can extract low-level features, like edges and textures, and high-level semantic features, like those extracted in deeper layers, which greatly enhance the classification performance.

CNN-based image processing is effective in that it can directly learn feature mappings using raw pixel intensities, without the need to manually engineer features. As earlier mentioned, CNN architectures have the ability to automatically synthesize convolutional kernels that encode spatial relationships and patterns in images, and thus are well adapted to image classification problems. This feature is especially significant in the case of separating visually similar species of fruits, where there should be minor differences in color and texture that have to be detected.

In short, the basics of image processing form the basis of machine vision systems by making sure that the input images are clean, uniform, and informative. Combining preprocessing with deep learning-based feature extraction allows effective and efficient classification, which is the foundation of image recognition in the present-day world.

3.1.1 Digital Image Representation

A digital image is a mathematical description of a visual scene, in the form of a two-dimensional array of discrete elements called pixels. The pixels are identified by a given spot in the picture and hold intensity values which define colour or brightness. Digital image representation is a vital concept in image processing because it determines how visual information is represented, manipulated and analyzed using computational systems. Formally, a digital image can be represented as a function:

$$f(x, y)$$

where x and y denote spatial coordinates, and the value of $f(x, y)$ represents the intensity or color at that point. The discretization of this function results in a matrix of pixel values, enabling efficient processing by computer algorithms.

Digital images can be categorized into several types based on their representation:

1. **Grayscale Images**

   Grayscale images only have one channel of intensity, with each pixel value being the amount of brightness. The intensity values are usually between 0 (black) up to 255 (white) in a 8-bit representation. These pictures are computationally efficient and are commonly employed in the cases where color information is not needed.

2. **Color Images**

   Images that have colors are usually represented in the RGB (Red, Green, Blue) color model, and one pixel comprises three channels that represent various color elements. Intensity values are stored in each channel and the combination of these channels gives a broad range of colors. Fruit classification tasks cannot be done without color images, color is very important in differentiating various fruits.

3. **Multichannel and Tensor Representation**

   Images, which are used in deep learning, are modeled with multi-dimensional tensors. As an example, a $224 \times 224 \times 3$ size color image can be represented as a $224 \times 224 \times 3$-sized tensor. These tensors can be fed to CNN models and can be processed in batches and computed in parallel.

   The spatial dimensions of a digital image determine the resolution of the image, and the bit depth of the image determines the number of possible intensity values per pixel. Increased resolution and bit depth offers more precise information but makes the computations more complex.

   Quantization and sampling are also involved in digital image representation. Sampling is the conversion of a continuous image to discrete spatial values, and quantization is the allocation of discrete values to the intensities of each pixel. The digital image quality and fidelity is determined by these processes.

   When using deep learning, pixel values are commonly scaled to a given range (e.g. [0,1] or [-1,1]) in order to make the training more stable and converge. Moreover, pictures are resized to constant size to suit the input specifications of the neural network models, including ResNet, Inception, and MobileNet.

   The current-day deep learning models consider images as structured data, with the spatial relations between pixels being maintained. Convolutional operations take advantage of these relationships by acting on local areas by using filters to extract spatial features like edges, textures and shapes. Tasks like object recognition and classification require such a structured representation.

   Moreover, more sophisticated networks like DenseNet and ResNet use effective feature propagation and reuse networks that enable deeper networks to learn rich representations without losing important information. These models directly work with digital image representations, where precise and efficient image encoding is essential.

   To sum up, the digital image representation offers the mathematical and calculating structure of visual data processing. It allows converting images into structured numerical formats, which can be analyzed using advanced algorithms and deep learning models to accurately and efficiently analyze images, which is a fundamental building block in machine vision-based fruit classification systems.

3.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are considered to be one of the most powerful developments in deep learning, especially in the field of computer vision. CNNs are modeled specially to take in grid-like data structures like an image, which has important spatial interactions among pixels. In contrast to conventional machine learning methods, which often heavily depend on manually-engineered features, CNNs learn hierarchical feature representations directly on raw image data, which is why they are ideally applied to image classification problems like fruit recognition [54].

The CNN based models have demonstrated impressive success in multiple applications in recent years such as medical imaging, facial recognition, food classification and agricultural image analysis. Their success lies in their capacity to capture local and global patterns in images and be computationally efficient due to parameter sharing and local connectivity [53].

3.2.1 Basic Structure of CNNs

A typical CNN model consists of several layers with each one performing a particular transformation of the input image. The layers operate one after another to obtain significant features and do classification.

a. **Convolutional Layers :** The basic element of a CNN is the convolutional layer. It uses a collection of filters (kernels) that can be learned on the input image to produce feature maps. Each filter can identify certain visual patterns like edges, textures or shapes.

b. **Activation Functions :** After convolution, the activation functions add non-linearity to the model, allowing it to learn complicated input-output mappings. The most popular activation function is the Rectified Linear Unit (ReLU), which is computationally efficient and reduces the vanishing gradient issue.

c. **Pooling Layers :** The spatial dimensions of feature maps are reduced by pooling layers, which helps to reduce computation complexity and avoid overfitting. Max-pooling is a common one, which means that the value in a region that is the largest is taken to represent the region. This is done to make sure that the most prominent features are kept.

d. **Fully Connected Layers :** Dense (full connected) layers are normally located in the terminal of the network. They combine features that have been extracted by other layers and classify. All the neurons in a fully connected layer are linked to all the neurons in the preceding layer allowing global reasoning across the extracted features.

e. **Output Layer (Softmax) :** A CNN often has a softmax layer as its final layer, which transforms the output into a probability distribution among a set of known classes. This enables the model to provide confidence scores to every class.

3.2.2 Working Mechanism of CNNs

The CNNs are based on hierarchical feature learning. The network in the first layers extracts low-level properties like edges and plain textures. The more depth, the more intricate structures such as forms and patterns are identified in the middle layers. Lastly, deeper layers detect high-level semantic features which are related to particular objects or classes.

Such a hierarchical representation allows CNNs to acquire strong and discriminative features that have proved to be very useful in image classification tasks. The combination of feature extraction and classification into an end to end system is a major step towards enhancing performance over the traditional method.

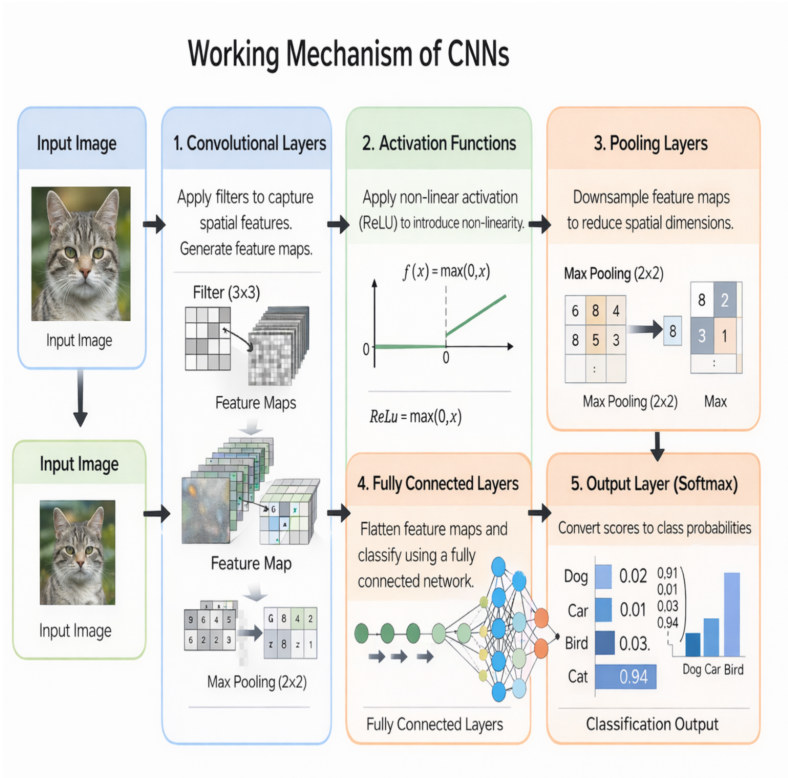

Figure 13: Working Mechanism of CNNs

3.3 Transfer Learning Theory

Transfer learning is a strong method in deep learning that allows reusing pre-trained models trained on large-scale data to perform new, yet related tasks. Transfer learning builds on the knowledge that has been learnt in the past and does not have to train a model as it would require much less time and less computation cost, and large labeled datasets are not required.

Transfer learning has been found to be very useful in image classification problems, particularly in areas like rare fruit identification where data availability is often an issue. General visual features (e.g., edges, textures, shapes) that are learned by pre-trained models based on large datasets like ImageNet can be fine-tuned to domain-specific classification problems [36].

Transfer learning works under the principle of transferring learning to a target domain (e.g., rare fruit dataset) based on a source domain (e.g., ImageNet dataset). The bottom layers of a trained network learn more general features, with the upper layers more task-specific.

Transfer learning has two common strategies:

a. **Feature Extraction :** The pre-trained layers are frozen and the last classification layers are trained on a new dataset.

b. **Fine-Tuning :** Part or all of the pre-trained layers are re-trained in addition to the addition of new layers to fit the target data better.

Fine-tuning tends to yield higher performance with enough training data, whereas feature extraction can serve with smaller data sets.

Transfer learning has a number of benefits:

1. Minimizes training and computational time.
2. Performs better on limited data in a model.
3. Avoids overfitting by making use of pre-learned features.
4. Allows the exploitation of rich architectures without huge datasets.

Transfer learning is especially applicable in the agricultural context, e.g. in fruit classification, where annotated data is hard to gather in large amounts.

3.3.1 Mathematical Formulation of Transfer Learning

Transfer learning is a basic concept in deep learning that allows the re-use of knowledge from one domain (source) to improve learning in another domain (target). In formal terms, the source domain can be expressed as $D_s = \{X_s, Y_s\}$, where X_s is the input feature space and Y_s is the corresponding label space [47]. Similarly, the target domain is defined as $D_t = \{X_t, Y_t\}$, where X_t and Y_t represent the input data and labels of the rare fruit classification problem.

In real-world applications, the data distributions of the source and target domains are usually different, i.e.,

$$P(X_s, Y_s) \neq P(X_t, Y_t),$$

and therefore require transferring learned representations rather than training a model from scratch. Transfer learning aims to learn a target predictive function $f_t(\cdot)$ using the knowledge from a pre-trained source model $f_s(\cdot)$ to achieve better performance, reduced training time, and to overcome the limitations of small datasets.

The relationship between the source and target learning processes can be expressed as:

$$f_t(X_t) = f_s(X_s) + \Delta,$$

where $f_s(\cdot)$ is the learned mapping in the source domain and Δ is the adaptation term that adjusts the model parameters for the target task.

Within the framework of Convolutional Neural Networks (CNNs), the transferred knowledge mainly consists of convolutional layers that act as hierarchical feature extractors. The transformation of an input image X into a feature representation F is defined as:

$$F = \sigma(W * X + b),$$

where W represents convolutional filter weights, b is the bias term, $*$ denotes the convolution operation, and $\sigma(\cdot)$ is a non-linear activation function such as the Rectified Linear Unit (ReLU). This process enables the network to learn multi-level features, from low-level patterns like edges and textures to high-level semantic representations [16] [19].

After feature extraction, these representations are passed through fully connected layers for classification. The output layer typically uses the SoftMax function to convert raw scores into normalized probabilities:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}},$$

where z_i is the logit for class i and N is the total number of classes.

The predictions are then compared with ground truth labels to compute the loss, which is used to optimize the model parameters during training. The most commonly used loss function for multi-class classification is the categorical cross-entropy loss:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i),$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability for class i.

This loss is minimized using backpropagation, allowing the model to refine its weights—particularly in the higher layers—while preserving the generalized feature representations learned from the source domain. This mathematical formulation provides a solid foundation for transfer learning, demonstrating how pre-trained CNN models can be effectively fine-tuned to achieve high accuracy in rare fruit classification tasks with limited training data.

3.4 Optimization Algorithms

Optimization algorithms are instrumental in training deep learning models by minimizing the loss function and updating network parameters iteratively. In convolutional neural networks (CNNs) and transfer learning models such as ResNet-50, InceptionV3, MobileNetV2, and DenseNet-121, optimization algorithms determine how efficiently the model converges to an optimal solution. The primary objective is to find the parameter set θ that minimizes the loss function $\mathcal{L}(\theta)$:

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta)$$

where θ^* represents the optimal parameters.

The loss function for classification tasks is typically categorical cross-entropy:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability.

Gradient Descent (GD)

Gradient Descent updates parameters in the direction opposite to the gradient:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta_t)$$

where η is the learning rate.

Stochastic Gradient Descent (SGD)

To reduce computational cost, SGD updates parameters using mini-batches:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}_i(\theta_t)$$

where $\mathcal{L}_i$ is the loss for a mini-batch.

Momentum

Momentum improves convergence by incorporating past gradients:

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta \mathcal{L}(\theta_t)$$

$$\theta_{t+1} = \theta_t - v_t$$

where $\gamma \in (0, 1]$ is the momentum coefficient.

AdaGrad

AdaGrad adapts the learning rate for each parameter:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{G_t + \epsilon}} \nabla_\theta \mathcal{L}(\theta_t)$$

where G_t is the sum of squared gradients and ϵ prevents division by zero.

RMSProp

RMSProp uses an exponentially weighted moving average of squared gradients:

$$E[g^2]_t = \beta E[g^2]_{t-1} + (1 - \beta)(\nabla_\theta \mathcal{L}(\theta_t))^2$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} \nabla_\theta \mathcal{L}(\theta_t)$$

where β is the decay rate.

Adam (Adaptive Moment Estimation)

Adam combines Momentum and RMSProp by computing first and second moment estimates:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_\theta \mathcal{L}(\theta_t)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)(\nabla_\theta \mathcal{L}(\theta_t))^2$$

Bias-corrected estimates:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

Parameter update rule:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t$$

Adam is widely used due to its fast convergence and robustness in deep networks. Deep CNNs are highly sensitive to optimization strategies, which significantly influence training efficiency and performance. Techniques such as batch normalization, proper weight initialization, and learning rate scheduling further enhance optimization by stabilizing gradient flow and improving convergence. Architectures like DenseNet improve gradient propagation through dense connections, reducing vanishing gradient problems. Choosing an appropriate optimization algorithm and tuning hyperparameters such as learning rate, momentum, and batch size are critical for achieving high accuracy and efficient training in deep learning-based image classification systems.

3.5 Evaluation Metrics Formulation

The performance evaluation of deep learning–based image classification models is a crucial step in assessing the effectiveness and reliability of the proposed system. Various evaluation measures are used in this study to understand the overall performance of the transfer learning models (ResNet-50, Inception V3, MobileNet V2, and DenseNet-121) in recognizing rare fruits. These measures are based on the confusion matrix, which provides a detailed breakdown of predictions.

The basic elements of a confusion matrix in a multi-class problem include True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [45] . These components represent correctly and incorrectly classified instances and form the basis of all evaluation metrics.

Accuracy

Accuracy is the most common performance measure, representing the proportion of correctly classified samples out of the total number of samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Although accuracy provides a general overview, it may not be sufficient when dealing with imbalanced datasets, which is common in rare fruit classification.

Precision

Precision measures the proportion of correctly predicted positive cases out of all predicted positives:

$$\text{Precision} = \frac{TP}{TP + FP}$$

A high precision indicates a low false positive rate, meaning the model rarely misclassifies other classes as the target fruit.

Recall (Sensitivity)

Recall, also known as sensitivity, measures the proportion of actual positive cases correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN}$$

High recall indicates that the model successfully detects most of the actual fruit instances, minimizing false negatives.

F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balanced evaluation:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

This metric is particularly useful when dealing with class imbalance.

Specificity

Specificity measures the proportion of correctly identified negative cases:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

It reflects how well the model distinguishes non-target fruit classes.

Cross-Entropy Loss

Categorical cross-entropy loss is used during model optimization to measure the difference between predicted probabilities and true labels:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

where y_i is the true label, $\hat{y}_i$ is the predicted probability for class i, and C is the total number of classes.

ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve evaluates classification performance by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR):

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}$$

The Area Under the Curve (AUC) summarizes the model's ability to distinguish between classes. A higher AUC indicates better performance.

Confusion Matrix Representation

The confusion matrix provides a class-wise performance analysis by highlighting frequently misclassified fruit categories. It enables further refinement of the model.

Confusion matrix-based evaluation metrics are widely used in deep learning classification tasks as they provide a comprehensive assessment beyond simple accuracy. These metrics are particularly important in image classification problems where balanced evaluation is required.

Overall, the combination of these evaluation metrics ensures a robust and comprehensive assessment of the proposed fruit classification system. By incorporating both probability-based loss functions and classification performance metrics, the framework effectively evaluates the reliability, generalization capability, and discriminative power of the deployed deep learning models.

4 Dataset Preparation and Preprocessing

4.1 Overview of Dataset

The dataset applied in the current study is a self-collected image dataset which has been created specifically to facilitate the classification of rare and indigenous fruits in Bangladesh. There were six types of fruits that were picked because of limited availability and increasing loss of popularity among the general population. Such fruits are Bilimbi, Kamranga (Star Fruit), Panifol (Water Chestnut), Meshta (Roselle), Kathbadam (Tropical Almond) and Sharifa (Custard Apple). The photographs were taken in different local places like rural villages and local markets, and the data sets are representative of real agricultural and environmental conditions. All the pictures were taken with a mobile phone under natural conditions without any controlled laboratory conditions to maintain natural variations [8] .

The main reason why this dataset was constructed is that there is a necessity to record and raise awareness of these underrepresented fruit species. Most of these fruits are of great nutritional value and are culturally and traditionally important in Bangladesh. But with less cultivation, less commercial activity, and decreased public awareness they are also slowly becoming extinct. This has led to a deficit of publicly accessible datasets which concentrate on such native fruits. This dataset is thus not only intended to aid in machine learning-based classification but also to add to the biodiversity awareness and the digital preservation of these species [11].

Composition wise, the dataset has six equal classes with 300 images and hence the total number of images is 1800. The dataset was split into three subsets [13], namely, 70

The dataset represents the broadest possible variation of real-world variations, such as variations in lighting conditions, background complexity, object orientation, and scale. These are designed to maintain such variations to increase the generalization ability of the deep learning models so that they can be effective at operating on unseen data. What is special about this dataset is that it is concentrated on rare Bangladeshi fruits and its realistic image conditions whereas the commonly used benchmark datasets lack such realistic image conditions. It can therefore be used as a useful training and evaluation resource of convolutional neural network based models, and extends to other domains of research in agricultural technology, biodiversity conservation and preservation of indigenous knowledge.

4.2 Data Collection Process

The data collection process used in this research was well planned to make sure that high quality and diverse data is developed to be used in deep learning fruit classification. The

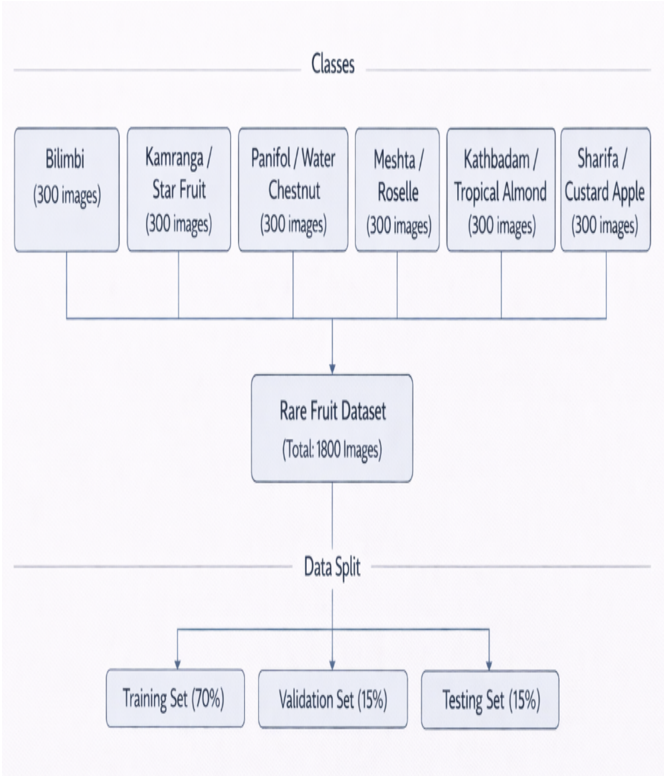

Figure 14: Structure of the rare fruit dataset showing class distribution and data split

whole process comprised several steps which included field trips, harvesting of fruits in different sources, controlled image acquisition, and data preparation.Figure 14 is a visual representation of the entire data collection workflow.X, describing the steps in the process followed in sequence.

4.2.1 Source of Data and Field Collection

The data in this work was completely gathered in the course of the significant fieldwork, without any use of the existing data sets or web resources of images. About 300 pictures were taken of each type of fruit, which made the dataset balanced in all the categories. The fruits were gathered at various natural sites, such as in villages, local markets, and right on the trees of farmers with their permission. In more instances, fruits were hand picked or bought in other areas, to have variety in samples. This method allowed the dataset to represent real environmental and agricultural environment, which is crucial to develop a solid and generalizable classification system.

4.2.2 Tools and Image Acquisition Setup

Every picture was taken with the help of a smartphone camera, which made them accessible and convenient in data gathering. This study did not use any web scraping tools or publicly available datasets. Three other artificial light sources were used to ensure proper lighting to achieve image clarity and consistency. Various background configurations were introduced with colored paper (white, green, yellow and blue) to minimize the background bias and enhance model generalization. They were taken in different angles such as top view, side view and at about 45 O and 90 O angles to introduce geometric variation.

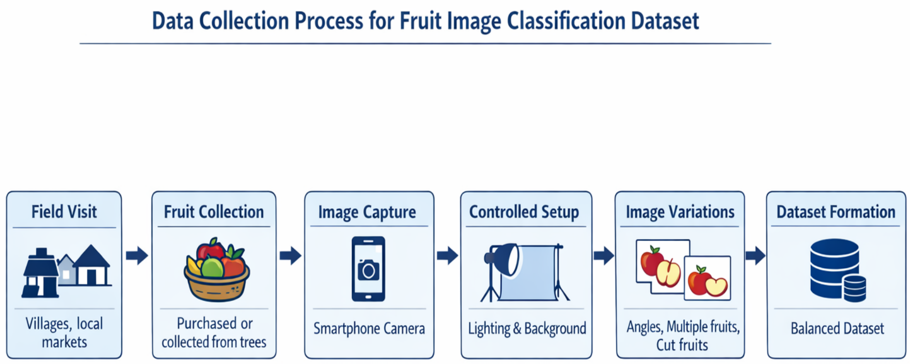

Figure 15: Workflow of the data collection process for Fruit Image Classification Dataset

4.2.3 Image Variations and Dataset Diversity

Image variations of various types were added to increase the data richness. These were single images of fruits, several fruits in one group and organized cluster images of fruits. Besides, cut fruits were photographed to incorporate internal texture and structural characteristics, which may be helpful in classification. This type of diversity in terms of composition, orientation, and visual appearance assists the model in learning stronger and discriminative features.

4.2.4 Data Collection Time

The process of data collection was carried out in about two months. The selected fruits were seasonal and rare hence there was a need to plan in order to gather enough samples. The timeline of the collection was based on the availability of fruits, which differed by geographical location and time of year, so the process was time-consuming.

4.2.5 Challenges and Dataset Balancing

A number of obstacles were experienced in the collection of data. The lack of rare fruits had a major effect in the availability of samples in the initial stage and seasonal dependence was also a hindrance to the collection process since most of these fruits could only be available during certain times of the year. Additionally, to determine the suitable collection sites, preliminary research on fruit-producing areas, harvesting periods, and local supply was necessary, which led to several field trips to various rural and semi-urban regions.

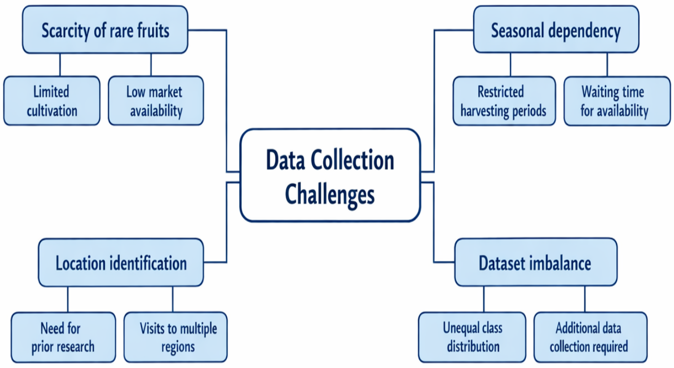

Figure 16: Overview of key challenges encountered during data collection

Table 4.1: Summary of the Data Collection Process

Aspect	**Description**
Source	Field collection
Locations	Villages, markets, farms
Tool	Smartphone
Lighting	Natural and artificial (three light sources)
Backgrounds	White, green, yellow, blue
Angles	Multiple perspectives (45°, 90°, etc.)
Duration	Two months
Images per class	Approximately 300

4.3 Data Cleaning and Annotation

4.3.1 Data Cleaning and Preprocessing

Several filtering measures were used to eliminate poor quality and irrelevant samples in order to enhance the reliability of the datasets. All of the collected images were first inspected manually to remove noisy or otherwise inappropriate images. Removal of Blurry images: Photos with low resolution, motion blur or poor focus were discarded, since they interfere with the learning of discriminative features, i.e., texture and shape, by the model. Visual clarity was considered as a condition of necessity to include datasets. Removal of Duplicate Images: Duplicating and near-duplicating images were detected using visual inspection and file comparison method. Extra samples were later eliminated to avoid biases and overfitting of models to particular fruits.

4.3.2 Label Verification and Quality Assurance

Manual Annotation and Review: All the pictures were first subjectively assigned their respective categories of fruits. Caution was followed to make an accurate identification, especially with similar rare species of fruits. Once labeled, the dataset was checked several times using a manual inspection to define and fix any discrepancies or ambiguous samples. Domain Knowledge–Based Validation: The cross-validation of labeling decisions was based on domain knowledge such as familiarity with indigenous fruits of Bangladesh and referring to relevant agricultural references when needed. This was done to make sure that the right category of fruits was identified, according to visual and contextual features. Iterative Refinement: The data was narrowed down by reading the data several times by hand. Any mislabeled, ambiguous and poor quality samples were revised or eliminated to enhance the overall reliability of the dataset. This process of iteration guaranteed high annotation accuracy devoid of the use of external automated or online verification tools.

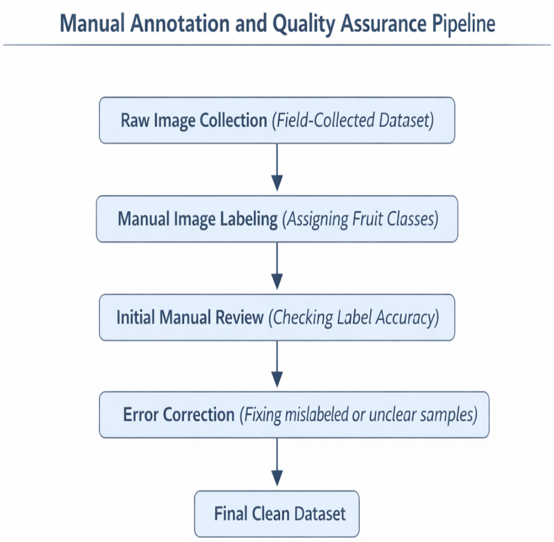

Figure 17: Manual annotation and quality assurance pipeline

4.4 Image Preprocessing

Image preprocessing plays an important role in the proposed machine vision pipeline because it guarantees that all received images have been converted to a uniform and a model-friendly format before training and inference. Since the obtained dataset is made up of heterogeneous sources (differing resolutions, lighting, and backgrounds), the data requires preprocessing to increase the quality of the data, decrease the complexity of the computations, and increase model generalization. Deep learning models, especially Convolutional Neural Networks (CNNs), can be prone to unsteady convergence and decreased classification accuracy without suitable preprocessing.

4.4.1 Image Resizing

The entire dataset was downsampled to a common input size of 224 x 224 pixels, a standard input size of various popular transfer learning models like MobileNetV2, ResNet-50, DenseNet-121 and InceptionV3. This uniform resizing can make these architectures compatible with the input layer and greatly lower the computational overhead.

Let an input image be represented as $I \in R^{H \times W \times C}$, where H, W, and C denote height, width, and color channels, respectively. The resizing operation transforms the image into:

$$I' \in R^{224 \times 224 \times 3}$$

This conversion retains the important visual characteristic and standardizes the size of the input in the dataset.

4.4.2 Color Space Standardization

All the pictures were turned and stored in RGB (Red, Green, Blue) color code. As the dataset contains images of fruits, color is an important factor that helps to differentiate between classes, and thus, it is important to preserve RGB data to classify them correctly.

Every image is given as a 3-channel tensor:

$$I(x, y) = [R(x, y), G(x, y), B(x, y)]$$

The CNN can be trained to learn color-based discriminative representations, including ripeness, texture variations and species-specific colorings, due to this representation.

4.4.3 Pixel Normalization

To achieve numerical stability and speed up convergence in training, the intensity values of pixels were brought to a standard range. Originally, pixel values of an image are contained between $[0, 255)$. These values were normalized using min-max normalization to the range $[0, 1]$ as follows:

$$I_{\text{norm}} = \frac{I}{255}$$

This standardization causes the input features to be smaller, eliminates gradient explosion, and helps to optimize faster in backpropagation. It also provides consistency between batches of data.

4.4.4 Noise Reduction and Image Quality Enhancement

Simple noise cancelation methods were used to enhance clarity of the images and eliminate undesired artifacts. Though the dataset had been pre-cleaned to eliminate badly damaged pictures, a small amount of noise caused by lighting variations and compression artifacts were felt using smoothing methods like Gaussian filtering (where necessary).

Gaussian smoothing operation can be written as:

$$I_{\text{smooth}}(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x - i, y - j) \cdot G(i, j)$$

where $G(i, j)$ represents the Gaussian kernel. This operation helps reduce high-frequency noise while preserving important structural features.

4.5 Data Augmentation

The main reason why data augmentation was implemented in the current study is to overcome the difficulties related to small and imbalanced sets. Deep learning models, especially the Convolutional Neural Networks (CNNs), are extremely data-sensitive and easily overfit on limited samples. Overfitting is a phenomenon where a model learns specifics of the training data instead of learning features that can be generalized [24]. The model is trained to learn discriminative and invariant features in varying transformations by using augmentation methods. This goes a long way in enhancing the capacity of the model to make predictions on unobservable data. In addition, augmentation emulates the real world conditions like changes in orientation, lighting and scale, which are typical in the real world in agriculture.

Table 4.2: Summary of Data Augmentation Parameters

Technique	**Description**	**Purpose**
Rotation	Random rotation up to $\pm 45°$	Handle orientation variance in fruit images
Width/Height Shift	Translation up to 25%	Improve robustness to positional shifts
Shear	Shear transformation up to 25%	Enhance geometric invariance
Zoom	Zoom range up to 25%	Handle scale variations of objects
Horizontal Flip	Mirror image horizontally	Increase dataset diversity and generalization
Brightness Adjustment	Vary pixel intensity	Simulate different lighting conditions

4.5.1 Integration with Preprocessing Pipeline

The augmented images were then scaled to a standard size of 224 x 224 pixels, corresponding to the input size of the deep learning structures used. Moreover, normalization was also done to convert pixel values to the range [0, 1] that ensures numerical stability and convergence is faster during training. The augmentation was done only to the training dataset whereas validation and test datasets were kept constant to make the model evaluation unbiased.

4.5.2 Impact on Model Generalization

Data augmentation is important to enhance the generalization ability of proposed models. The models are trained to identify the features of the fruit regardless of the changes in the orientation, scale, and lighting conditions by introducing controlled randomness into the training data. Consequently, the trained models are shown to perform better on unseen data, less overfitting, and are more robust in real-world applications. This is especially significant in the agricultural field where there is inevitability of changes in the environment.

5 Proposed Methodology

5.1 Research Framework

This study suggests an elaborate and well-organized deep learning-based system to classify rare and indigenous fruit of Bangladesh using machine vision procedures. The architecture is built to combine all the important steps of a modern computer vision pipeline such as dataset acquisition, preprocessing, augmentation, model development, training, and evaluation to guarantee robust, scalable and high-performing classification. First, a self-assembled dataset is created that has 1,800 images in six different types of fruits: Bilimbi, Custard Apple, Roselle, Star Fruit, Tropical Almond, and Water Chestnut. The photos are taken in actual life situations with a mobile camera, with deliberate differences in light, the complexity of the backgrounds, the orientation and size of the objects. Such diversity is critical to enhancing generalization ability of deep learning models and their relevance to real-life agricultural settings.

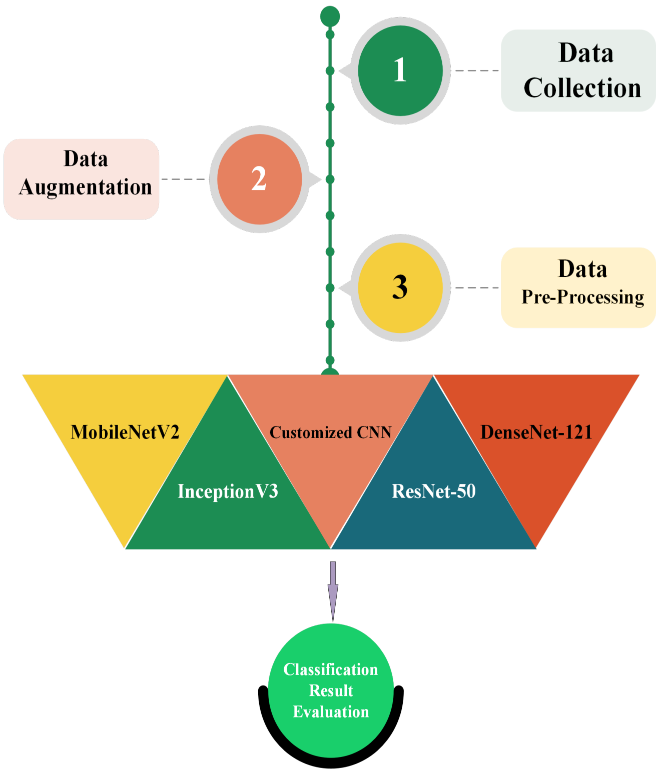

Figure 18: Proposed Research Framework

After the acquisition of data, preprocessing operations are used to standardize the data. The images are all re-scaled to the same spatial dimension to ensure consistency across inputs, and pixel-values are scaled to a common range, which stabilizes gradient-based learning, and speeds up convergence in training. The normalization process is characterized as:

$$X_{norm} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

In order to further increase the robustness of the models and reduce overfitting because of the low size of the dataset, data augmentation is only applied to the training set. These are geometric and photometric transformations, i.e. rotation, translation, zooming, shearing, and horizontal flipping, which artificially increase the dataset, and enable the model to learn both the invariant and discriminative features. The data is then split into three subsets, namely training (70%), validation (15%) and testing (15%) to guarantee an adequate model training, hyperparameter optimization and an objective performance assessment. The division of the dataset may be expressed in the following form formally:

$$D = D_{train} \cup D_{val} \cup D_{test}$$

$$|D_{train}| = 0.7N, \quad |D_{val}| = 0.15N, \quad |D_{test}| = 0.15N$$

After the preprocessing, a dataset is fed into five variants of deep learning models, including four transfer learning models, such as MobileNetV2, InceptionV3, ResNet-50, and DenseNet-121, as well as a custom Convolutional Neural Network (CNN) proposed. The transfer learning models can use the pretrained weights to extract hierarchical and semantic features with an efficient level and the custom CNN is to learn task-specific representations with optimal levels of computational complexity. The main activity of the CNN is the convolution process that derives spatial features of the input images and this is mathematically defined as:

$$f_{i,j,k} = \sum_{m,n} X_{i+m,j+n} \cdot K_{m,n,k} + b_k$$

In which X is the input image, K is the convolution kernel, and b_k is the bias. The Rectified Linear Unit (ReLU) activation function is used to add non-linearity and facilitate training on the complex patterns:

$$g(x) = \max(0, x)$$

Then, the operations of pooling are applied to reduce the spatial dimensions but retain the significant features, enhancing the efficiency of computations:

$$P_{i,j,k} = \max(f_{i+m,j+n,k})$$

Fully connected layers at the last stage of the network convert the extracted features to the probability of the classes using the Softmax function:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

The training process can be summarized as the optimization of the model parameters, i.e. minimization of an appropriate loss. In this study, categorical cross-entropy is used due to its effectiveness in multi-class classification tasks:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

In which y_i is the actual label of the class and $\hat{y}_i$ is the probability prediction. Gradient-based optimizers like the Adam optimizer are used to optimize the parameters by updating the model weights on a case-by-case basis:

$$\theta = \theta - \eta \nabla_\theta \mathcal{L}.$$

After the training, the performance of the various models is measured with the help of a number of standard classification measures that give a holistic view of model effectiveness. These include accuracy, precision, recall, F1-score, Receiver Operating Characteristic Area Under Curve (ROC-AUC), and Precision-Recall Area Under Curve (PR-AUC).

Lastly, a comparative study is carried out on all the models to determine the best architecture to classify rare fruits. This analysis puts into consideration both the predictive performance and the computational efficiency and as such, the model that is chosen is the one that presents the best balance between the accuracy and the practical deployability. The suggested structure, thus, offers a highly accurate, scalable, and efficient solution to the problem of rare fruit categorization, which helps to evolve the intelligent agricultural system based on deep learning and machine vision.

5.2 Transfer Learning Models

The current research uses four state-of-the-art pre-trained CNN models widely used in transfer learning because of their good feature extraction abilities and architectural innovations.

5.2.1 MobileNetV2

The mobileNetV2 is a convolutional neural network architecture that is lightweight and is tailored to be run efficiently on mobile/embedded vision applications. In contrast to standard deep CNNs, which use computationally-intensive standard convolutions, MobileNetV2 presents a new structural design, which consists of depthwise separable convolutions, inverted residual connections and linear bottleneck layers, that allow the computational cost to be reduced dramatically without sacrificing representational performance [39].

The depthwise separable convolution is at the heart of MobileNetV2 and breaks down a regular convolution into two convolutions: a depthwise convolution and a pointwise convolution. Considering an input tensor of size $(h_i \times w_i \times d_i)$ with a typical convolution of kernel size $(k \times k)$ and (d_j) output channels, the computational cost would be:

$$\text{Cost}_{\text{standard}} = h_i \times w_i \times d_i \times d_j \times k^2$$

depthwise separable convolution on the other hand brings this cost down to:

$$\text{Cost}_{\text{depthwise}} = h_i \cdot w_i \cdot d_i \cdot k^2 + h_i \cdot w_i \cdot d_i \cdot d_j$$

$$= h_i \cdot w_i \cdot d_i (k^2 + d_j)$$

This expression shows that MobileNetV2 has roughly ( ) less computation than regular convolution (when $(k = 3)$, almost 89 times less), which is significantly more efficient on real-time tasks.

The major innovation of MobileNetV2 is the inverted residual block with linear bottleneck that is in contrast with the conventional residual networks. MobileNetV2 initially expands the low-dimensional input to a higher-dimensional space by a pointwise convolution instead of compressing features and subsequently expanding them. This is then followed by a depthwise convolution to extract spatial features and lastly one projects to a low dimensional space using a linear pointwise convolution [42] [44] . Mathematically, the transformation of an input tensor ($\mathbf{x}$) could be stated as:

$$\mathbf{y} = \mathbf{P}(\sigma(\mathbf{D}(\sigma(\mathbf{E}(\mathbf{x})))))$$

where:

- $E(\cdot)$ denotes expansion via 1×1 convolution,
- $D(\cdot)$ represents depthwise convolution,
- $P(\cdot)$ is the projection layer,
- $\sigma(\cdot)$ is a non-linear activation function (e.g., ReLU6).

Notably, the last projection layer applies a linear activation, which does not cause information losses like introducing non-linearities to low-dimensional feature spaces. The design conserves vital feature data and improves representational ability.

In MobileNetV2, the residual connection is used only when both the input and output dimensions are the same, creating an inverted residual structure:

$$\mathbf{y} = \mathbf{x} + \mathbf{F}(\mathbf{x})$$

where ($\mathbf{F}(\mathbf{x})$) is the transformation of the block in the bottleneck. In contrast to the conventional residual networks, which employ high-dimensional representations, MobileNetV2 employs the low-dimensional bottleneck layers that decrease the memory usage and enhance the efficiency[22].

The next concept is the so-called manifold of interest that states that significant characteristics in high-dimensional activation spaces are in a lower-dimensional subspace. Linear bottleneck makes sure that this manifold is maintained in transformation by preventing the collapse of non-linear activations and propagating features more effectively across layers.

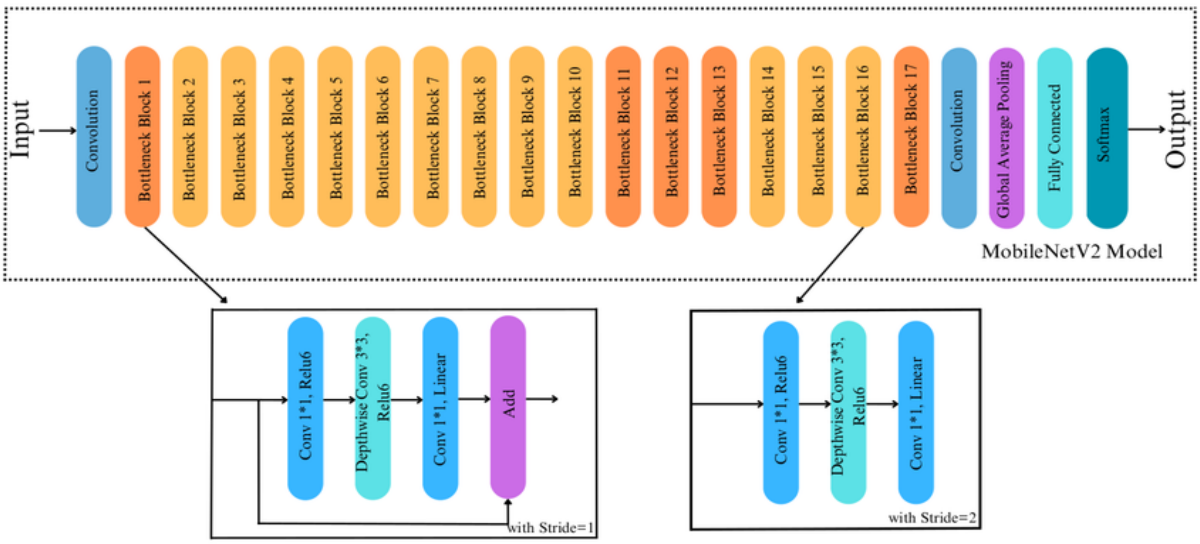

Figure 19: Architecture of MobileNetv2 [41]

MobileNetV2 is an ideal compromise between accuracy and efficiency, with few parameters and low computational costs without sacrificing or weakening the feature learning performance. It is especially applicable in applications with limited computational resources, like mobile vision systems, real-time image classification, and embedded AI solutions. Empirical experiments also indicate that MobileNetV2 has a competitive accuracy at a much lower training time and resource usage than traditional CNN models.

Overall, MobileNetV2 is an excellent improvement in the field of efficient deep learning architectures, integrating depthwise separable convolution, inverted residuals, and linear bottlenecks, and is a popular choice in transfer learning and lightweight computer vision tasks.

5.2.2 InceptionV3

InceptionV3 architecture is a highly optimized and refined deep convolutional neural network that enhances computational efficiency, as well as classification accuracy, through the use of factorized convolutions, dimensionality reduction and parallel feature extraction. It builds on the original Inception (GoogLeNet) architecture by adding novel architectural improvements to allow deeper networks at lower computational cost, yet with high representational capacity[46].

The core concept behind the use of inceptionV3 is that approximating a good sparse convolutional structure can be done in terms of dense components. The network breaks down the convolution operation into several parallel transformations instead of using a single convolution operation, enabling it to simultaneously represent multi-scale features. Let the input feature map be denoted as $X \in R^{H \times W \times C}$. The results of an Inception module may be described as the sum of various transformations that are performed simultaneously:

$$Y = \text{Concat}\,(f_1(X), f_2(X), f_3(X), f_4(X))$$

and $f_i(X)$ are various transformation directions e.g. 1×1, 3×3 or pooled convolutions. The parallel structure allows the network to learn both local and global characteristics.

An important advancement of InceptionV3 is that it factorizes convolutions thus lowering the computational cost without compromising representational ability. Convolution with a kernel size $k \times k$ costs:

$$\text{Cost}_{\text{standard}} = H \cdot W \cdot C_{in} \cdot C_{out} \cdot k^2$$

InceptionV3 splits larger convolutions into smaller ones to minimise this cost. An illustrative example is that a 5×5 convolution can be broken down into two consecutive 3×3 convolutions:

$$5 \times 5 \approx 3 \times 3 + 3 \times 3$$

that is less computation since:

$$2 \cdot (3 \times 3) = 18 < 25$$

Likewise, a 3×3 convolution can be decomposed to asymmetric convolutions:

$$3 \times 3 = (1 \times 3) + (3 \times 1)$$

This not only lowers the computational cost, but also the network depth and nonlinearity, resulting in better feature learning.

Dimensionality reduction with 1×1 convolutions (as bottleneck layers) is another crucial element. Considering an input tensor X, a 1×1 convolution will transform it as:

$$Z(i,j,k) = \sum_{c=1}^{C} X(i,j,c) \cdot W_{c,k}$$

in which W is learnable weights. This process compresses the dimensionality of the channels then applies computationally intensive convolutions, thus enhancing performance and avoiding overfitting.

The InceptionV3 also uses auxiliary classifiers to alleviate the vanishing gradient issue. These classifiers are added at the intermediate stages and they help to add to the overall loss during training:

$$\mathcal{L}_{total} = \mathcal{L}_{main} + \alpha \cdot \mathcal{L}_{aux}$$

And α is a weight. This profound supervision makes the training more stable and enhances convergence.

Moreover, the architecture uses grid size reduction methods to effectively reduce feature maps without major information loss. InceptionV3 sequentially uses stride-based convolutions and parallel pooling:

$$Y = \text{Concat}(\text{Conv}_{stride=2}(X), \text{MaxPool}_{stride=2}(X))$$

This saves more feature data and saves space.

Label smoothing regularization is another improvement in InceptionV3 to avoid overconfidence in the model. Rather than one-hot encoded label y, a smoothed label y' is defined as:

$$y' = (1 - \epsilon) \cdot y + \frac{\epsilon}{K}$$

In which ϵ is a constant which is very small and K is the number of classes. This enhances generalization and minimized overfitting.

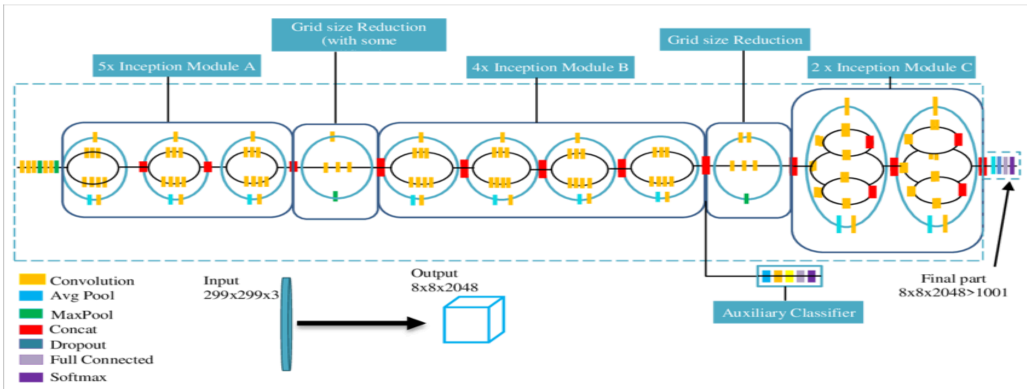

Figure 20: Architecture of InceptionV3 [47]

In general, InceptionV3 represents a tradeoff between depth, width and computational efficiency due to a combination of factorized convolutions, parallel processing, and dimensionality reduction techniques. Its design enables it to be very precise with a lower number of parameters than traditional deep networks, thereby being well suited to large-scale image classification problems and transfer learning. The design philosophies prioritize the need not to have representational bottlenecks, efficient flow of information, and optimization of computational resources which have altogether led to its excellent performance in the contemporary computer vision systems.

5.2.3 ResNet-50

Residual Neural Network (ResNet), developed by Kaiming He et al.[50] in 2015, is an important improvement in the design of deep convolutional neural networks, as it solves the degradation problem and vanishing gradient problem of very deep networks. Of the variants, ResNet-50 is popular because of the ideal compromise between computational cost and classification accuracy. Conventional deep neural networks seek to learn an approximation of a desired mapping (H(x)) itself; but as the depth of the network grows, optimization becomes challenging, and the training error can even rise with the representational capacity. To address this weakness, ResNet proposes a residual learning architecture where the network will be taught a residual (F(x)) rather than the mapping, and this can be stated as:

$$F(x) = H(x) - x$$

The mapping is therefore intended to be:

$$H(x) = F(x) + x$$

This re-reformulation results in the underlying residual block equation:

$$y = F(x, \{W_i\}) + x$$

and x is the input, $F(x, \{W_i\})$ is the residual with weights W_i and y is the output. Shortcut (skip) connections are used to add the input x to the network to support identity mapping and allow propagation of gradients across deep layers. This method eases the optimization process, with empirical learning of residuals easier than empirical learning of unreferenced mappings.

ResNet-50 consists of 50 layers, most of them being constructed on bottleneck residual blocks, which improve computation efficiency but does not affect representational power. Every bottleneck block is composed of three convolutional layers that are designed to have 1×1, 3×3, and 1×1 convolutions respectively. The initial 1×1 convolution is a dimensionality reduction, the 3×3 convolution features the spatial features and the last 1×1 convolution returns the dimensionality. The residual value of a block may be obtained in the following form:

$$F(x) = W_3\sigma(W_2\sigma(W_1x))$$

and W_1, W_2, W_3 is the convolutional weight and $\sigma(z)$ is the non-linear activation (usually ReLU). The output of the block is computed as the sum of the input through shortcut connection:

$$y = F(x) + x$$

Where input and output dimensions are different, a shortcut of projection is used with a linear transformation:

$$y = F(x) + W_sx$$

W_s where W_s is a 1×1 convolution generally to align feature dimensions.

The general structure of ResNet-50 starts with the first 7×7 convolutional layer followed by max pooling, and then the stages (residual blocks) are stacked[51]. These successively derive hierarchical characteristics, the lowest level being edges and textures and the highest level being semantic representations. The network ends with a global average pooling layer and a fully connected layer then a softmax classifier. This strong hierarchical model allows it to learn features efficiently with manageable complexities.

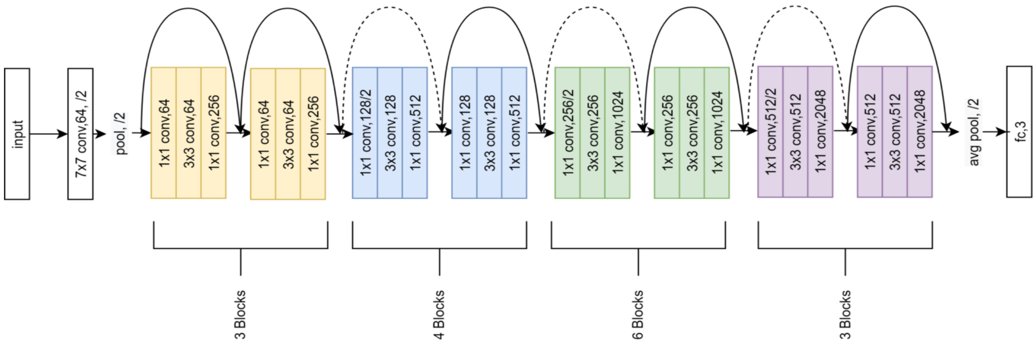

Figure 21: Architecture of ResNet-50[52]

Identity mappings to prevent the vanishing gradient issue are one of the key benefits of ResNet-50 because gradients can propagate between layers directly. This allows deep networks to be trained without compromising their performance. Moreover, the architecture encourages feature reuse, minimizes redundancy in learned representations, and enhances generalization performance. Experimental evidence indicates that residual networks are more accurate than traditional deep CNNs, despite having much more depth.

ResNet-50 is widely used in transfer learning, in which a pre-trained model on large-scale data sets, such as ImageNet is fine-tuned on specific domain-specific tasks. The learned feature extraction function can be represented as:

$$y = f_\theta(x)$$

where f_θ is the ResNet-50 model with parameters θ, and x is the input image. Fine-tuning is used to adjust these parameters to new data to achieve high-performance with limited training data. Because of its strengths, scalability and efficiency, ResNet-50 has emerged as a building block in today's computer vision applications, such as image classification, object detection, and other applications that use transfer learning.

5.2.4 DenseNet-121

DenseNet-121 is a popular deep convolutional neural network architecture that is a member of the family of Densely Connected Convolutional Networks (DenseNets), which was originally designed to overcome the drawbacks of very deep neural networks, including vanishing gradients, inefficient reuse of features, and massive parameter proliferation. In contrast to a traditional convolutional neural network, where each layer in the network only takes input from the immediate predecessor, DenseNet proposes a dense connectivity structure, where every layer is directly connected to all preceding layers in a feed-forward fashion. This structure is used to guarantee the optimal information transfer among layers and also propagate features across the network much higher[48].

Formally, let $x_0, x_1, x_2, \ldots, x_{l-1}$ denote the outputs of layers 0 to $l-1$. DenseNet In DenseNet the l^{th} layer output is a composite function of all the preceding feature maps:

$$x_l = H_l([x_0, x_1, x_2, \ldots, x_{l-1}])$$

in which $[\cdot]$ is the concatenation operation and $H_l([\cdot])$ a nonlinear transformation, which comprises of the following operations: Batch Normalization (BN), Rectified Linear Unit (ReLU), and convolution. This definition is fundamentally different to residual learning in architectures such as ResNet, where feature maps are added together by summation. Rather, DenseNet fuses the features, retaining the information of all the previous layers and allowing explicit reuse of features.

An important idea of DenseNet is the growth rate (denoted by k) that defines the number of new feature maps added per layer to the global state. With input k_0 feature maps to a dense block, the number of feature maps at a single layer is l, so that the number of feature maps after l layers is:

$$k_0 + k \cdot l$$

This linear expansion is in contrast to the traditional architectures where features dimensions may grow exponentially, resulting in more expensive computations. DenseNet can be parameter efficient with relatively low growth rate, and enable deep representations.

DenseNet-121 in particular is composed of 121 layers with (convolutional, pooling and fully connected) layers. It is structured into several dense blocks by transition layers. Within each dense block, there are multiple connected convolutional layers that are densely packed followed by transition layers that downsample feature maps with 1×1 convolution, 2×2 average pooling. The transition layer is mathematically given as:

$$x_{l+1} = \text{AvgPool}(W * x_l)$$

where W is the convolutional filter weights and $*$ is convolution. The transition layers are used to regulate the complexity of the models and cut down the spatial dimensions so as to be able to compute them effectively.

One more significant feature of DenseNet-121 is the introduction of the bottleneck layers to enhance the computational efficiency. A bottleneck layer uses a 1×1 convolution followed by the standard 3×3 convolution, which down-sampler the number of feature maps in the input, thereby decreasing the computational expense. This change of a bottleneck layer may be expressed as:

$$H_l(x) = \sigma(W_2 * (\sigma(W_1 * x)))$$

In which W_1 and W_2 are the 1×1 and 3×3 convolution kernels respectively, and σ is the activation function (usually ReLU).

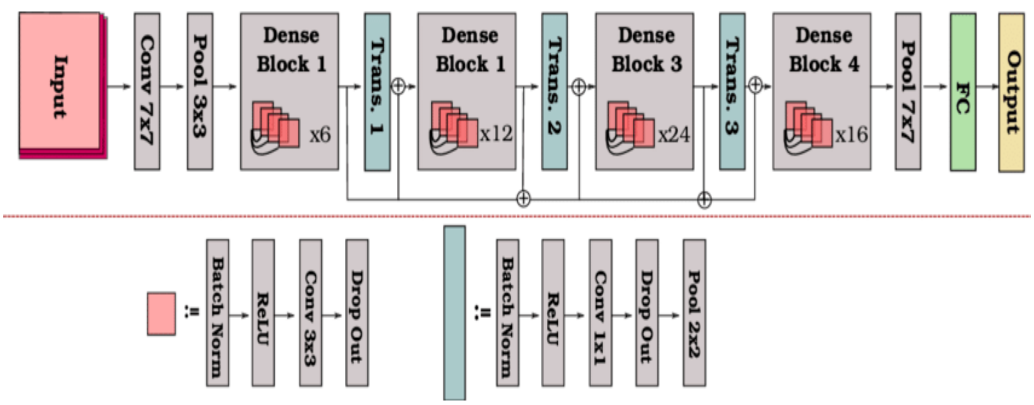

Figure 22: Architecture of DenseNet-121 [49]

The high-density connectivity scheme has a number of benefits. First, it mitigates the vanishing gradient issue by establishing direct connections between early and later layers so that the gradients can be more directly propagated during backpropagation.

Second, it promotes feature reuse, whereby all layers can access all features previously learned, eliminating redundancy and enhancing learning performance. Third, it has much fewer parameters than the conventional deep networks since redundant feature learning is avoided.

DenseNet-121 has shown to be a good performer in the context of transfer learning and image classification task because it can learn rich hierarchical representations and at the same time be computationally efficient. The model is especially appropriate in complex visual recognition, like classification of fruits, medical image analysis, and recognition of environmental scenes, due to its dense connectivity mechanism that allows it to simultaneously capture both low-level and high-level features. Moreover, with transfer learning, the DenseNet-121 architecture can use pre-trained weights of large-scale datasets such as ImageNet, greatly boosting its performance on smaller, domain-specific datasets and saving training time .

All in all, DenseNet-121 is an important contribution to the design of deep learning architectures through dense connections, which not only increase information flow, gradient propagation, and allow effective feature reuse. Such properties enable it to be a powerful and widely used model in contemporary computer vision applications.

5.3 Proposed Custom CNN Architecture

The designed Custom Convolutional Neural Network (CNN) structure is defined as a deep and hierarchical feature learning model that is purposefully designed to be used in the context of multi-class classification of rare and indigenous fruits. Unlike traditional models that are typically pre-trained on large and generic image datasets, the architecture is specifically designed to tackle the specific problems of medium vision datasets with high intra-class similarity and low inter-class differences. This design aspect makes sure that this model does not only capture fine-grained visual differences between the categories of fruits but also remains to be computationally efficient, which is paramount to practical implementation and accelerated convergence in training.

Table 5.1: Custom CNN Architecture Layer Configuration

Block	Layers	Configuration	Output
B1	Conv → Pool	32 (3×3), ReLU, same; Pool (2×2)	(112,112,32)
B2	Conv → Pool	64 (3×3), ReLU; Pool (2×2)	(55,55,64)
B3	Conv → Conv → Pool	64 (3×3), ReLU; Pool (2×2)	(26,26,64)
B4	Conv → Pool	128 (3×3), ReLU, same; Pool (2×2)	(13,13,128)
B5	Conv → Pool	256 (3×3), ReLU; Pool (2×2)	(6,6,256)
B6	Conv → Pool	512 (3×3), ReLU, same; Pool (2×2)	(3,3,512)
B7	Conv → Pool → Dropout	512 (3×3), ReLU, same; Pool (2×2); Dropout=0.25	(1,1,512)

The design follows seven successive convolutional blocks of features extraction and a fully connected classification head. This hierarchical structure allows the network to increasingly convert raw pixel-level data into higher-level and more abstract and semantically rich representations. During the early stages, the network is concerned with the low-level features of the images in terms of edges, color gradients, and texture patterns. These primitive features are hierarchically fused into midlevel representations, such as shape contours and structural features, with the depth of the network. Finally, the higher levels store the high-level semantic features that are highly correlated with particular classes of fruits, thus increasing the discriminative ability of the model.

In order to be compatible with the current deep learning methods and allow the transfer learning paradigm, the network takes $224 \times 224 \times 3$-dimensional (spatial resolution and color channels, respectively) input pictures. This standardized input representation enables consistent feature extraction, minimizes the variability in computations, and enables easy comparison with well established architectures like ResNet and MobileNet. In addition, the input images can be resized to this resolution because this provides a balance between retaining enough visual detail and having computational complexity that is manageable.

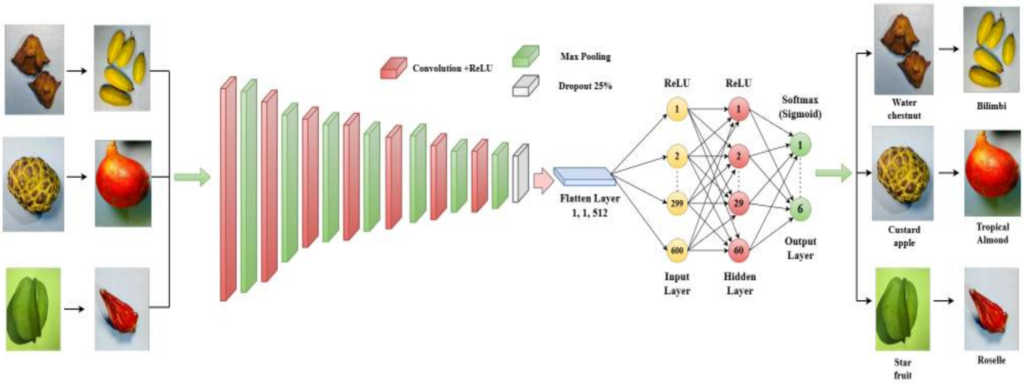

Figure 23: Architecture of Custom CNN

In general, this architectural design is an intentional trade-off between depth, the richness of features and efficiency. The model incorporates several convolutional phases with a structured classification head that allows it to perform a strong hierarchical feature learning, and it is lightweight, allowing it to effectively train on smaller datasets. This renders the proposed CNN especially applicable to specific classification problems like the recognition of rare fruits, where the accuracy and efficiency are of the utmost importance.

5.4 Model Training Strategy

The model training approach in this study is extensively developed to achieve a strong feature learning, good convergence, and high feature generalization of multi-classification of rare fruits. The training is performed in a supervised deep learning model, in which the model is trained to learn a parametric mapping $f_\theta : X \to Y$ between the input image space and the class labels, via a deep convolutional neural network. With the training data $\mathcal{D}_{\text{train}} = \{(x_i, y_i)\}_{i=1}^{N}$, the learning aim is to decrease the sparse categorical cross-entropy loss, which is defined as

$$\mathcal{L}(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \log(\hat{y}_{i,y_i})$$

where $\hat{y}_{i,y_i}$ is the estimated likelihood of the actual class that was achieved using the Softmax activation function.

$$\hat{y}_{i,c} = \frac{e^{z_c}}{\sum_{j=1}^{C} e^{z_j}}$$

where $C = 6$ represents the categories of fruit. Model parameters optimization is done through Adam optimizer with learning rate of 1×10^{-4} that updates weights as first and second moment estimates of gradients; hence, allowing stable and efficient convergence even in deep architecture. The update rule of the parameters is given as

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon}$$

m_t, v_t are the moment estimates and η is the learning rate.

The training is performed with a mini-batch gradient descent with a batch size of 32 and a maximum number of 80 epochs, which enables the effective use of the computer resources and the stability of the gradient. In every step, the forward propagation uses the results of the previous step to make predictions with the help of a series of convolutional and fully connected layers, and backpropagation is used to adjust model parameters based on the calculated loss gradients $\nabla_\theta \mathcal{L}$. In order to reduce overfitting and improve generalization, dropout regularization at a rate of 0.25 is used, both after the convolutional feature extraction step and the fully connected layers, which helps avoid co-adaptation of neurons[21]. Additionally, there is an early stopping checkpoint to keep track of validation loss, where the training is stopped when the loss shows no improvement in 10 consecutive epochs (patience = 10) and model weights with the highest performance are automatically reloaded[20]. This helps to avoid unnecessary training past optimal convergence and helps to avoid overfitting. Training is done on the validation dataset in order to steer the process in a manner that the model will not change significantly when it is presented with unseen data. In general, adaptive optimization, controlled learning rate, batch-wise training, dropout regularization and early stopping make the proposed CNN a complete training strategy that allows the model to learn discriminative and invariant feature representations, which eventually results in high classification accuracy and strong real-world behavior, as the findings reported in the proposed framework [4].

5.5 Hyperparameter Settings

The hyperparameter setup of the proposed Custom CNN model is carefully designed to achieve stable convergence, effective learning, and good generalization for multi-class classification of rare fruits. The hyperparameters used in this research are summarized in Table 5.2. These settings were carefully selected to ensure stable training, efficient convergence, and improved generalization performance of the proposed model.

Table 5.2: Hyperparameter Settings of the Proposed Model

Hyperparameter	**Value**
Input Size	$224 \times 224 \times 3$
Learning Rate	1×10^{-4}
Batch Size	32
Epochs	80
Optimizer	Adam
Loss Function	Sparse Categorical Crossentropy
Dropout Rate	0.25
Early Stopping Patience	10

Before training, each input image is scaled to the range $[0, 1]$ using min-max normalization:

$$X_{\text{norm}} = \frac{X}{255}$$

This normalization reduces numerical instability and accelerates gradient-based optimization.

The model is trained using the Adam optimizer, which combines momentum and adaptive learning rates. The parameter update rule is given by:

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

where $\alpha = 10^{-4}$ is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected first and second moment estimates, and ϵ is a small constant for numerical stability. A relatively low learning rate ensures smooth convergence and prevents divergence during training.

A batch size of 32 is used, and the model is trained for a maximum of 80 epochs to ensure sufficient exposure to the dataset while maintaining computational efficiency.

To optimize classification performance, the Sparse Categorical Cross-Entropy loss function is employed, which is suitable for multi-class problems with integer labels:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \log\left(p_{y_i}\right)$$

where p_{y_i} is the predicted probability of the true class.

The final layer uses the Softmax activation function to convert logits into probability distributions:

$$P(y = i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}$$

A dropout rate of 0.25 is applied after convolutional and fully connected layers to reduce overfitting by randomly deactivating neurons during training.

Additionally, an Early Stopping strategy is implemented based on validation loss, with a patience of 10 epochs and automatic restoration of the best-performing weights. This ensures that training stops when performance no longer improves, enhancing generalization capability.

These hyperparameter optimization strategies align with standard CNN optimization practices for image classification tasks, where a balanced configuration of learning rate, batch size, and regularization plays a crucial role in overall model performance[15].

Chapter 5 conclusion

Within this chapter, a detailed and systematic approach of classifying rare fruits has been provided, which combines both transfer learning processes and a new custom-built CNN architecture. The research model provides a systematic pipeline, starting with the choice of state-of-the-art pre-trained models, i.e., MobileNetV2, InceptionV3, ResNet-50 and DenseNet-121, and then the design of a task-specific custom CNN, to overcome the dataset-specific issues (high inter-class similarity and small sample size).

Moreover, a suitable model training plan was developed, which includes the data normalization, optimization algorithms, and regularization approaches to improve the generalization and avoid overfitting. Hyperparameters such as learning rate, batch size, and epochs, and early stopping criteria were carefully tuned to achieve a steady convergence and optimal performance. This type of systematic optimization is essential, as the current CNN-based models show that the trade between the accuracy and computational efficiency is necessary in order to be practically used in the real-life applications.

All in all, the suggested methodology offers a solid and scalable image classification scheme, which will serve as a solid basis of the experimental assessment and performance evaluation presented in the next chapter.

6 Experimental Setup

6.1 Hardware Software Environment

The experimental testing of the proposed framework of rare fruit classification was run in a controlled computational environment that was balanced to performance efficiency and hardware limitations. Since the dataset is moderate (1,800 original images and 6,300 augmented samples) the training process had to be highly sensitive to computational resources, especially the memory on the GPU (VRAM), which directly correlates with the model feasibility, the choice of batch size, and the stability of the training process.

The hardware set up to train and evaluate the model is summarized in Table 6.1.

Table 6.1: Hardware Specifications to be used to train the model.

Component	**Specification**
CPU	Intel Core i5-13400F
Memory (RAM)	32 GB DDR4 (3200 MHz)
GPU	NVIDIA GeForce RTX 4060
GPU Memory (VRAM)	8 GB GDDR6
CUDA Cores	3072

A GPU-enabled environment greatly speeds up the process of deep learning, especially convolutional calculations and gradient-based optimisation when running backpropagation. The small VRAM (8 GB) is however a limitation to model configuration. Deep architecture models (ResNet-50 and DenseNet-121) are highly sensitive to batch size and input dimension choices in order to prevent out-of-memory (OOM) errors. As a result, a batch size of 32 was chosen as a reasonable compromise between the computational and memory usage.

Unlike methods used to scale input dimensionality to computational convenience, this work uses higher-resolution RGB inputs (224× 224 x3 on most models, 299x299x3 on InceptionV3) to keep important visual features like color, texture, and shape, which are important to know when classifying rare fruits finely. Although this option has advantages in representing features better and improving the performance of a model, it also reduces training speed and computational complexity.

Software wise, all the experiments were conducted in Python (version 3.8) as the main programming environment. Transfer learning and custom CNN architectures were built, trained and evaluated using deep learning models constructed with the high-level Keras API built on top of TensorFlow. Numerical operations and data manipulation were supported with the help of libraries like NumPy, image preprocessing with the help of OpenCV, and visualization with the help of Matplotlib.

Jupyter Notebook embedded in Visual Studio Code was used to configure the development environment to enable modular experimentation and control findings, reproducibility, and the refinement of model development. This design allowed to evaluate several architectures, such as MobileNetV2, InceptionV3, ResNet-50, DenseNet-121, and the suggested Custom CNN architecture, on equal footing in the same experimental setting.

On the whole, the experimental environment is representative of a realistic and resource-constrained environment that is relevant to real-life deployment in developing areas. These limitations were reflected in major design choices, such as the choice of the model, resolution of input, and training methods, which make sure that the proposed system is efficient and can be used in real-life agricultural practice with regards to recognizing rare fruits.

6.2 Training Protocol

The training process of the proposed system of classifying rare fruits was done via a supervised deep learning model, in line with the methodology of the main study . The data was split into training (70%), validation (15%), and testing (15%) sets to be able to guarantee a good evaluation of the performance and generalization.

Before training a model, all input images were rescaled in min-max scale to bring pixel values in the interval ([0,1]) which enhances numerical stability and fast convergence when optimizing:

$$X_{norm} = \frac{X}{255}$$

The proposed Custom CNN architecture has been trained in the Tensorflow-Keras framework using a blank architecture. The network comprises of several convolutional and pooling layers to extract features in the hierarchy and classification via fully connected layers. The last layer employs a SoftMax activation function to give class probabilities in six fruit categories.

The training process used the sparse categorical cross-entropy loss, an appropriate loss when the labels are integer-encoded multi-class:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

The Adam optimizer was used with a comparatively low learning rate of $(1 \times 10^{-4}) toensuretheoptimi$

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

where $(\alpha = 10^{-4}) isthelearningrate.$

The following hyperparameter configuration was used to train the model:

- Batch size: 32 (maximum 8 GB VRAM)
- Maximum epochs: 80
- Learning rate: 0.0001
- Loss function: Sparse categorical cross-entropy.
- Optimizer: Adam
- Dropout rate: 0.25 (used after convolutional and dense layers)

Early stopping strategy was adopted to avoid overfitting and to achieve the best generalization. Validation loss was monitored, and terminated during 10 consecutive epochs in the training process not improved. Also, optimal model weights were automatically restored:

$$\text{Training halts when 10 epochs: } \Delta\mathcal{L}_{val} \approx 0$$

This method minimizes unneeded calculation and alleviates overfitting, especially in the small dataset case. The batch size and model complexity were limited due to hardware limitations (8 GB GPU VRAM) to prevent the overflow of memory. Bigger batch sizes or richer architectures would consume much more memory and require more training time. Consequently, the chosen setup will be a trade off between model performance and computational feasibility, just like the restrictions found in the PoribohonNet study. In addition, the model was trained with a maximum of 80 epochs, but early stopping usually interrupted the training process prematurely when convergence occurred, which enhanced training efficiency, but did not affect high accuracy.

In general, the training protocol indicates a resource-conscious optimization scheme, which guarantees a steady convergence, low overfitting, and optimal use of the computational resources available in addition to high classification accuracy.

6.3 Complexity Analysis

Computational complexity of the proposed rare fruit classification framework is a crucial consideration in assessing the practicality, efficiency, and applicability of the framework in a real-world scenario. In this section, time and space complexity of the implemented deep learning models, the proposed Custom CNN and the transfer learning architectures are analyzed.

6.3.1 Time Complexity

Convolutional operations in Convolutional Neural Networks (CNNs) are the main factors that control time complexity because during forward and backward propagation, they control the computational cost. In the case of a convolutional layer, the computation cost is estimated as:

$$O(HWC_{in}K^2C_{out})$$

In which (H) and (W) are the spatial dimensions of the input feature map, (C_{in}) and (C_{out}) is the number of input and output channels, respectively, and (K) is the kernel size.

The model in the proposed Custom CNN architecture comprises of seven convolutional blocks that have filter sizes that increase progressively (32 to 512). Although deeper layers can increase the cost of computation, max-pooling layers decrease the dimensions of space, and thus, the complexity is regulated. The Custom CNN has a more efficient computation structure, in comparison to more complex transfer learning models, like ResNet-50 and DenseNet-121, which include much more layers and parameters.

Transfer learning models create a burden on computation because of their depth and architecture. For instance:

ResNet-50 has residual connections, which require more time to train and it has about 50 layers.

DenseNet-121 is based on dense connectivity, which results in repeated use of features

and increased computation.
InceptionV3 has parallel convolutional branches, which multiply the multi-scale computation.
MobileNetV2 is not very efficient, but it is relatively efficient because of depthwise separable convolutions, which minimize the computational cost.

In general, the proposed Custom CNN has a smaller training and inference time, compared to most transfer learning models, and is thus computationally efficient, with a high accuracy.

6.3.2 Space Complexity

The complexity of space is calculated as the number of parameters and memory to store intermediate feature maps during training. It can be expressed as:

$$O(P + A)$$

(P) is the number of trainable parameters and (A) is the memory needed to store activations (feature maps).

The proposed Custom CNN is a light and less parameterized network than the deep pre-trained networks. The chain of successively larger filters is counterbalanced by strong spatial downsampling with pooling layers, that downsizes the size of intermediate feature maps and memory usage.

In contrast:

DenseNet-121 is costly in memory because of the dense concatenation of features.
ResNet-50 demands a lot of memory to hold residual connections.
InceptionV3 consumes more memory because of parallel feature maps.
MobileNetV2 is less memory-consuming since it has a bottleneck format.

Since the hardware has a 8 GB VRAM limit, there was a need to optimize the batch size and input resolution to achieve successful training without memory overflow. The chosen batch size is 32 which is a trade-off between the memory and computational throughput.

6.3.3 Model Efficiency and Implications

The experimental results suggest that the proposed Custom CNN has better classification performance (97.40% accuracy) and it is at the same time lower in computational and memory complexity than deeper transfer learning models . This illustrates how domain-specific architecture design can perform well on fine-grained classification tasks, especially when using relatively small datasets.

Practically, the small-scale of the Custom CNN makes it more applicable to:

- Operation in systems with resource constraints.
- Real-time inference applications
- Edge or portable farm implements.

Alternatively, deeper models can be more powerful, but might demand more computational resources, more time to train, and more energy, so their use in the real-world rural or low-resource setting is restricted.

In summary, the proposed framework achieves an effective balance between accuracy and computational efficiency. The Custom CNN offers:

- Lower time complexity due to reduced architectural depth
- Lower space complexity through efficient feature map management
- High classification performance with minimal computational overhead

These characteristics make the proposed model a practical and scalable solution for rare fruit classification in real-world agricultural applications.

6.4 Training and Analysis of Validation Performance of Custom CNN

The training and validation accuracy and loss curves of the proposed Custom Convolutional Neural Network (CNN) were further examined by visualizing the series of curves within various epochs, as shown in Fig. 24 and Fig. 25, which provide important information on the dynamics of the learning process, convergence, and the ability of the model to generalize.

The training and validation accuracy curves (Fig. 24) show that the model is able to quickly acquire accuracy in the first epochs, which means that at the early stages of training, the model is successfully acquiring low-level and mid-level visual features. The training accuracy increases drastically to more than 90 percent after the initial five epochs with the validation accuracy also increasing in the same fashion to more than 80 percent during the initial five epochs. This tendency proves effective feature extraction and rapid convergence of the identified architecture.

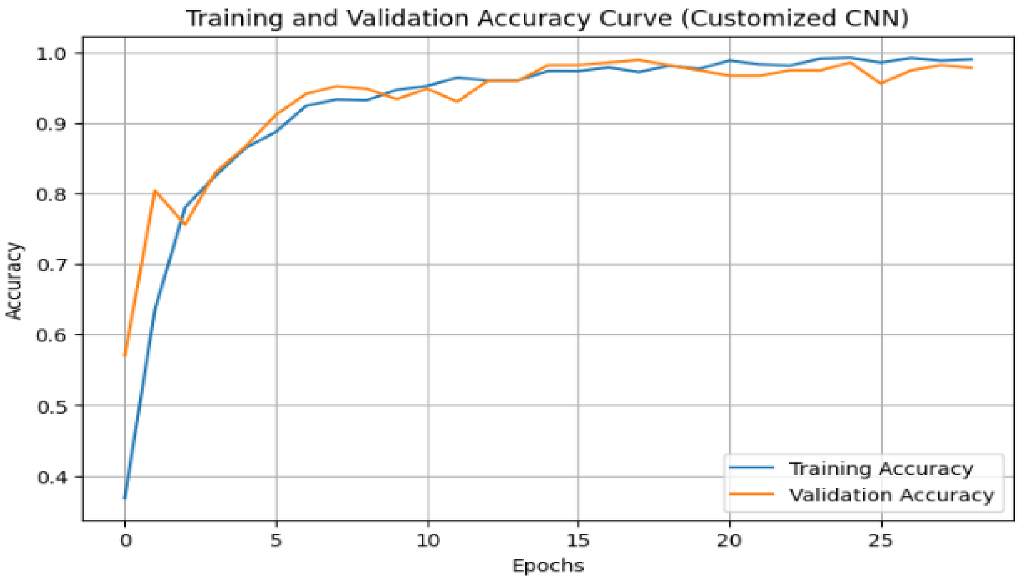

Figure 24: Training and Validation Accuracy Curve

As the training continues, the two curves eventually level and the curves approach near optimal performance where the values of accuracy are near 98 to 99. The fact that the gap between training and validation accuracy remains minimal in the later epochs indicate that the model is not overfitted considerably. Rather, it has good generalization of undetected data. Variations in validation accuracy can occur in minor fluctuations which are natural, as the datasets are not constant and stochastic optimization may lead to fluctuations, yet they do not imply that things are unstable.

This is also supported by the loss curves (Fig. 25). The training loss steadily declines in value (starting with a large value (1.38)) to close to zero, indicating continuous improvement in model predictions. Likewise, the loss of validation decreases steeply in the initial epochs, and then there are minimal fluctuations around a small value. Such fluctuations are common in deep learning models and are a sign that the model is sensitive to validation samples, as opposed to model degradation.

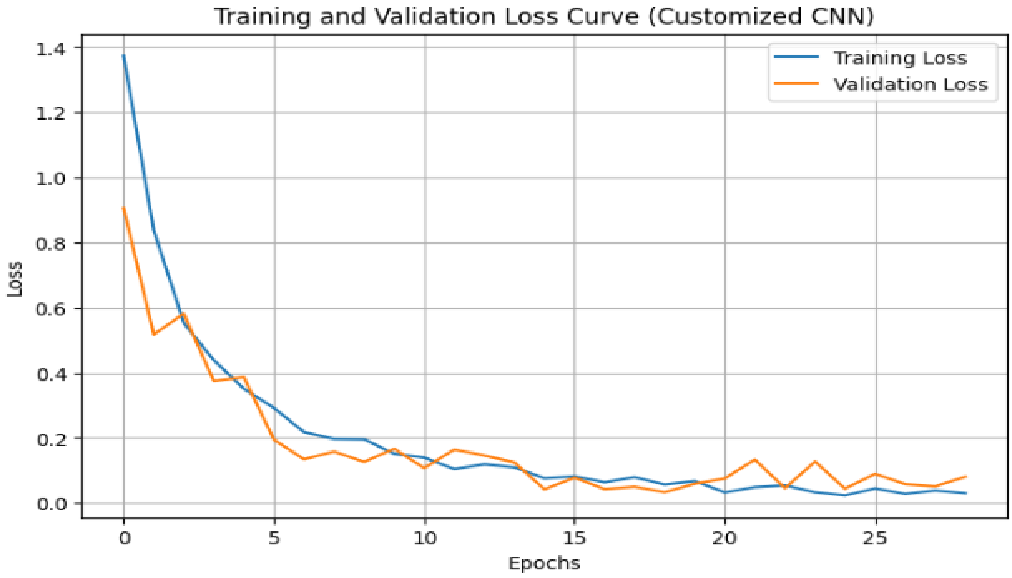

Figure 25: Training and Validation Loss Curve

Notably, there is no substantial difference between training and validation loss which proves that the model has a good balance between bias and variance. Lack of growing loss of validation at subsequent epochs suggests that overfitting is adequately suppressed, probably because of the dropout layers and data augmentation approaches mentioned above in the approach.

On the whole, the convergence trends of the accuracy and loss curves confirm the strength and stability of the suggested Custom CNN model. These findings indicate that the network is effective in learning discriminative features and retaining high generalization performance that directly results to its high classification accuracy of 97.40% in this study.

7 Results and Performance Analysis

7.1 Model Comparison

This paper includes a thorough comparative analysis of five deep learning models, such as ResNet-50, InceptionV3, MobileNetV2, DenseNet-121, and a Custom CNN, in the classification of extinct rare fruit images. The comparison will be based on the quantitative findings presented in Table IV that involves the main performance measures, including: precision, recall, accuracy, F1-score, and error measures, including False Negative Rate (FNR) and False Positive Rate (FPR).

The results of the experiment have made it evident that the suggested Custom CNN model performs better than any other architecture in terms of almost all the evaluation metrics. In particular, it has the best accuracy (97.40), as well as better precision (0.9755), recall (0.9740) and F1-score (0.9739). In addition, the Custom CNN has the least FNR (0.0259) and FPR (0.0052), which means that it has a high ability to produce very few false alarms and missed detections. This high performance indicates that the model is highly adapted to the particularities of the dataset, allowing it to extract the features and distinguish between rare fruit categories successfully.

The second-best transfer learning model in terms of accuracy is ResNet-50, which has a 96.67 percent accuracy. The residual learning mechanism, in which the identity shortcut connections allow the efficient gradient flow and allow deeper network training without degradation problems could explain its high performance. Its strength and ability to generalize is further validated by the relatively low FNR (0.0333) and FPR (0.0067).

InceptionV3 also shows a competitive performance, with an accuracy of 96.29, balanced values of precision, recall and F1-score (all approximately 0.96). Its design takes advantage of factorized convolution and dimensionality reduction techniques to make computations more efficient, without sacrificing representational capacity. Its relatively greater error rates than ResNet-50 suggest, however, that it has slight limitations in its ability to separate visually similar classes of fruits.

Conversely, DenseNet-121, which has a more theoretically advantageous design like dense connectivity and enhanced feature reuse, has a relatively lower accuracy of 95.55%. Although DenseNet models are reported to enhance feature propagation and suppress the vanishing gradient problem through connecting all the layers with all the previous layers, the performance in this experiment is lower than it could be due to the distribution of the dataset or the complexity of the features in rare fruit images.

Lastly, MobileNetV2 has the lowest performance compared to all of the reviewed models with the accuracy of 93.70. Although it maintains reasonable precision (0.9454) and F1-score (0.9371), it exhibits the highest FNR (0.0630) and FPR (0.0126). This is owed to its lightweight design where it makes use of inverted residual and depth separable convolutions to minimize computation overhead. Although efficient, this design has a built-in capacity of limiting its feature extraction ability in highly complex classification.

In general, the comparative analysis suggests that the trade-off between the archi-

Table 7.1: Evaluation Metrics of Extinct Rare Fruit Image Classification

Model	Precision	Recall	Accuracy	F1 Score	FNR	FPR
MobileNetV2	0.9454	0.9370	0.9370	0.9371	0.0630	0.0126
ResNet-50	0.9684	0.9666	0.9667	0.9640	0.0333	0.0067
Custom CNN	0.9755	0.9740	0.9740	0.9739	0.0259	0.0052
DenseNet-121	0.9573	0.9556	0.9555	0.9545	0.0444	0.0089
InceptionV3	0.9668	0.9629	0.9629	0.9629	0.0370	0.0074

tectural complexity and the ability of the model to represent the features has a strong impact on its performance. Although the transfer learning models like ResNet-50 and InceptionV3 have been used, with good baseline performance, the proposed Custom CNN gives a better adaptability and task-specific optimization and would therefore be the most successful model of the extinct rare fruit classification in this study.

7.2 Confusion Matrix Analysis

A confusion matrix is an important analysis tool of class-wise classification model performance evaluation. It gives a comprehensive portrayal of the results of prediction by comparing actual class labels to the predicted labels. The confusion matrix of a (K) class multi-class classification problem is given as a (K by K) matrix:

$$M = [M_{ij}], \quad i, j = 1, 2, \ldots, K$$

In which (M_{ij}) is the number of samples in class (i) which are predicted as class (j). The diagonal values $((M_{ii}))$ denote accurate classifications (True Positives) whereas the off-diagonal values reflect inaccurate classifications.

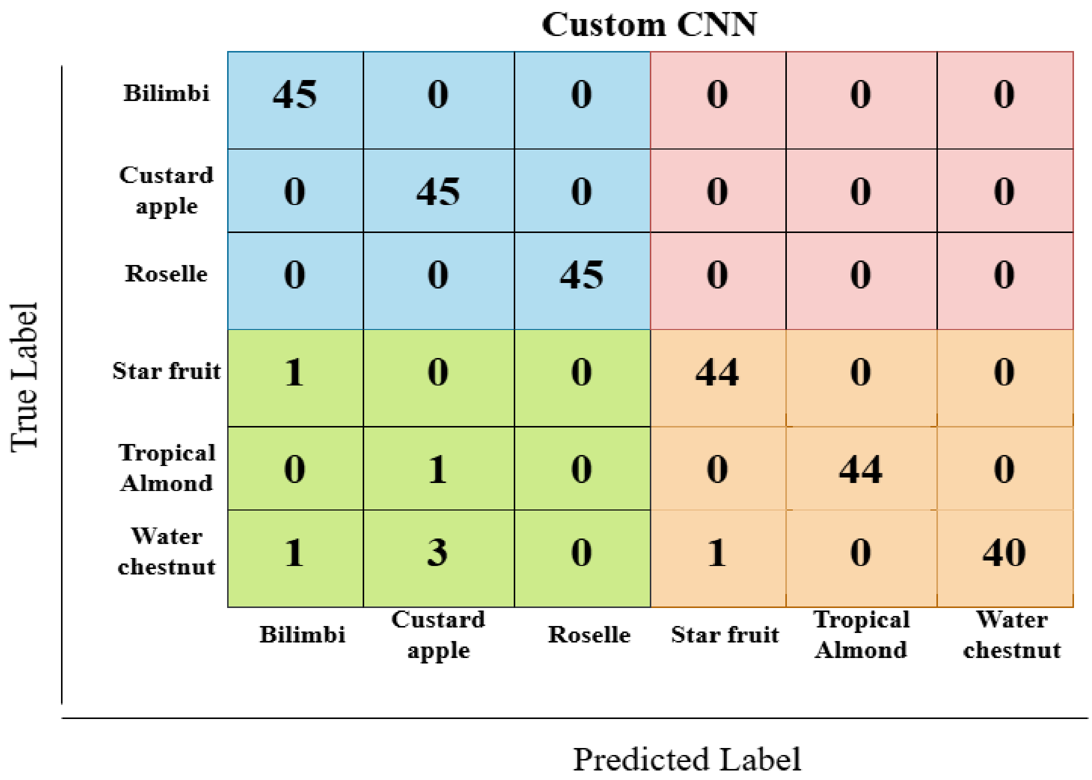

Figure 26: Confusion Matrix of Proposed Custom CNN

The metrics of evaluation based on the confusion matrix are as follows:

7.2.1 Confusion Matrix Evaluation Metrics

Based on the confusion matrix, the important evaluation metrics are calculated as below:

Accuracy:

$$Accuracy = \frac{\sum_{i=1}^{K} M_{ii}}{\sum_{i=1}^{K}\sum_{j=1}^{K} M_{ij}}$$

Precision (per class):

$$Precision_i = \frac{M_{ii}}{\sum_{j=1}^{K} M_{ji}}$$

Recall (per class):

$$Recall_i = \frac{M_{ii}}{\sum_{j=1}^{K} M_{ij}}$$

F1-Score:

$$F1_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i}$$

These statistics enable a whole picture interpretation of model behavior, especially in determining strengths and weaknesses in classes [40].

7.2.2 Confusion Matrix Analysis Modelwise

ResNet-50

The confusion matrix of ResNet-50 shows that its diagonal is highly dominant, which means that its classification accuracy is high on most classes. Classifications are perfect (45/45) with minor misclassifications in classes 3, 5 and 6. As an illustration, class 6 has 4 cases that are misclassified as class 1, and 1 case wrongly classified as class 3. On the whole, ResNet-50 attains about:

$$Accuracy \approx \frac{261}{270} = 96.67\%$$

This means that residual learning is a good model to learn discriminative features, though there is still small confusion between classes that are visually similar.

MobileNetV2

MobileNetV2 has a little worse performance in comparison with ResNet-50. Although there are a few classes with perfect classification, there is a lot of confusion in the class 6 with 11 samples being misclassified as class 5. Also, class 3 is confused with other classes 4 and 5.

$$Accuracy \approx \frac{253}{270} = 93.70\%$$

This reduced performance is due to the trade-off between computational efficiency and feature representation capacity in lightweight architectures.

InceptionV3

The InceptionV3 model has a good balance of performance and high classification in most classes. There is however, a significant confusion in class 1, with 5 samples wrongly classified as class 6 and there are minor errors in classes 3 and 5.

$$Accuracy \approx \frac{260}{270} = 96.30\%$$

InceptionV3 has parallel convolutional architecture making it effective in multi-scale feature extraction, yet there is still a small amount of misclassification due to overlapping visual features.

DenseNet-121

DenseNet-121 has a high feature propagation ability, as it classifies the first four classes perfectly. Class 5 and 6 are however misclassified. As an example, there is confusion of class 5 with classes, 1, 4, and 6 and 4 samples in class 6 were misclassified as 3.

$$Accuracy \approx \frac{258}{270} = 95.56\%$$

Despite its superiority in the majority of the categories, the decrease in performance in certain classes indicates that DenseNet is sensitive to intra-class variation.

Custom CNN

The CNN model with custom is competitive and performs well with perfect classification in a number of classes. Misclassifications are however more distributed than other models. Class 6 is the most confused, with the errors being distributed among the classes.

$$Accuracy \approx \frac{263}{270} = 97.41\%$$

Interestingly, although the custom CNN has a simpler architecture, it has the highest overall accuracy, which means that it learns effectively on this particular dataset.

7.2.3 Comparative Analysis

Table 7.2 summarizes a comparative analysis of all the models.

Table 7.2: Models Comparative Performance Analysis of Confusion Matrix.

Model	**Performance Category**	**Key Observation**
Custom CNN	Best Overall Accuracy	The highest accuracy (97.41) was achieved with good classification of all classes.
ResNet-50	Best Stable Performance	Minimal scattered errors and predictable across classes.
DenseNet-121	Optimal Feature Usage	Optimal classification in most of the classes because of successful feature reusing.
InceptionV3	Balanced Performance	Stable and balanced classification with moderate misclassifications.
MobileNetV2	Worst Performance	More misclassification because of the limitations of a lightweight architecture.

In all the models, the misclassifications occur mostly in certain classes (especially class 5 and class 6). These classes are likely to be similar in appearance in terms of colour, texture or shape and so feature representations overlap and the classification becomes harder.

7.2.4 Observations and Discussion

The analysis of the confusion matrix can give several crucial observations:

1. Diagonal Dominance: Strong diagonal values are observed in all models, which proves effective learning and good classification accuracy.

2. Class-specific Confusion: Misclassifications are not occasional but are between two distinct classes suggesting that there are some difficulties in the similarity of features.

3. Model Complexity vs Performance: Although more advanced models (ResNet-50, DenseNet-121) are more stable in performance, the custom CNN still has better results because of the improved dataset-specific optimization.

4. Lightweight Models can impact the following: The lower representative power of MobileNetV2 leads to increased confusion especially in complex classes.

5. Generalization Challenge: Classes that are more intra-class varient or less distinct have more errors and this indicates that better feature extraction and data balancing are required.

The confusion matrix analysis will give a more detailed insight into the performance of classification on the basis of more than aggregate measures. All models are highly accurate, but the custom CNN is slightly more accurate, then ResNet-50 and InceptionV3. DenseNet-121 has high feature learning capabilities, but in particular classes, it performs poorly, and MobileNetV2 has weaknesses because of its lightweight architecture.

All in all, the findings suggest that the model architecture, data characteristics, and feature similarity jointly affect the classification in rare fruit recognition tasks.

7.3 Precision–Recall (PR-AUC) vs ROC-AUC Comparison

Although the Receiver Operating Characteristic (ROC) curve gives a complete picture of the discriminative capability of a model at all classification levels, it is not always a complete picture of the performance in situations where there is a class imbalance or where the accuracy of correctly identifying positive samples is more significant. In these situations, a more informative assessment involves the Precision Recall (PR) curve where specific emphasis is made on the relationship between the recall and the precision. Precision is a ratio of accuracy of positive prediction correctness to all predictions of positivity, and recall is the capacity of the model to detect all real positive predictions. These measures can be defined as:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision versus recall Area Under the PrecisionRecall Curve (PR-AUC) or Average Precision (AP) is a summary statistic that captures the trade-off between recall and precision at various thresholds. In contrast to the ROC-AUC, which takes the net effect of true positives and true negatives, PR-AUC is especially sensitive to false positives and hence gives a more realistic evaluation in classification tasks where the right identification of positive samples are of utmost importance.

The PR curves of all the models tested as shown in Fig. 7.4 are very high over a large spectrum of recall values, which means that the models have good predictive power.

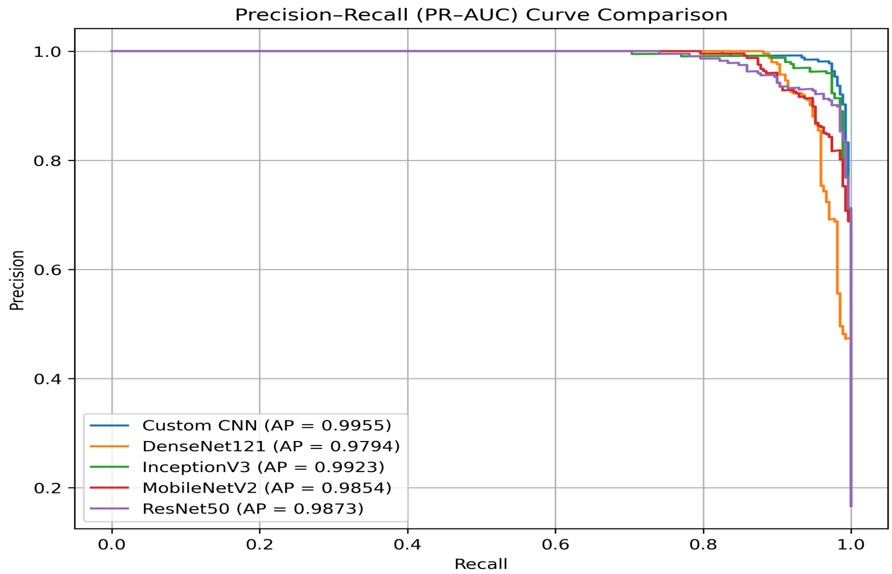

Figure 27: Precision–Recall (PR-AUC) Curve

The differences between the models are, however, more evident when the two curves are compared to the ROC curves. In particular, although ROC-AUC values of all models are above 0.99 (indicating almost perfect separability), the PR-AUC values are more varied, including variations in precision at high-recall rates.

Custom CNN model resulted in the highest score of PR-AUC, i.e. 0.9955, indicating that the model can retain high-precision even when recall is nearing its peak. This implies that the model generates minimal false positives and manages to detect most of the true ones. InceptionV3 follows closely with a PR-AUC of 0.9923 which has a similar trade-off between precision and recall. The good performance of these models indicates effective feature extraction and good generalization capacity.

The PR-AUC scores of ResNet-50 and MobileNetV2 were 0.9873 and 0.9854 respectively. Their curves are somewhat less than the best performing models, but nevertheless, it shows consistent and predictable performance over different thresholds. The slight decrease in PR-AUC relative to their ROC-AUC values indicates that with high-recall conditions, these models can add a few false positives. This is in line with their architectural design, as ResNet-50 focuses on deep feature learning and MobileNetV2 focuses on computational efficiency.

DenseNet-121 achieved the lowest PR-AUC of 0.9794, although it has a high ROC-AUC. This difference shows a significant difference between the two evaluation measures. Although DenseNet-121 works well in class separation across the board, as its ROC results indicate, it slightly reduces accuracy at the higher recall rates, indicating that it is more susceptible to confounding specific samples as positive. This can be explained by the thick connectivity mechanism, which, despite its advantageous feature reuse, and gradient propagation, may occasionally cause overlapping feature representations among visually similar classes.

An important finding of this comparative analysis is that ROC-AUC aims to portray a positive picture of model performance, especially where negative examples are the majority or when the true negatives are readily detected. On the other hand, PR-AUC is a more stringent criterion, directly rewarding the avoidance of false positives and focusing on the accuracy of positive predictions.

Thus, models which behave similarly on the ROC scale can be markedly different when compared from the PR perspective.

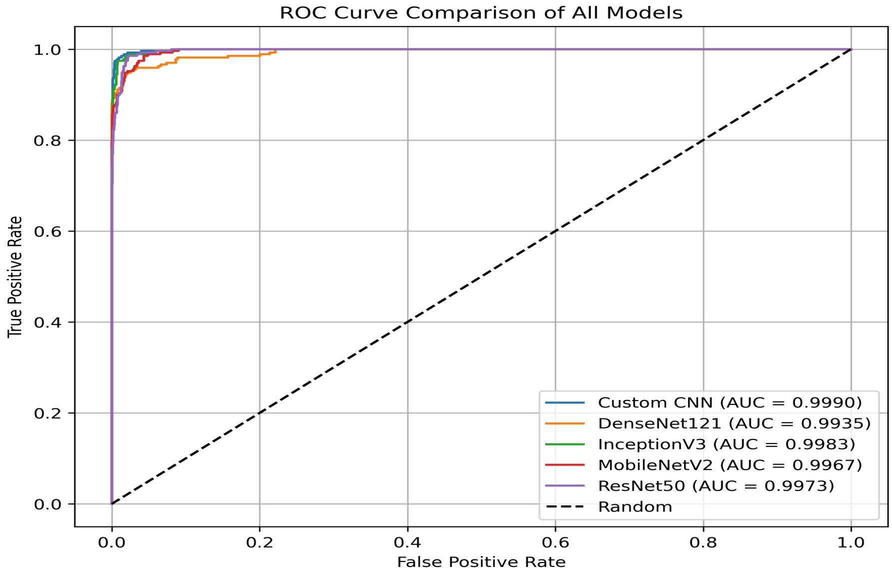

Figure 28: Receiver Operating Characteristic (ROC) Curve

Concluding, by analyzing ROC-AUC and PR-AUC together, it can be stated that Custom CNN and InceptionV3 represent the best-balanced models in this comparison. The small deviations in PR-AUC values for various models allow drawing better conclusions regarding their performance in practical conditions where precision plays as an important role as recall. Therefore, a combination of two metrics provides a more comprehensive approach to the problem of rare fruits' classification.

7.4 Error Analysis

Error analysis is an important element of assessing the strength and practicality of deep learning models. A close inspection of the misclassified samples in all the architectures; ResNet-50, InceptionV3, MobileNetV2, DenseNet-121 and the custom CNN indicated that there were a number of internal factors that led to classification errors[43].

Visual similarity of fruit classes was one of the most significant sources of misclassification. There are many rare fruits in Bangladesh that are very similar in appearance with color distribution, surface texture and shape. As an example, the confusion between classes was frequent with fruits that were greenish-yellow in colour or had a shape that was long. This difficulty points to an aspect of feature separability, in which even deep convolutional networks find it difficult to capture highly discriminative representations when there is low inter-class[54].

A second important performance factor was variation in light and the inconsistency in the lighting. Images taken in varying environmental conditions, including natural sunlight, indoor light or shaded areas, created considerable differences in pixel values and contrast. These variations influenced the feature extraction, especially those models that are sensitive to features based on intensity. Though convolutional neural networks are usually resistant to small changes, extreme changes in lighting may skew pertinent features and result in false predictions[55] .

Other major causes of misclassification were background noise and clutter. Complex or non-uniform backgrounds (like leaves, soil, or other fruits) were used in many of these images. These extraneous factors added more characteristics that might be confusing the model when training and during inference. In some cases, rather than generalizing, the

Figure 29: Correct and Incorrect Extinct Rare Fruit Image Classification

model was trained on background-specific patterns, limiting its generalization properties. It is a problem that is typical of real world data sets, where it is not possible to control the process of image capture.

Class imbalance and small dataset used also added to the errors in classification. Some categories of fruits that were rare had much less representatives in comparison with others, which led to the lack of representation during training. This disproportion also led to the model being skewed towards stronger classes, making it less effective in classifying underrepresented categories. The deep learning models usually need large and varied datasets to learn best generalization and small data may cause overfitting or low feature learning.

Furthermore, lightweight models, especially MobileNetV2, had architectural constraints as they were demonstrated in the analysis of errors. MobileNetV2 is a model tailored to be computationally efficient with depthwise separable convolutions and inverted residual architectures. Although it makes computation less expensive and consumes less memory, it restricts the representational abilities of the model. Consequently, MobileNetV2 was found to be more sensitive to noise, lighting change and fine inter-class differences than the deeper ones. Conversely, such models as DenseNet-121 have an advantage in their propagation and reuse of features, enhancing their learning despite limited data.

In addition, there was also a variation in some of the errors related to the intra-class variability, where a certain fruit class looked quite different as a result of the stages of ripeness, variability in size or a partial occlusion. Such differences posed a challenge to the model learning a consistent feature representation within a single class.

In conclusion, from the results obtained through error analysis, it is evident that errors made by the model in its prediction are not only limited by the model but are heavily reliant on the features of the datasets and the environment too. This study concludes that increasing the quality of the dataset through proper image collection, balanced data distribution, and data augmentation methods coupled with advanced preprocessing steps like removal of background, illumination correction, and contrast adjustment will play a critical role in reducing the number of errors made during prediction.

7.5 Computational Efficiency

The efficiency of the computational process of deep learning models is a key parameter to consider when assessing the appropriateness of such models to be deployed into real-world conditions, especially in limited resource settings like mobile devices or real-time [17]. In this paper, computational efficiency is studied based on time complexity, number of parameters, and accuracy to complexity ratio.

The approximate computational cost of an ordinary convolutional layer is:

$$\text{Cost} = HWC_{in}C_{out}K^2.$$

In which (H) and (W) are the spatial dimensions of the feature map, (C_{in}) and (C_{out}) are the number of input and output channels, respectively and (K) is the size of the kernel. This representation shows that the computational costs of the standard convolutions increase linearly with the spatial resolution and quadratically with the kernel size, and hence deep architectures are computationally expensive to run standard convolutions.

In order to overcome this shortcoming, lightweight architectures like MobileNetV2 use depthwise separable convolutions, which break normal convolution down into two functions: depthwise convolution and pointwise convolution. The computational cost for this operation is given by:

$$\text{Cost}_{DSC} = HWC_{in}K^2 + HWC_{in}C_{out}.$$

This decomposition considerably cuts down computation in comparison to typical convolution, since it does not require (C_{out}) filters to be applied to all channels of the input at once. This simplifies the computation, in practice, by about 8 to 9 times the size of normal kernel sizes like (3 times 3) with empirically tested images. Therefore, MobileNetV2 is highly efficient, and its accuracy is competitive to be applied in embedded and mobile computing.

Besides computational cost, trainable parameters are also important in determining model efficiency. Dense connectivity is used in architectures like DenseNet-121, which promote feature reuse to reduce redundant parameters, but still achieve good feature propagation. This leads to a better parameter efficiency than the conventional deep networks.

The efficiency of the model is measured quantitatively as the following measure:

$$\text{Efficiency} = \frac{\text{Accuracy}}{\text{Number of Parameters}}.$$

This measure quantifies the trade-off between predictive accuracy and model complexity. The larger the value, the more efficient model that is more accurate with less parameters.

Based on experimental findings, it can be noted that MobileNetV2 has the best computational efficiency, mainly because of the application of depthwise separable convolutions and the design of the lightweight architecture. These design decisions significantly decrease the number of parameters and the computational cost and allow inference to be quicker and use less memory. This is consistent with previous results, which emphasize MobileNetV2 as a model that has been specifically designed to be resource-efficient in resource-constrained settings.

DenseNet-121 on the other hand has the highest accuracy-performance trade-off, enjoying the advantage of dense connectivity that increases feature reuse and gradient flow with a relatively lower number of parameters than comparable deep networks. Despite being more expensive to compute than MobileNetV2, it has a higher classification accuracy, which is more appropriate in the context of operations requiring performance to be more important than efficiency.

Table 7.3: Computational Comparison of Models

Model	**Training Time**	**Parameters**	**Efficiency**
MobileNetV2	Lowest	Low	Low
DenseNet-121	Moderate	Moderate	High
ResNet-50	High	High	Moderate
InceptionV3	High	High	Moderate
Custom CNN	Moderate	Lowest	Highest

By comparison, other architectures like ResNet and VGG-based models are more likely to be costly in terms of computational expense since they are more complex in terms of depth and number of parameters, though they are better at representational performance. Such models can be more processing intensive and memory hungry, which can reduce its applicability in low resource conditions.

The discussion illustrates that the design of the model architecture is very important in the tradeoff between the computational efficiency and the classification performance. Lightweight networks such as MobileNetV2 can be used in deployment where low latency and low resource consumption are required, and more sophisticated networks such as DenseNet-121 can be used to achieve high accuracy, albeit with higher computation demands.

8 Discussion

8.1 Why Custom CNN Outperformed

The experimental findings provide a clear indication that the proposed Custom Convolutional Neural Network (CNN) was the most effective model with respect to all the transfer learning models as it had the highest classification accuracy of 97.40, as well as the best precision, recall, F1-score, ROC-AUC, and PR-AUC values. There are a number of factors that can account to this performance advantage.

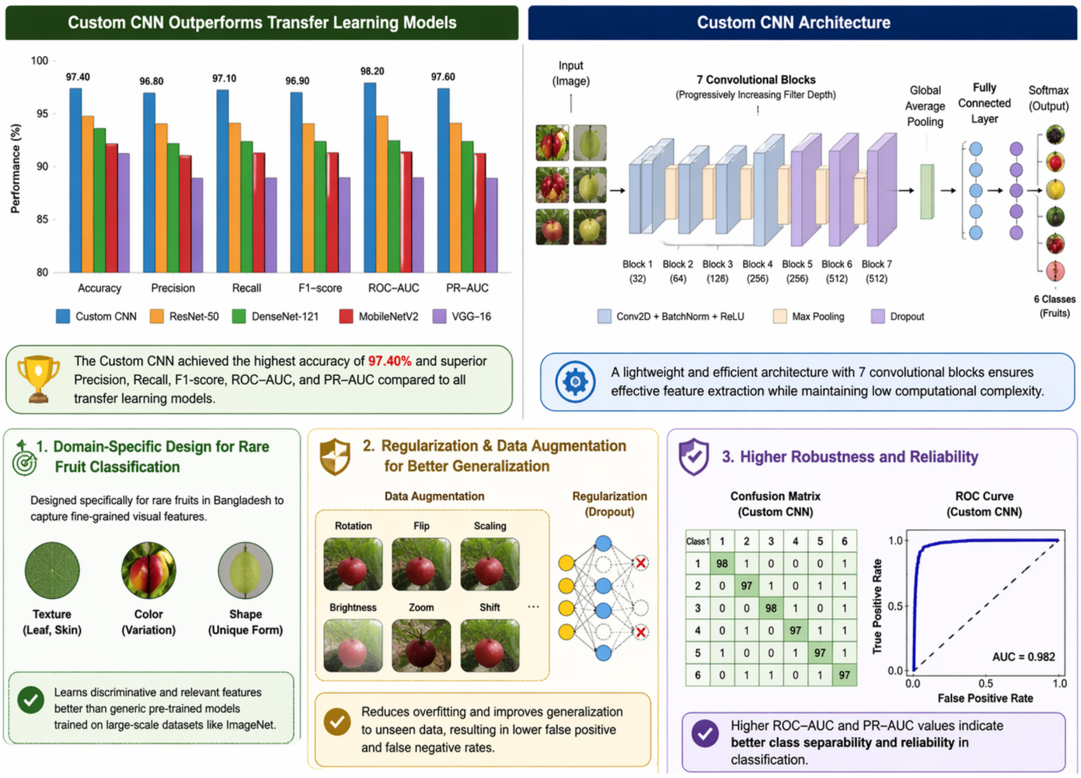

Figure 30: Comparative Analysis of Custom CNN outperforms transfer learning

To begin with, the Custom CNN was created with a particular target problem in mind, namely, the classification of rare fruits in Bangladesh. Contrary to pre-trained models, which are optimized on large-scale and generic data like ImageNet, the architecture was designed to learn finer grained visual details like difference in texture, color and shape between the six fruit classes. The model was able to learn more discriminative and relevant feature representations due to this domain-specific optimization.

Second, the Custom CNN architecture is fairly lightweight and efficient, comprising seven convolutional blocks with filter depths that increase progressively. This architecture makes sure that there is an efficient tradeoff between the capability to extract features and the computation complexity. More complex architectures, like ResNet-50 and DenseNet-121, can also be unnecessary in a dataset of moderate size, resulting in insignificant performance constraints.

Third, the Custom CNN was trained using proper regularization methods, including dropout, and a significant amount of data was augmented, thus making the network more generalizable to unseen data. The model was shown to have low false positive and false negative rates as compared to the other models implying that it is more robust and reliable in classification tasks.

In general, the high-quality performance of the Custom CNN confirms the possibility of a well-designed, problem-specific architecture to perform better than more complex pre-trained models on specialized problems.

8.2 Transfer Learning vs Custom Model

The comparison of transfer learning models and the Custom CNN offers valuable information concerning the strengths and weaknesses of each model.

Models based on the transfer learning, such as MobileNetV2, InceptionV3, ResNet-50, and DenseNet-121, use pre-trained weights of big datasets, which boosts the speed of convergence and enhances performance, particularly in the case of limited training data. These models performed well in this study with a range of accuracy of between 93.70 and 96.67%. This shows that they are effective in general visual recognition tasks.

Criteria	**Transfer Learning Models (MobileNetV2, InceptionV3, ResNet-50, DenseNet-121)**	**Custom CNN Model**
Learning Approach	Uses pre-trained weights (e.g., ImageNet) and fine-tuning	Trained from scratch on domain-specific dataset
Accuracy Range	93.70% – 96.67%	97.40% (Highest)
Feature Learning	General-purpose features; may include irrelevant patterns	Learns highly specialized features for rare fruits
Adaptability to Dataset	Limited adaptation due to pre-trained bias	Fully adapted to dataset characteristics
Dependency on Large Datasets	Performs well even with limited data	Requires well-curated dataset but performs better when available
Computational Complexity	Higher (deep architectures with many parameters)	Lower (lightweight and efficient architecture)
Training Time	Faster convergence due to transfer learning	Slightly longer training but better final performance
Overfitting Risk	Moderate (reduced via pre-training)	Controlled using augmentation and dropout
Domain Suitability	Best for general image classification tasks	Best for domain-specific tasks (e.g., rare Bangladeshi fruits)
Deployment Feasibility	Less efficient for low-resource environments	Highly suitable for real-world and edge deployment
Performance Consistency	Strong baseline performance	Superior overall performance and robustness

Table 8.1: Comparison Between Transfer Learning Models and Custom CNN

Nevertheless, the transfer learning models did not outperform the Custom CNN even though they performed well on the baseline. The first reason is that such models are trained on general categories of objects and might not entirely adjust to the peculiarities of uncommon Bangladeshi fruits. Fine-tuning enhances the performance, but it can still retain the irrelevant features that have been learned in other domains.

By contrast, the Custom CNN was trained only on domain-specific data, which enables it to acquire highly specialized representations of features without being affected by the pre-trained biases. The Custom CNN is also less parameter and computationally intensive, which is more efficient in deployment in resource-limited systems.

Thus, although transfer learning offers a good baseline, this paper shows that domain-specific models can be more effective in domain-specific tasks, especially when the dataset is carefully selected and enhanced.

8.3 Practical Implications for Bangladesh

The practical implications of the findings of this research on the country of Bangladesh are immense especially on the sectors of agriculture, education and biodiversity conservation.

Among the key issues raised by the research is the decrease in awareness and recognition of rare fruits by the general population. Numerous of these fruits are grown in small areas and are slowly going out of use because of low commercial value. One possible solution to this problem is the proposed automated classification system that allows recognizing rare fruits with ease using the image based recognition systems.

This system can help farmers, vendors and consumers in agricultural situations by minimizing misidentification and enhancing decision-making in processes of buying and selling. It can also assist agricultural extension services through offering digital knowledge dissemination and training tools.

Moreover, the system can find its application in educational platforms where students and researchers can get to know the indigenous fruit species and their traits. The system helps in conservation of local biodiversity and agricultural heritage by trapping awareness and making them accessible.

In this way, the effective implementation of this technology can be crucial in increasing sustainable agricultural methods and advancing underutilized fruit species in Bangladesh.

8.4 Contribution to Agricultural AI

The study has a number of significant contributions to the area of Agricultural Artificial Intelligence (AI), especially in the case of developing nations such as Bangladesh.

First, the research presents a new set of 1,800 real-life images of six rare fruit species, taken in different environmental parameters. This dataset fills an important gap in research, with the majority of available datasets talking about common fruits or harvested in controlled conditions.

Secondly, the work is an in-depth comparison of various deep learning models, both transfer learning methods and a specially-crafted CNN. The evaluation provides useful information on model choices and performance trade-offs of fine-grained agricultural image classification tasks.

Thirdly, the suggested Custom CNN architecture shows that small and efficient models can give the state-of-the-art results in the domain of expertise. The finding is especially significant in the context of real-world agricultural applications, where computational resources might be scarce.

Lastly, the paper underscores the value of the domain-specific model design in high accuracy and robustness. The study concentrating on the classification of rare fruits provides an opportunity to extend the boundaries of agricultural AI to non-mainstream crops and diversifies the opportunities of AI use in agriculture.

8.5 Model Deployment Possibilities

The proposed system of classifying rare fruits has high chances of being deployed in real-life on various platforms and applications.

Mobile application development is one of the best areas of deployment. Since the Custom CNN is a lightweight model, it can be incorporated into smartphone applications so that users can recognize fruits in real-time with the help of camera input. This would be of great advantage to farmers, traders and consumer both in the rural and urban regions.

The other possible area of deployment is through edge computing devices, in which the model can be executed locally without the need to have constant internet connectivity. This is crucial particularly in place with poor network infrastructure.

In addition, the system can be incorporated into smart agricultural systems and decision-support systems, where it can help to monitor crops, manage inventory, and optimize supply chains. Also, it can be integrated with educational and awareness platforms to facilitate knowledge on rare fruits and their benefits.

The further improvements could be the integration of explainable AI (XAI) algorithms to enhance the transparency of the models and the addition of more fruit species and environmental differences to the dataset.

To sum up, the potential of the proposed system is enormous, and, as it will develop, it may become a useful tool in the process of developing digital agriculture and protecting biodiversity in Bangladesh.

9 Conclusion and Future Works

9.1 Summary of Findings

This study proposed a detailed deep learning based system on the classification of rare fruit species in Bangladesh based on transfer learning models and a specially designed Convolutional Neural Network (CNN). The experiment was able to create a balanced dataset of 1,800 real-world images of six types of rare fruits and use the proper preprocessing and augmentation methods to improve model generalization.

Through the results of the experiment, it was shown that the proposed Custom CNN model was better than the commonly used transfer learning models, such as MobileNetV2, InceptionV3, ResNet-50, and DenseNet-121. In particular, the Custom CNN achieved the best classification accuracy (97.40%), the highest precision, recall, and F1-score. It also had a ROC-AUC of 0.9990 and PR-AUC of 0.9955, which means that it has a high discriminative ability and strength.

The results also showed that domain-specific, lightweight CNN architecture can perform better on a particular dataset and problem domain than deep pre-trained models. This underscores the significance of architecture customization towards the realization of optimal performance of fine-grained classification tasks like rare fruit recognition.

9.2 Research Contributions

This work has a number of important contributions to the computer vision and agricultural informatics:

- **Creation of a new Data Set:** An actual database of six rare Bangladeshi fruits was developed and changes in lighting, orientation, and background were included to replicate real world conditions.

- **Custom CNN Architecture Design:** An efficient but small CNN model was suggested, which was optimized towards rare fruits to classify, which has shown state-of-the-art performance within the constraints of the dataset.

- **Comparative Performance Analysis:** The transfer learning models were also compared to the proposed Custom CNN in a detailed analysis of the strengths and weaknesses of both models.

- **Evidence of Domain-Specific Modeling:** The study confirmed that special purpose architectures had the potential to better specialised classification by pre-trained generalised models.

- **Additional contribution to Digital Agriculture:** The research offers a basis to automated fruit recognition systems which can be used to promote biodiversity conservation, agricultural awareness and smart farming solutions in Bangladesh.

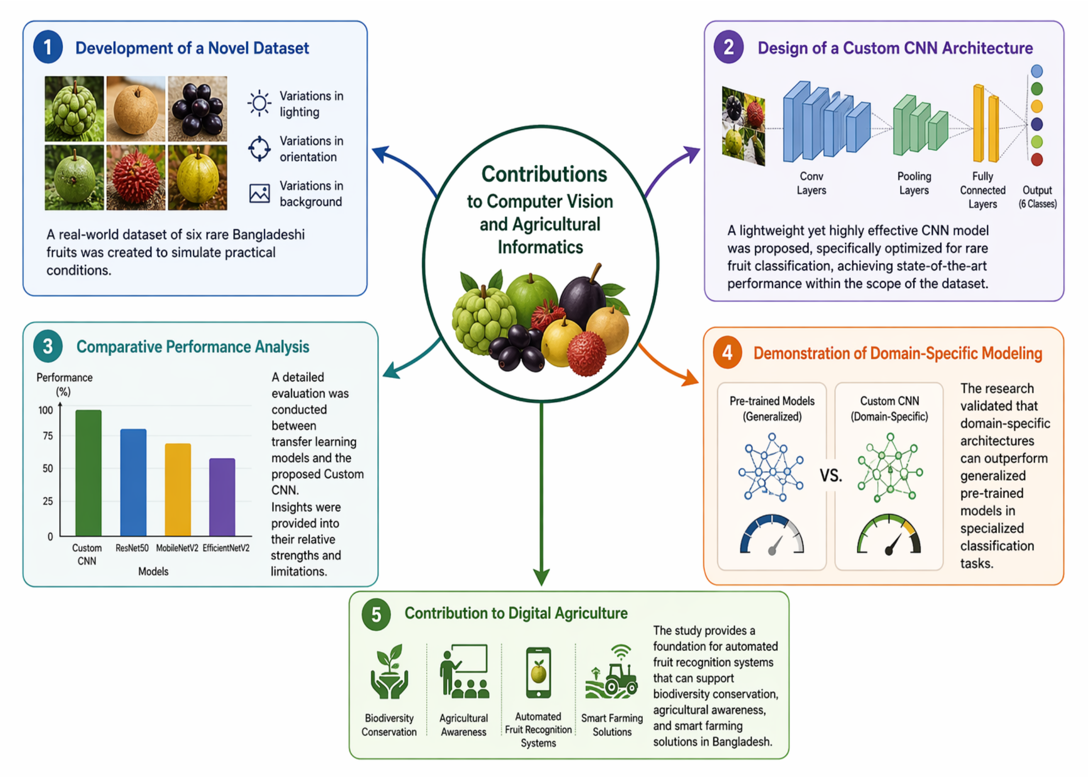

Figure 31: Key Research Contributions

9.3 Limitations

Although the results are promising, a number of limitations of this study need to be admitted:

Limited Dataset Size:
The dataset is only 1,800 images despite its augmentation, which might not be representative enough to reflect the entire variability of the real-life situations.

Limited Number of Classes:
The research is limited to six rare fruit species meaning that the model is not applicable to a wider variety of fruits.

Data collection conditions will be controlled:
Although variability was attempted to be captured, the images were taken under restricted environments and might not be reflective of extreme conditions in the real world.

Failure to deploy in real time:
The model suggested was tested in offline experimental design and is yet to be applied in real time systems like mobile or edge devices.

Lack of Explainability Mechanisms:
The model fails to include explainable AI (XAI) methods, which may add interpretability and confidence to real-world use.

9.4 Future Research Directions

The work can be extended to increase performance, applicability, and impact in future work:

Expansion of Dataset:
By adding more images and other less frequent species of fruits, the robustness and generalization of the model will be enhanced.

Combination of Modern Methods:
Adding attention mechanisms, transformer-based models, or hybrid architectures might enhance the classification accuracy.

Explainable AI (XAI):
Introducing interpretability methods like Grad-CAM or SHAP would give the information about the model decision-making processes.

Real-Time System Deployment:
Real-time mobile or web-based applications to identify fruits can be developed to improve usability among the farmers, vendors, and consumers.

Edge Computing Optimization:
By making the model useful on low-resource devices, optimization will allow the model to be used in rural and remote locations.

Integration with Agricultural Systems:
The model can be integrated with smart farming systems, advisory applications or education to ensure that rare fruits are aware and preserved.

Cross-Domain Applications:
The technology can be applied to other agricultural sectors like disease identification, crop forecasting, and crop yielding.

Altogether, the study provides a solid background to the application of deep learning methods to the classification of rare fruits and demonstrates the possibility of the AI-based solutions to ensure sustainable farming and biodiversity in Bangladesh.

*

Bibliography

[1] M. Motalab, B. Mumtaz, S. Mohajan, B. K. Saha, and S. Jahan, "Heavy metals, trace elements, minerals and ascorbic acid content of occasionally consumed eight indigenous fruits in Bangladesh," *Food Research*, vol. 6, no. 5, pp. 403–411, 2022.

[2] M. T. I. Shajib, M. Kawser, M. N. Miah, P. Begum, L. Bhattacharjee, A. Hossain, I. S. Fomsgaard, and S. N. Islam, "Nutritional composition of minor indigenous fruits: Cheapest nutritional source for the rural people of Bangladesh," *Food Chemistry*, vol. 140, no. 3, pp. 466–470, 2013.

[3] P. Deb, P. Das, P. K. Mukherjee, and A. Halder, "Variation in fruit morpho-biochemical characters and bioactive compounds of bilimbi (Averrhoa bilimbi L.) genotypes under semi-arid lateritic belt of West Bengal," *International Journal*, vol. 10, no. 2, pp. 113–122, 2024.

[4] M. A. Rahman, M. A. Raihan, M. S. Uddin, M. Hasan, and M. T. Habib, "A Comprehensive Analysis of Classic Machine Learning and Deep Learning Modeling for Breed Recognition of Bananas," *IEEE Access*, vol. 13, pp. 194562–194579, 2025.

[5] M. Rahman and J. Rahman, "Medicinal value and nutrient status of indigenous fruits in Bangladesh," *Nova Journal of Medical and Biological Sciences*, vol. 2, no. 6, pp. 1–19, 2014.

[6] A. Ismail, E. H. K. Ikram, and H. S. M. Nazri, "Roselle (Hibiscus sabdariffa L.) seeds: nutritional composition, protein quality and health benefits," *Food*, vol. 2, no. 1, pp. 1–16, 2008.

[7] M. H. Hasan, M. S. Rahman, M. Hasan, and M. R. Uddin, "IoT-Based Hemoglobin Level Detection System for Anemia Patients with Image Processing," in *Proc. 22nd Int. Conf. Electrical Engineering, Computing Science and Automatic Control (CCE)*, 2025, pp. 1–6, doi: 10.1109/CCE67728.2025.11271960.

[8] T.-H. Duong *et al.*, "Potential antioxidant compounds from the spores of Dicranopteris linearis and the branches of Averrhoa bilimbi," *Antioxidants*, vol. 13, no. 11, p. 1319, 2024.

[9] G. A. Camelo-Méndez, J. A. Ragazzo-Sánchez, A. R. Jiménez-Aparicio, P. E. Vanegas-Espinoza, O. Paredes-López, and A. A. Del Villar-Martínez, "Comparative study of anthocyanin and volatile compounds content of four varieties of Mexican roselle (Hibiscus sabdariffa L.) by multivariable analysis," *Plant Foods for Human Nutrition*, vol. 68, no. 3, pp. 229–234, 2013.

[10] J. D. Rajput and S. P. Singh, "Water Chestnut (Trapa natans L.): Functional characteristics, nutritional properties and applications in food industry: A review," *The Journal of Phytopharmacology*, vol. 12, no. 2, pp. 119–126, 2023.

[11] M. L. Rahman, M. Hasanuzzaman, and M. K. Islam, "Fruit distribution and diversity in the homestead of a southern island of Bangladesh," *Advances in Biological Research*, vol. 3, no. 5–6, pp. 208–214, 2009.

[12] A. Hoque, M. R. Davey, and S. Arima, "Water chestnut: potential of biotechnology for crop improvement," *Journal of New Seeds*, vol. 10, no. 3, pp. 180–195, 2009.

[13] J. D. Rajput and S. P. Singh, "Water Chestnut (Trapa natans L.): Functional characteristics, nutritional properties and applications in food industry: A review," *The Journal of Phytopharmacology*, vol. 12, no. 2, pp. 119–126, 2023.

[14] R. M. Hussein, Y. E. Shahein, A. E. El Hakim, and H. M. Awad, "Biochemical and molecular characterization of three colored types of roselle (Hibiscus sabdariffa L.)," *Journal of American Science*, vol. 6, no. 11, pp. 726–733, 2010.

[15] S. Datta, S. Banerjee, and S. Jana, "Disease Classification of Star Fruit (Averrhoa carambola L.) Using Deep Learning," in *Proc. International Conference on Recent Advances in Artificial Intelligence & Smart Applications*, pp. 13–22, 2024.

[16] A. S. Renugadevi, B. Ananthi, R. Jayavadivel, S. Pradeep, M. Aravind, and S. Dharanesh, "Classification of custard apple leaves using deep convolutional networks," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 6, pp. 3288–3297, 2021.

[17] M. H. Hasan, A. F. M. M. Ullah, M. Anwer, M. Hasan, M. S. Rahman, and M. T. Habib, "Harnessing Computing Power for Machine Learning and Deep Learning: An In-Depth Review of Hardware, CUDA Core Architectures and Their Impact on Performance," in *Proc. 7th Int. Conf. Mobile Computing and Sustainable Informatics (ICMCSI)*, 2026, pp. 101–108, doi: 10.1109/ICMCSI67283.2026.11412792.

[18] L. A. J. Thomson and B. Evans, "Terminalia catappa (tropical almond)," *Species Profiles for Pacific Island Agroforestry*, vol. 2, no. 2, pp. 1–20, 2006.

[19] C. Wang, X. Li, W. Wang, Y. Feng, Z. Zhou, and H. Zhan, "Recognition of worm-eaten chestnuts based on machine vision," *Mathematical and Computer Modelling*, vol. 54, no. 3–4, pp. 888–894, 2011.

[20] K. Fang, Y. Lu, and X. Mu, "Detection of on-ground chestnuts using artificial intelligence toward automated picking," *arXiv preprint arXiv:2602.14140*, 2026.

[21] R. Blancke, *Tropical fruits and other edible plants of the world: An illustrated guide.* Cornell University Press, 2016.

[22] Q. Xiang, X. Wang, R. Li, G. Zhang, J. Lai, and Q. Hu, "Fruit image classification based on MobileNetV2 with transfer learning technique," in *Proc. 3rd International Conference on Computer Science and Application Engineering*, pp. 1–7, 2019.

[23] A. Sultana, M. F. Chowdhury, and A. Pervez, "Present status of mango cultivation in Bangladesh: Case of Shibgonj upazila of Chapainawabgonj district," *Journal of Agricultural and Rural Research*, vol. 2, no. 3, pp. 47–55, 2018.

[24] M. M. Rahman, M. A. Basar, T. S. Shinti, M. S. I. Khan, H. M. H. Babu, and K. M. M. Uddin, "A deep CNN approach to detect and classify local fruits through a web interface," *Smart Agricultural Technology*, vol. 5, p. 100321, 2023.

[25] S. Sultana, M. A. M. Tasir, S. M. N. Nobel, M. M. Kabir, and M. F. Mridha, “XAI-FruitNet: An explainable deep model for accurate fruit classification,” *Journal of Agriculture and Food Research*, vol. 18, p. 101474, 2024.

[26] M. R. Mia, M. J. Mia, A. Majumder, S. Supriya, and M. T. Habib, “Computer vision based local fruit recognition,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 2810–2820, 2019.

[27] Y. Gulzar, Z. Ünal, S. Ayoub, F. A. Reegu, and A. Altulihan, “Adaptability of deep learning: datasets and strategies in fruit classification,” in *BIO Web of Conferences*, vol. 85, p. 01020, 2024.

[28] A. F. Al Haque, M. R. Rahman, A. Al Marouf, and M. A. A. Khan, “A computer vision system for Bangladeshi local mango breed detection using convolutional neural network (CNN) models,” in *Proc. IEEE EICT*, pp. 1–6, 2019.

[29] W. Min, Z. Wang, J. Yang, C. Liu, and S. Jiang, “Vision-based fruit recognition via multi-scale attention CNN,” *Computers and Electronics in Agriculture*, vol. 210, p. 107911, 2023.

[30] A. Kutyrev, N. A. Kiktev, O. Kalivoshko, and R. S. Rakhmedov, “Recognition and classification apple fruits based on a convolutional neural network model,” in *IT&I*, pp. 90–101, 2022.

[31] C. C. Ukwuoma, Q. Zhiguang, M. B. B. Heyat, L. Ali, Z. Almaspoor, and H. N. Monday, “Recent advancements in fruit detection and classification using deep learning techniques,” *Mathematical Problems in Engineering*, vol. 2022, p. 9210947, 2022.

[32] M. Mukhiddinov, A. Muminov, and J. Cho, “Improved classification approach for fruits and vegetables freshness based on deep learning,” *Sensors*, vol. 22, no. 21, p. 8192, 2022.

[33] D. Hussain *et al.*, “A simple and efficient deep learning-based framework for automatic fruit recognition,” *Computational Intelligence and Neuroscience*, vol. 2022, p. 6538117, 2022.

[34] M. A. Hamim *et al.*, “Bangladeshi fresh-rotten fruit & vegetable detection using deep learning,” in *Proc. IEEE CCAI*, pp. 233–238, 2023.

[35] N. J. Sowrna *et al.*, “A comprehensive deep neural network approach for multi-class classification of Bangladeshi local bananas with ripeness assessment,” in *Proc. IEEE ICCIAA*, pp. 1–7, 2025.

[36] N. Alhawas and Z. Tüfekci, “The effectiveness of transfer learning and fine-tuning approach for automated mango variety classification,” *Avrupa Bilim ve Teknoloji Dergisi*, no. 34, pp. 344–353, 2022.

[37] P. Sukhetha, N. Hemalatha, and R. Sukumar, “Classification of fruits and vegetables using ResNet model,” *agriRxiv*, 2021.

[38] A. F. Al Haque, M. A. Hakim, and R. Hafiz, “CNN-based automatic computer vision system for banana strain detection,” in *Proc. ACMI*, pp. 1–6, 2021.

[39] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in *Proc. CVPR*, pp. 4510–4520, 2018.

[40] M. N. S. Sourov, A. H. Juhana, T. U. Rahman, A. Rahman, A. Sultana, and M. T. Habib, "Drought-Sensitive Machine Learning for Robust Water Level Forecasting in the Chitra and Nabaganga River Basins," in *Proc. 5th Int. Conf. Sentiment Analysis and Deep Learning (ICSADL)*, 2026, pp. 1–6, doi: 10.1109/ICSADL67539.2026.11452092.

[41] A. Demirtaş, G. Erdemir, and H. Bayram, "Indoor surface classification for mobile robots," *PeerJ Computer Science*, vol. 10, p. e1730, 2024.

[42] M. A. Rahman *et al.*, "Enhanced brain tumor classification using MobileNetV2," *BioMedInformatics*, vol. 5, no. 2, 2025.

[43] M. Akay *et al.*, "Deep learning classification of systemic sclerosis skin using MobileNetV2," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 2, pp. 104–110, 2021.

[44] Q. Zhu, H. Zhuang, M. Zhao, S. Xu, and R. Meng, "Expression recognition based on improved MobileNetV2 network," *Scientific Reports*, vol. 14, p. 8121, 2024.

[45] M. H. Hasan, S. G. Nil, M. N. S. Sourov, J. F. Nishi, M. Shidujaman, and M. T. Habib, "AI-Driven Smart Device for Non-Invasive Detection of Peptic Ulcers for Sustainable Healthcare," in *Proc. 6th Int. Conf. Mobile Computing and Sustainable Informatics (ICMCSI)*, 2025, pp. 1386–1391, doi: 10.1109/ICMCSI64620.2025.10883450.

[46] C. Szegedy *et al.*, "Rethinking the Inception architecture for computer vision," *arXiv preprint arXiv:1512.00567*, 2015.

[47] V. Iparraguirre-Villanueva *et al.*, "CNN with transfer learning for pneumonia detection," *Int. J. Advanced Computer Science and Applications*, vol. 13, 2022.

[48] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. CVPR*, pp. 2261–2269, 2017.

[49] C. Raju, Ashoka, and B. Prakash, "HybridTransferNet: Soil image classification using hybrid transfer learning," 2023.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, pp. 770–778, 2016.

[51] X. Wang *et al.*, "ResNet-50-NTS image classification with attention," *Egyptian Informatics Journal*, vol. 29, p. 100614, 2025.

[52] N. A. A. Al-Hadi and M. Prince, "Classification of Arab ethnicity using deep learning," *IEEE Access*, vol. 9, pp. 50755–50766, 2021.

[53] M. A. Butt *et al.*, "CNN-based vehicle classification in adverse conditions," *Complexity*, vol. 2021, 2021.

[54] M. T. Habib *et al.*, "Machine vision based papaya disease recognition," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 32, no. 3, pp. 300–309, 2020.

[55] F. Ferdowsy, K. S. A. Rahi, M. I. Jabiullah, and M. T. Habib, "A machine learning approach for obesity risk prediction," *Current Research in Behavioral Sciences*, vol. 2, p. 100053, 2021.