



**DETECTION OF IMPURITIES FROM ETHIOPIAN INJERA
USING DEEP LEARNING**

A Thesis Presented

by

Getachew Bualew Digafe

to

The Faculty of Informatics

of

St. Mary's University

**In Partial Fulfillment of the Requirements
for the Degree of Master of Science**

in

Computer Science

June, 2024

ACCEPTANCE

**DETECTION OF IMPURITIES FROM ETHIOPIAN INJERA
USING DEEP LEARNING**

By

Getachew Bualew Digafe

**Accepted by the Faculty of Informatics, St. Mary's University, in partial fulfilment of
the requirements for the degree of Master of Science in Computer Science**

Thesis Examination Committee:

Internal Examiner

Signature and Date

External Examiner

Signature and Date

Dr. Alembante Mulu Kumlign

Dean, Faculty of Informatics

Signature and Date

June, 2024

DECLARATION

I, the undersigned, declare that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for the thesis work have been fully acknowledged.

Getachew Bualew Digafe

Addis Ababa

Ethiopia

This thesis has been submitted for examination with my approval as advisor.

Dr. Alembante Mulu Kumlign

Addis Ababa

Ethiopia

June 2024

Acknowledgment

First and foremost, I want to express my deepest gratitude to God for His blessings and guidance throughout this research. His presence and support have been a constant source of strength and inspiration. My heartfelt thanks go to Alembante Mulu (PhD), my research advisor, for his unwavering support, invaluable insights, and continuous encouragement. His expertise and guidance have been crucial in shaping the direction and quality of this research. I deeply appreciate his commitment to his availability to discuss ideas, provide feedback, and address challenges throughout this journey.

Special thanks to my wife, Rediet Gebremedhin, for her patience and meticulous assistance in the data preparation phase. Her dedication to ensuring the quality of the data has been instrumental in the successful completion of this research. I would also like to extend my gratitude to Tessfu Geteye (PhD) for his valuable support in specific parts of my research. His assist has significantly contributed to improving the overall quality of the work.

I am deeply grateful to my colleagues, Yigeremu Yohanness and Temesgen Demilew, for their crucial contributions. Yigeremu's support and idea-sharing have provided fresh perspectives and innovative solutions, enhancing the research's robustness. Temesgen's collaborative spirit and provision of necessary hardware tools have been essential in the experimental phases, enabling smooth and efficient progress. Additionally, I thank my office manager, Henok Doni, for creating an environment conducive to my research.

In conclusion, I acknowledge and appreciate all the individuals whose contributions, support, and encouragement have been integral to the success of this research. Thank you for being part of this journey.

Table of Contents

DECLARATION.....	II
ACKNOWLEDGMENT.....	III
LIST OF ACRONYMS.....	VI
LIST OF FIGURES	VIII
LIST OF TABLES	IX
ABSTRACT.....	X
CHAPTER ONE	1
1. INTRODUCTION	1
1.1 BACKGROUND OF THE STUDY	1
1.2 MOTIVATION OF THE STUDY	3
1.3 STATEMENT OF PROBLEM.....	3
1.4 RESEARCH QUESTIONS.....	5
1.5 OBJECTIVE OF THE STUDY	5
1.5.1 General Objective	5
1.5.2 Specific Objective	5
1.6 SCOPE AND LIMITATION OF THE STUDY	6
1.7 SIGNIFICANCE OF THE STUDY	6
1.8 EVALUATION AND TESTING METRICS	7
1.9 MODEL DEVELOPMENT TOOLS	7
1.10 ORGANIZATION OF THE THESIS	8
CHAPTER TWO	9
2. LITERATURE REVIEW	9
2.1 INTRODUCTION.....	9
2.2 FUNDAMENTALS OF DIGITAL IMAGE PROCESSING.....	9
2.2.1 Low-level Image processing	10
2.2.2 Intermediate level Image processing.....	13
2.2.4 High level Image processing.....	17
2.3 NEURAL NETWORKS	18
2.4 DEEP LEARNING AND FOOD IMAGE CLASSIFICATION	19
2.4.1 CNN For Food Image Detection and Classification	26
2.4.2 Transfer Learning.....	27
2.5 EVALUATION TECHNIQUES	28

2.6	RELATED WORKS	30
2.6.1	Global Studies on Food Detection and Classification.....	30
2.6.2	Local Studies on Injera Impurities Detection and Classification.....	32
2.7	SUMMARY	34
CHAPTER THREE		35
3.	DESIGN AND METHODOLOGY	35
3.1	INTRODUCTION.....	35
3.2	PROPOSED SYSTEM ARCHITECTURE FOR DETECTION OF IMPURITIES IN INJERA	36
3.2.1	Data Collection and Preparation	37
3.2.2	Data Pre-processing	39
3.2.3	Deep and Handcrafted Feature Extraction	43
3.2.4	Feature Fusion.....	44
3.2.5	Dimensionality Reduction.....	45
3.3	EXPERIMENTAL DESIGN	45
3.3.1	Hyperparameter Tuning	46
3.3.1.1	HYPERPARAMETER TUNING TECHNIQUES	47
3.4	SUMMARY	48
CHAPTER FOUR.....		49
4.	EXPERIMENTAL RESULTS AND DISCUSSIONS	49
4.1	INTRODUCTION.....	49
4.2	DATA PREPARATION AND PRE-PROCESSING.....	49
4.3	HYPER PARAMETER TUNING	50
4.4	MODEL BUILDING AND EXPERIMENTAL RESULTS.....	51
4.5	EVALUATION OF MODEL.....	54
CHAPTER FIVE		60
5.	CONCLUSION AND RECOMMENDATION	60
5.1	CONCLUSIONS	60
5.2	CONTRIBUTIONS OF THE STUDY	61
5.3	RECOMMENDATIONS	61
REFERENCE.....		62

List of Acronyms

Abbreviations	Definition
AHE	Adaptive Histogram Equalization
AI	Artificial Intelligence
ANN	Artificial Convolutional Networks
API	Application Programming Interface
CHE	Classical Histogram Equalization
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CV	Computer Vision
DAE	De-noising autoencoder
DHE	Dynamic Histogram Equalization
DL	Deep Learning
DNN	deep neural network
FBC	Fana Broadcasting Corporation
FN	False Negatives
FP	False Positives
GAN	Generative Adversarial Network
GB	Gigabyte
GDP	Gross domestic product
GHz	Gigahertz
GIST	Image features from Spatial Histograms
GLCM	Gray level co-occurrence matrix
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
HOG	Histogram of Oriented Gradients
HSV	Hue Saturation Value
IEEE	Institute of Electrical and Electronics Engineers
ILSVRC	Large Scale Visual Recognition Challenge
KG	Kilogram
KNN	K-Nearest Neighbours
LBP	Local Binary Pattern
LSTM	Long Short-Term Memory
LTP	Local Ternary Patterns

ReLU	Rectified Linear Unit
ML	Machine Learning
MLP	Multi-Layer Perceptron
MP	Megapixel
MSMVFA	Multi-Scale Multi-View Feature Aggregation
macOS	Mac Operating Systems
PCA	Principal Component Analysis
PDF	Portable Document Format
PoI	Pixel of Interest
RAM	Random Access Memory
RBM	Restricted Boltzmann Machine
RCNN	Region-Based Convolutional Neural Network
RF	Random Forest
RGB	Red Green Blue
RNN	Recurrent Neural Network
RoI	Region of Interest
SIFT	Scale-Invariant Feature Transform
SLGC	SURF-Local and Global Colour
SURF	Speeded-Up Robust Feature
SVM	Support Vector Machine
TB	Terabyte
TN	True Negatives
TP	True Positives
UEC	University of Electro-Communications
YOLO	You Only Look Once

List of Figures

Figure 1- 1 Smart phone used for Image capturing [113]	8
Figure 2- 1 Different levels in image processing process	9
Figure 2- 2 The architecture of Deep learning technologies R. Dastres et al. [59]	19
Figure 2- 3 Relationship between Human vision, CV, ML, DL, and CNN	20
Figure 2- 4 Schematic diagram of a basic CNN architecture [73]	22
Figure 2- 5 Example of convolution operation	23
Figure 2- 6 A convolution operation with one zero padding.....	24
Figure 2- 7 Schematic diagram of pool layer operation type.....	24
Figure 2- 8 Fully Connected Layer [67]	25
Figure 2- 9 Learning Process of Transfer Learning	27
Figure 3- 1 Proposed general System architecture.....	37
Figure 3- 2 Sample collected ‘injera’ images.....	39
Figure 3- 3 Resized sample Image.....	40
Figure 3- 4 Augmented sample Image	41
Figure 3- 5 Segmented Image using thresholding (a) and Masked outer most parts (b)	42
Figure 3- 6 data splitting process	43
Figure 3- 7 multi-feature fusion strategy	45
Figure 4- 1 (a)Training, Testing and Validation data size, (b) Original and augmented images.....	50
Figure 4- 2 a) Accuracy vs Epoch for AlexNet b) Loss vs Epoch for AlexNet	
c) Accuracy vs Epoch for ResNet50 d) Loss vs Epoch for ResNet50	52
Figure 4- 3 Accuracy vs Epoch for CNN and Handcrafted Features Fusion Approach b) Loss vs Epoch for CNN and Handcrafted Features Fusion Approach c) Accuracy vs Epoch for Pure CNN-based approach d) Loss vs Epoch for Pure CNN-based approach.....	53
Figure 4- 4 Comparison of Accuracy vs. Model.....	54
Figure 4- 5 Confusion matrix of CNN Model in pure CNN-based approach model	56
Figure 4- 6 Confusion matrix of CNN model for CNN and Handcrafted Features Fusion Approach..	57
Figure 4- 7 Some correctly classified images	58
Figure 4- 8 Some incorrectly classified images	58

List of Tables

Table 2- 1 Confusion Matrix.....	29
Table 2- 2 Summary of related works.....	33
Table 5- 1 hyperparameter search space	50
Table 5- 2 Determined Optimal Hyper parameters value	51
Table 5- 3 comparison of the two CNN models in a table format	53
Table 5- 4 Accuracy, F1 score, Precision and Recall for all 4 classes in the full CNN model	55
Table 5- 5 Precision, recall, F1 score and Accuracy values for Concatenated feature CNN model	55

Abstract

Injera is a fermented Ethiopian traditional food usually prepared from teff flour. Even though teff is the most popular cereal for injera preparation, other cereals such as sorghum, maize, barley, wheat, and rice flour, or combinations of these, are sometimes used. Since not everyone has the means or time to make injera at home, it is often purchased from shops, supermarkets, hotels, and restaurants. Some producers adulterate injera by mixing teff flour with cheaper cereals or harmful substances for motives such as market surplus and Cost reduction to compete and desire for higher profit margins. This poses health risks to consumers and marketing challenges for the country. Although many researchers have worked on food detection and classification, their datasets often lacked sufficient class similarity and did not quantify the proportion of impurities added, making it unreliable for real-life testing. The visual similarities between pure and adulterated injera make manual identification of impurities difficult, and there is no existing research on identifying impurities in Teff injera using deep learning. In this thesis, we developed models using both deep learning algorithms alone approach and deep learning algorithms by combining deep learning features with handcrafted features. The injera dataset was prepared at home traditionally by mixing 15% sawdust flour with 85% pure white teff flour, and 15% sorghum flour with 85% pure red teff flour. Additionally, we prepared 100% pure white teff injera and 100% pure red teff injera. After 12 hours, we captured images of the prepared injera using a Samsung Galaxy M13 50-megapixel camera and labelled them into 'pure white teff injera,' 'white teff with sawdust injera,' 'pure red teff injera,' and 'red teff with sorghum (zengada) injera' classes. We applied various pre-processing techniques, including resizing, filtering, segmentation, enhancement, and augmentation. Then, hyperparameter values were identified for each model using the random search tuning method. For experimentation, we utilized pretrained models such as AlexNet and ResNet50, and we built CNN, LSTM, and YOLO models from scratch for both approaches. Handcrafted features were extracted using Gray Level Co-occurrence Matrix and Local Binary Pattern methods. The experiment results showed that using deep learning algorithms alone, we achieved accuracy of 79% with CNN, 58% with AlexNet, and 54% with ResNet50. When combining deep learning features with handcrafted features, the CNN model achieved 77% accuracy. Overall, the CNN built from scratch attained the highest accuracy in both approaches compared to the other models.

Keywords: Injera, Impurity detection, Image processing, Convolutional Neural Network, Handcrafted features, Concatenated features.

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the study

In the realm of food consumption, various regions across the globe rely on diverse cereal crops as their dietary staples. In Ethiopia, “Injera”, a traditional Ethiopian flatbread, usually prepared from teff flour. It stands as a staple food in Ethiopia, Eritrea, and other parts of East Africa. Other cereals such as sorghum, maize, barley, wheat, and rice flour, or combinations of these, are sometimes used. It plays a crucial role in the livelihoods of rural producers, processors, and consumers. Furthermore, Ethiopia generates substantial revenue through the export of injera.

In Ethiopia, the issue of food impurity has emerged as a pressing concern, posing potential health hazards and contributing to various risks. Foods such as milk, beef, honey, butter, juices and including injera have been subject to alterations, both within Ethiopia and globally.

Despite its widespread popularity, injera can be susceptible to adulteration during its production process. As media reports in Ethiopia, injera adulteration may include harmful substances such as sawdust, Jasso, and even mold, and mixing other cheaper cereals flour with teff, thereby posing risks of food poisoning and health problems. The studies A. Choudhary et al. [1] indicate, that a significant portion of the food consumed is susceptible to food fraud, involving the addition of undesirable ingredients or cheaper substitutes. Motives for such practices include market surplus, cost reduction to match competitors, desire for higher profit margins, lack of awareness regarding the impact on nutritional value, and inadequate regulatory control.

Teff, Ethiopia's most extensively cultivated cereal crop, encompasses three primary categories: white, red, and mixed (sergegna). Teff stands as a gluten-free cereal crop that offers potential health benefits, including diabetes prevention, management of iron deficiency due to its high iron content, and assistance in managing celiac disease M. L. Habte *et al.* [2]. The human body requires varying amounts of iron daily to function optimally, and a deficiency in this essential nutrient can lead to maternal mortality in pregnant women, as well as increased fatigue and weakness. Studies of M. L. Habte *et al.* [2], indicates that men typically need approximately 10 milligrams of iron per day, while women require around 15 milligrams. Iron is essential for the health of children and pregnant women. In 2021, the International Diabetes Federation reported that approximately 537 million adults aged 20-79 were living with diabetes

worldwide. This figure is expected to increase, especially in low- and middle-income nations, including Ethiopia.

Studies [3], have revealed that red teff possesses the highest iron concentration (2472.7 milligrams(mg) /kilogram(kg)), followed by mixed teff (1440.9 mg/kg), while white teff exhibits the lowest concentration (881 mg/kg). Teff injera boasts a higher iron content compared to other Ethiopian cereal foods. For instance, in sorghum (zengada), the iron content decreases approximately fivefold compared to red teff. Consequently, the mixture of red teff flour with red sorghum flour compromises the necessary iron content, thereby leading to health issues, including maternal mortality in pregnant women. Due to the difficulty in visually identifying pure injera, the development of an automatic image-based food detection model becomes crucial.

The accurate description of visual information, also known as feature extraction, and the selection of an appropriate classifier represent significant steps in addressing image-based food detection and classification challenges S. Jiang *et al.*[4]. Feature extraction can be achieved using manually created techniques or convolutional neural networks (CNNs). Examples of manually engineered features include Local Binary Pattern (LBP), Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), and Gray Level Co-occurrence Matrices (GLCM). These hand-crafted feature extraction techniques are commonly used in computer vision and image processing applications. However, these features may possess limited discriminative power, thereby hindering their ability to distinguish between different classes accurately X. Huang *et al.* [5]. On the other hand, convolutional neural networks, a popular deep learning technique, have demonstrated remarkable performance by directly extracting features from the original pixels of the image X. Huang *et al.*[5], D. Liu *et al.* [6] and [7]. Nonetheless, convolutional neural networks often lack semantic information and suffer from the issue of overfitting. Additionally, they may struggle to extract comprehensive feature information necessary for image content characterization [7], J. Lu et al.[8] . Relying solely on a single feature for image classification can be susceptible to inaccuracies caused by variations in object detection, angle, scale, noise interference, and other factors, ultimately leading to reduced classification accuracy. Therefore, an effective integration of classical machine learning (ML) and deep learning features becomes essential for improving food picture categorization using machine learning classifier X. Huang *et al.*[5], D. Liu et al. [6], [7] and J. Lu et al. [8].

The specific objective of this research is to identify impurities in injera, with a particular focus on different types of injera: pure white teff (Magna) injera, pure white teff (Magna) flour mixed with sawdust injera, pure red teff flour injera, and injera made from a combination of red teff flour and sorghum (zengada) flour. By utilizing image processing techniques and deep learning algorithms, the research aims to develop a model for detecting and classifying impurities in injera. This method will leverage visual information extracted from images of injera samples. The ultimate goal is to address the health and economic issues associated with impurities, ensuring the well-being of individuals who consume injera. Furthermore, this research will explore and compare the performance of two deep learning models for impurity detection and classification in injera, one that utilizes only deep learning features, and another that incorporates the fusion of deep learning features and handcrafted classical machine learning (ML) features. By comparing the accuracy of these two approaches, the study will provide insights into whether the combination of deep features and classical ML features, can lead to improved impurity detection and classification in injera compared to using deep learning features alone.

1.2 Motivation of the study

One possible motivation for identifying impurities in Teff injera is to ensure food safety and quality. Millions of people in Ethiopia and other parts of East Africa eat Teff injera every day [9], [10].

For example, the addition of sawdust flour to pure white teff flour can lead to injera with a lower nutritional value and potentially harmful contaminants. Similarly, the use of a combination of red teff flour and, sorghum flour with low iron content, can result in injera with a lower iron content, which can be a concern in areas where iron deficiency is prevalent. The burden of foodborne disease is very high, especially in developing countries, and can have major economic and social implications, such as lost productivity, medical costs, and decreased quality of life A. H. Havelaar *et al.* [11].

Overall, since no research has been done on the identification of such Injera impurities, the development of food impurity detection model may greatly impact public health by ensuring food safety and quality and can support the expansion of local food industries and businesses, making it an important area for research and innovation.

1.3 Statement of Problem

Injera, a traditional Ethiopian food usually made from teff at home. However, not everyone has the means or time to prepare injera at home, leading to its frequent purchase from various commercial sources such as shops, supermarkets, hotels, and restaurants. Unfortunately, this convenience comes with a risk as there is a possibility of adulteration or alteration, which can compromise the original quality and content of the injera A. Choudhary *et al.*[1].

In a recent incident, the residents of Addis Ababa city were unfortunately exposed to adulterated commercial injera products. The occurrence was brought to public attention through various communication media channels, including the Fana Broadcasting Corporation (FBC), as well as radio and social media platforms [12]. Following the incident, the Addis Ababa Police announced that they have acted by arresting individuals involved in the illicit practice of mixing teff powder with foreign objects and wooden sawdust for distribution to the public. During a search of their warehouse, authorities discovered 47 quintals of unidentified plant roots and wood chips (Segatura), along with 167 quintals of teff powder.

When non-food items or contaminants have a similar physicochemical composition to the food itself, such as in the case of bread, it becomes more challenging to detect food impurities. This is corroborated by the findings of A. Choudhary *et al.* [1]. To overcome this difficulty, it is straightforward to assume that the development of food impurity detection model becomes essential B. S. Advisor *et al.*[13]. This will also help in the prevention of further food fraud acts that may spread and put consumers at risk.

Previous research has primarily focused on analysing the texture and colour features of food images using traditional methods, which alone may not provide sufficient characterization of the items. To obtain a more comprehensive and discriminatory representation of food images, it may be essential to incorporate other local binary pattern descriptors of local features, in addition to texture and colour features, as suggested in [9]. In Local Binary Pattern, the study of local structures (similar to structural approaches) and the analysis of occurrences (similar to statistical methods) are combined [9]. In the context of Ethiopian injera, the term "local structure" refers to specific small-scale physical features and texture patterns observed in injera bread. These include eyes/holes, which are tiny pores or holes scattered across the injera surface, as well as cell walls and other significant elements that contribute to its regional structure. The structure between the eyes is made up of the thin cell walls, which also give it its spongy texture and colour.

Most existing methods for food recognition directly extract deep visual features using convolutional neural networks or traditional machine learning. While individual handcrafted features or combinations of handcrafted features with deep learning features may contribute valuable information for the deep learning model decision-making, there is a need for a reliable approach to combine multiple traditional machine learning features with deep learning features to enhance overall accuracy [14], S. Jiang et al. [15].

Despite the work of many researchers on food recognition, detection, and classification, past studies have mostly relied on traditional machine learning algorithms without utilizing image enhancement techniques. Furthermore, the datasets had insufficient class similarity and failed to quantify the proportion of impurities, making them unreliable for real-life testing. Therefore, this study aims to address these gaps by employing deep learning-based feature fusion methods that combine local image features, texture features, and features automatically extracted by convolutional neural networks.

1.4 Research Questions

The study aims to address the following research questions:

- i. Which approach for image classification is more accurate, using deep learning algorithms alone or combining them with handcrafted features?
- ii. How do pretrained models perform compared to custom-built models when combined with handcrafted features?
- iii. How do the performance metrics (accuracy, precision, recall, F1-score) of various deep learning models compare when classifying injera food images?

1.5 Objective of the Study

1.5.1 General Objective

The overall objective of this study is to develop a deep learning model capable of classifying injera samples based on the presence of impurities.

1.5.2 Specific Objective

To accomplish the study's primary objective, the following specific objectives have been formulated:

- To conduct a thorough review of the scholarly literature related to food detection and classification using computer vision and deep learning techniques.

- To prepare an injera dataset according to the specified combination ratio and carry out pre-processing operations.
- To examine how discriminative hand-crafted features affect the accuracy of injera classification based on impurities.
- To build models for injera classification based on impurities.
- To evaluate the performance of injera classification models based on impurities.

1.6 Scope and Limitation of the Study

In this research, we have focused on building models for classifying injera into pure white teff injera, teff with sawdust injera, pure red teff injera, and red teff flour mixed with sorghum (zengada) flour injera categories. However, this study does not extend to examining the types of teff and sorghum. Additionally, it does not address other mixture such as mold, rice and Jasso in injera.

1.7 Significance of the Study

Impurities in food can lead to various health problems threatening to society after consumption and is seen as one of the major problems of daily life. Thus, altered foods can be cost-effective for the provider, but even though it is inexpensive for the consumer, it leads to health as well as life-threatening issues. When the problem is detected, acceptance of foods in the market decreases, which also leads to economic losses for the supplier as well as for the country.

Celiac disease, a chronic digestive and immunological illness, can cause harm to the small intestine when gluten is consumed. The protein known as gluten is present in wheat and barley. Celiac patients have a difficult time absorbing nutrients due to gluten damage to the small intestine's lining [16]. Teff is usually recommended as the best grain for celiac disease patients since it is nutrient-rich and gluten-free [16]. teff-based foods are expected to have outstanding contributions to the prevention of diabetes and anaemia which are occurred due to deficiency of dietary iron throughout the world, Hence, dietary iron intake through the consumption of naturally iron-rich food such as teff injera can be considered as a more efficient and safer strategy E. Tsalera *et al.*[17].

The primary benefit of this study is to create a model for detecting Injera impurities that will aid customers and suppliers and since deep learning approaches face significant challenges in lake of data, the data set we will gather will be utilized by other researchers working in this area for further improvement. The creation of a reliable method for classifying food images

can also has the potential to have broad uses in the food business, quality assurance, and food safety.

1.8 Evaluation and Testing Metrics

The accurate classification of food images into the appropriate categories has a significant impact on the effectiveness of food identification models. Therefore, in order to thoroughly assess the accuracy of the food identification models, the use of a confusion matrix and various evaluation metrics is essential. Specifically, metrics such as Accuracy, Precision, Recall, and F1-score, computed based on the correctly and incorrectly classified samples in the confusion matrix, provide a comprehensive evaluation of the model's performance.

1.9 Model Development Tools

Software tools play a crucial role in implementing deep learning and machine learning tasks. Among the popular choices, TensorFlow, Keras, and PyTorch are widely preferred by data scientists and beginners. These platforms are highly regarded for their support of the Python programming language, which is extensively utilized in deep learning applications [18]. TensorFlow, an open-source deep learning library developed by Google, offers versatility across various platforms, including Windows, Mac operating systems (macOS), Linux, cloud services, and mobile/embedded platforms [18]. It facilitates data pre-processing, model construction, training, and estimation, and incorporates Tensor Board for visualizing computational processes. Keras, a high-level neural network Application Programming Interface (API), seamlessly integrates with TensorFlow, the Microsoft Cognitive Toolkit. It provides a user-friendly interface for constructing and training deep learning models, supporting Central Processing Unit (CPU) and graphics processing unit (GPU) computing, and offering comprehensive support for Convolutional Neural Network [19]. Additionally, Microsoft Word serves as a popular word processor, allowing users to create high-quality documents, reports, letters, and applications. It includes functions such as spell and grammar checkers for accurate and well-presented written information. Mendeley, a desktop software, streamlines citation of papers in various formats and functions as a portable document format (PDF) viewer. It automatically fills in citation data, saving time and effort for researchers, and offers a plugin that syncs with Microsoft Word for improved citation and reference management. In terms of hardware tools, we utilized the Galaxy M13 camera to capture the injera image. This camera features a triple rear camera system, which includes a 50 megapixel (MP) main camera, a 5MP ultra-wide camera, and a 2MP depth camera. The main camera has

a resolution of 50.0 MP and an F1.8 aperture. For the implementation of the research, a personal computer was used. The computer type is a laptop, specifically an HP Notebook, running Microsoft Windows 10 Pro as the operating system. The processor is an Intel(R) Core (TM) i7-8650U CPU@ 1.9 gigahertz (GHz), with a system type of 64-bit operating system and an x64-based processor. The computer is equipped with 16.00 gigabyte (GB) of installed random access memory (RAM) and has storage disks totalling 474 GB.



Figure 1- 1 Smart phone used for Image capturing [4]

1.10 Organization of the Thesis

The thesis is organized into five chapters. Chapter one provides an introduction, background, significance, and research objectives. Chapter two provides a thorough literature review, examining previous studies from both local and international sources related to injera and food production, impurities, and quality control. Chapter three focuses on the design and methodology employed in the research, including details of data collection, image processing techniques, and deep learning approaches used for impurity detection. Chapter four presents the findings, results, and data analysis derived from the research. In conclusion, chapter five covers the conclusions derived from the study, as well as suggestions for future research and practices in the field.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 Introduction

This section reviews existing literature on food classification employing diverse methods and algorithms, alongside discussing image processing, feature extraction, feature fusion, and deep learning techniques utilized in detecting impurities within food using images. Due to limited research on Ethiopian injera impurity detection, we gathered relevant documents from local sources and academic repositories like Google Scholar, IEEE, ScienceDirect, Springer, and ResearchGate to support the development of an injera impurity detection model.

2.2 Fundamentals of Digital Image Processing

The primary objective of digital image processing, as stated by A. M. Hambal *et al.*[20], [21], is to enhance the quality of images or extract valuable information from images obtained from different sources, including static objects, photographs, or video frames. Image processing encompasses three levels of processing: low-level processing, intermediate-level processing, and high-level processing, as depicted in Figure 2-1. The selection of appropriate processing techniques greatly impacts the accuracy of the results A. M. Hambal *et al.*[20], [22].

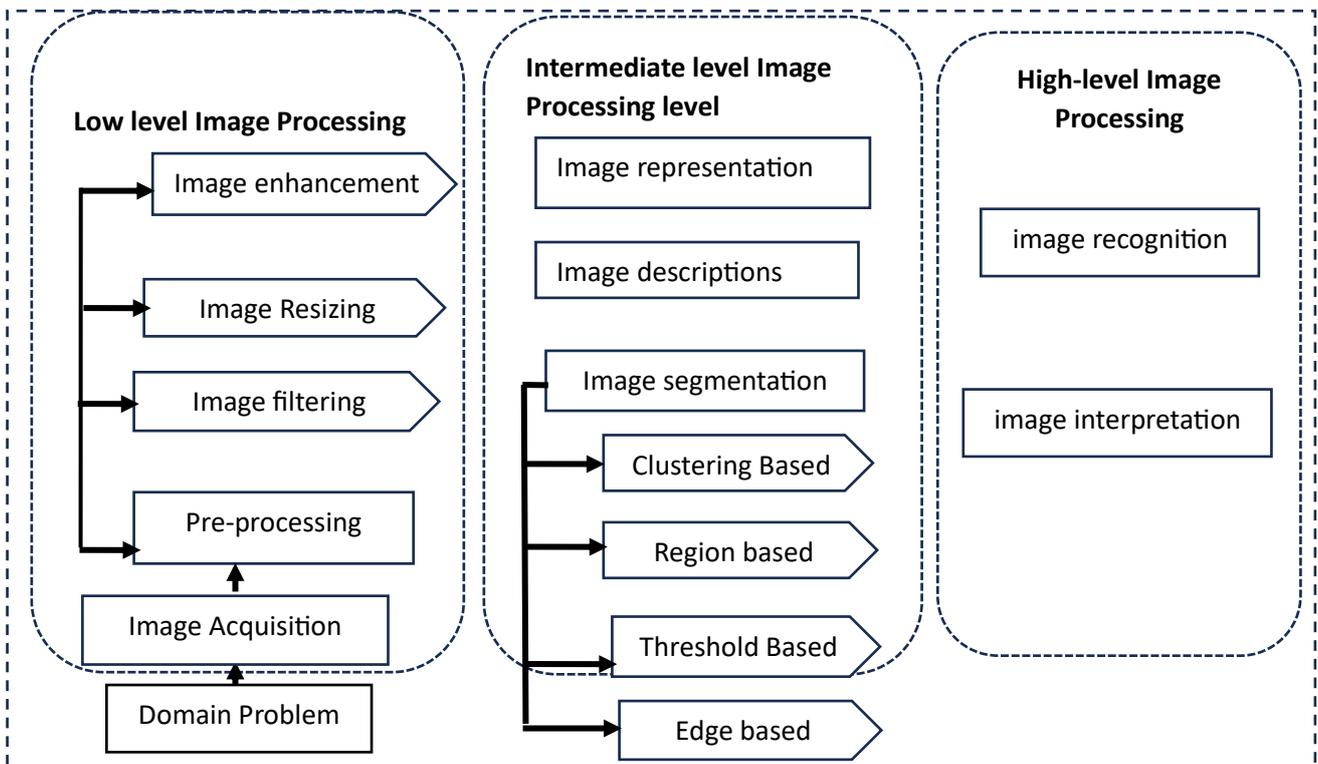


Figure 2- 1 Different levels in image processing process

2.2.1 Low-level Image processing

2.2.1.1 Image acquisition

Image acquisition is the first and most fundamental stage in every image processing activity. It involves taking pictures with cameras, scanners, or other imaging equipment. It may be done in a variety of settings, including indoor, outdoor, and low-light ones, which might have an impact on the images' utility and quality. After gathering the necessary images, the input image undergoes pre-processing to enhance its quality before being fed into the model A. M. Hambal *et al.*[30], [31] ,[32].

2.2.1.2 Image Pre-processing

The pre-processing steps have a significant impact on the classification of images using deep learning, according to research by C. Swathi *et al.* [23]. It involves removing low-frequency background noise, controlling brightness and contrast variations in the image, and normalizing the intensity of the individual particle images to enhance convolutional neural networks performance A. M. Hambal *et al.*[30], [31] ,[32], C. Swathi *et al.* [33], S. Sonawane *et al.* [34].

Pre-processing techniques help improving image quality, enhancing contrast, reduce unwanted variations and make suitability of input images by applying various pre-processing steps such as image enhancement, noise reduction, image resizing, and colour space conversion, making them more suitable for subsequent analysis and classification S. Sridevy *et al.*[25].

A. Image Filtering

Image filtering is a technique used to enhance or remove certain features from the degraded image. The most common cause of image degradation occurs during the acquisition such as sensor noise, camera-misfocus, relative object-camera motion, random atmospheric disturbance. random variation of brightness or transmission of the image which makes the image usually blur or noisy R. Al-taie *et al.*[26], A. M. Hambal *et al.*[30], [31] . The majorly employed filtering techniques includes, Gaussian Filter, Mean Filter, Median Filter, Wiener filter and wavelet-based technique [28],[31],[32].

Gaussian Filter: It executes the average value of neighbouring pixels based on the Gaussian function. The paper of [32], [34], [37] describe that the Gaussian smoothing is very effective for removing Gaussian noise and speckle noise model, However, Gaussian filtering blurs images, take time and reduces fine details while removing noise.

Median Filter: In this filter, each pixel in the image will be replaced with the median of adjacent pixels for each pixel that is being evaluated. The median filter has the advantage of

being simple to use and fast, to keep edges while reducing noise and having good performance in filtering out salt and pepper noise R. Al-taie *et al.*[29], C. Swathi *et al.* [33], [35], [37],[39]. The benefit of the median filter over mean filters is that the median filter can remove the effect of input noise values with huge magnitudes [28], [29]. However, its disadvantages are it remove both noises and image details [23], [27].

Mean filter: is often known as the linear filter which can be done by calculate the average value of the image with noise in a predefined area and the centre pixel intensity value is then changed by average value of pixels in the neighbourhood. This process is repeated for all pixel values in the entire image and it is better with speckle noise [28], R. Al-taie *et al.* [29].

Wiener filter: is an effective filter for removing unwanted noise and blur, but it is not appropriate for images with a lot of edges. The filtered image takes longer to process and has less visual quality [13],[14],[15].

Wavelet: Wavelet-based techniques can achieve good noise reduction while preserving important image details. However, the drawback of these methods is the complex calculations involved, which leads to high time consumption. Additionally, wavelet-based approaches do not provide accurate information about the analysed surface [15].

B. Image Resizing

Maintaining uniform dimensions across the dataset is made easier by resizing the acquired raw images to the same recommended dimensions. According to S. Sridevy *et al.*[25], certain classification algorithms, particularly those based on deep learning architectures, have input size limits or call for standard dimensions. In such circumstances, image resizing techniques can be applied to transform the image to the desired sizes. Resizing also reduces storage and improves image visibility by reducing noise, making manual identification simpler. However, there are also limitations to resizing. Important information may be lost and distortions like blurring, jagged edges, or pixelation may happen if the aspect ratio of the primary images is not maintained. Resizing low-quality images can enhance noise, artifacts, and other distortions in the intensity values of the image, therefore the original image quality must also be considered.

There are several techniques available for resizing images, such as bilinear interpolation, Lanczos interpolation, nearest neighbor interpolation, and bicubic interpolation. Each of these techniques offers its own advantages and limitations when it comes to effectively resizing images S.Sridevy *et al.*[25].

Nearest Neighbour- Of all the interpolation methods, this one is the simplest and takes the least amount of time to process S. Sridevy *et al.*[25], [30].

Bilinear interpolation- Bilinear interpolation employs a weighted average of four neighbouring pixels to determine the interpolated value, yielding superior outcomes compared to nearest neighbour interpolation while demanding less computational resources than bicubic interpolation [30], [31].

Bicubic Interpolation- Bicubic interpolation takes an additional step compared to bilinear interpolation by considering the nearest 4x4 neighbourhood of known pixels, totalling 16 pixels. As these pixels are at different distances from the unknown pixel, the closer pixels are given greater importance in the calculation through higher weights. Bicubic interpolation results in significantly sharper images compared to the preceding methods and strikes a balance between processing time and output quality, making it an optimal choice [31], [32].

Lanczos interpolation- has the best properties in terms of detail preservation and minimal generation of aliasing artifacts for geometric transformations not involving strong down sampling [31], [32].

C. Image Enhancement

Image enhancement improves image clarity and readability for specific applications by correcting contrast, brightness, and reducing noise. It aims to enhance visual appeal by sharpening, reducing noise, and brightening images affected by different imaging devices S. Sonawane *et al.*[34], [43], [33], [45].

Image enhancement involves adjusting contrast, saturation, and sharpness. Contrast affects the distinction between dark and bright areas, while increasing it intensifies shadows and highlights. Saturation widens the gap between shadows and highlights. Sharpness enhances edge contrast. A common technique used to improve contrast is histogram equalization, which redistributes the pixel intensities to achieve a greater dynamic range. Different methods of histogram equalization offer various trade-offs. Selecting appropriate enhancement techniques based on image properties and application requirements is crucial for effectively improving contrast, saturation, and sharpness for subsequent analysis [25],[40], [31],[32] [43].

Classical Histogram Equalization (CHE): is a popular traditional method for enhancing contrast in image processing, particularly for grayscale images, due to its simplicity and low CPU usage. It works by uniformly distributing Gray levels across the intensity range, thereby

improving overall image contrast. However, a drawback is that it can significantly alter image brightness, leading to over-saturation with very bright or dark values. To address this issue, adaptive histogram equalization (AHE) is employed[34].

Adaptive Histogram Equalization: differs from traditional methods by adjusting specific parts of an image rather than the entire image at once. It enhances contrast by interpolating regions using a bilinear approach, redistributing brightness values to improve local contrast and reveal fine details. However, AHE tends to amplify noise in homogeneous areas and may not preserve overall brightness accurately. To address these issues, Contrast Limited Adaptive Histogram Equalization (CLAHE), a more advanced variant of AHE, was proposed by [32], [33], [35], [36].

Contrast Limited Adaptive Histogram Equalization: It is an improved form of AHE that prevents noise amplification. This technique operates on the idea that the image is separated into a number of sections, and the enhancement on each area is examined separately. The total outcome of the component regarded as a whole is the resulting component [37].

Dynamic Histogram Equalization (DHE): it improves contrast by partitioning the histogram into sub-histograms based on local minima and then applying histogram equalization to each sub-histogram. However, it may compress low histogram components, leading to some loss of information [37].

Multi-Histogram Equalization: Multi-histogram equalization decomposes the image into sub-images and applies histogram equalization to each sub-image. It improves image contrast while preserving brightness and generating natural-looking images. But it requires additional computational resources due to the decomposition process [37].

2.2.2 Intermediate level Image processing

Image segmentation, image representation, and image description are all part of intermediate level processing, as described by L. Zhu and colleagues [38].

2.2.2.1 Image Segmentation

One of the most important processes in image processing is image segmentation since it greatly determines whether image analysis concentrates on the target sample. It is the technique of breaking a digital image up into several pieces or divisions. Image segmentation performs the function of separating the desired target from irrelevant image information, lowering the computing cost of subsequent image analysis and boosting the relevance of the images for

classification with better accuracy L. Zhu *et al.*[50], R. Sarki *et al.* [52]. According to [40]–[42], the four approaches of image segmentation are threshold-based, edge-based, region-based, and cluster-based.

Region based: region-by-region segmentation of an image (small groups of connected pixels having similar properties). In noisy images where boundaries are hard to see, region-based approaches often perform better. Operations are quick and calculations are straightforward and works well for high contrast images [40], [43].

Threshold Based: is a straightforward method that divides an image into object and background based on specific threshold criteria. It creates uniform regions within the image by classifying pixels within a certain intensity range into one category, while the remaining pixels belong to the other category. Despite its simplicity, thresholding can be computationally intensive S. Saxena *et al.*[43], [44].

Edge based: The method here begins by identifying the image edges. Since the edges are not completely connected to one another and there are some gaps between them, these edges are then joined to create the border of the regions of interest in order to further segment the regions L. Zhu *et al.* [38]. It aids in maintaining Gray tones in edges and for images with great contrast, but it is inappropriate if edges are numerous S. Saxena *et al.*[43].

Clustering Based: Clustering is the technique of putting things in groups based on characteristics so that each cluster has related objects that are different from the objects in other clusters [40], [42], [43]. Effectively produces outstanding clusters from very little datasets S. Saxena *et al.*[43].

2.2.3 Image Representation and Descriptions

2.2.3.1 Feature Extraction

Feature extraction is an important step in the construction of any classification and aims at the extraction of the relevant information like colour, shape, and texture of images that characterizes each class to perform the desired task [45]–[47]. Feature extraction enhances the accuracy of the entire image classification process, making it crucial to select the most appropriate feature selection approach for precise results. This process involves extracting relevant features from objects to form feature vectors. Classifiers then use these feature vectors to recognize the input unit and match it with the target output unit. By examining these features, classifiers can more easily distinguish between different classes.

Deep visual features and handcrafted features are the two primary categories for feature extraction approaches.

A. Handcrafted Features extraction

The human visual system often uses criteria such as colour, shape, and texture to determine an image's content. Therefore, in computer vision, extracting texture, colour, and shape characteristics is crucial [14], [45], [48]

A1.Texture-Based Handcrafted Feature Extraction

In the image processing, texture represents the surface and the structure of an object that can be defined as a function of spatial variation of the brightness intensity of the pixels. [45], [48]. According to [12], [14], F. Marpaung et al. [61], several methodologies are commonly utilized for extracting different texture features through texture analysis. These methodologies encompass statistical methods, structural methods, and filter-based approaches.

Statistical-based approaches- Within the realm of computer vision, statistical texture analysis methods have long been relied upon for effectively characterizing textures in images. These techniques are, among the earliest techniques for image texture analysis, and have been widely used to complete a various task [14], A. Ramola et al. [50].

In statistical-based approaches, the classification into first-order, second-order, and high-order statistics are three categories depending on the number of pixels in the local features in the statistical-based approach A. Ramola et al. [50]. From the viewpoint of human visual sight, first-order statistics do not offer adequate information A. Ramola *et al.*[50], L. O. Osapoetra *et al.*[51], [64]. Second-order statistics use the neighbourhood relationship between a pixel of interest (PoI) and their neighbourhood pixel C. Reischauer *et al.*[52]. The human beings are quite sensitive to second-order statistics as they provide sufficient information from the point of human visual perception which is the limitation of first-order statistics A. Ramola *et al.*[50]. Second-order statistics include GLCM which is popularly used to describing the texture [12], F. Marpaung et al. [61] and is explained below. High-order statistics are not taken into account for image interpretation since they don't offer any information from a spectral and spatial point of view A. Ramola *et al.*[50].

Gray level cooccurrence matrix: is a common second-order A statistically based technique establishes the relationships between two pixels in an image's texture and offers details on the spatial distribution, brightness changes, and descriptions of how surfaces are arranged structurally [57], W. M. Thu et al. [65]. The following list includes a number of texture features

that were retrieved from the GLCM, including contrast, correlation, energy, entropy, and homogeneity.

Contrast- Contrast evaluates the intensity of a pixel and its neighbour over the whole image, and it is at its lowest when the pixels have the same value for the greyscale [54].

$$Contrast = \sum_{i,j=0}^N |i - j| 2p(i, j) \text{-----Equation 2- 1}$$

Correlation – It quantifies the similarity patterns between a pixel and its neighbouring pixels. By examining it, one can understand how pixel values change in relation to their neighbours. This measure is useful in image processing tasks like denoising, edge detection, and texture analysis, as it provides insights into the spatial relationships and patterns within an image.

$$Correlation = \frac{1}{(\sigma_i \sigma_j)} \sum_{i,j=0}^N (i - \mu_i)(j - \mu_j)p(i, j)\text{-----Equation 2- 2}$$

Entropy- quantifies the level of randomness or uncertainty in the Gray-Level Co-occurrence Matrix (GLCM). It reaches its maximum value when all elements within the GLCM are equal, indicating high randomness in texture properties. Conversely, when the image lacks uniformity in its texture characteristics, the elements of the GLCM attain their minimum values [54]. By calculating entropy, we can gain insights into the complexity and variability of the texture patterns present in the image.

$$Entropy = \sum_{i,j=0}^N p(i, j) \log_2 p(i, j)\text{-----Equation 2- 3}$$

Energy- is the sum of squared elements in the GLCM.

$$Energy = \sum_{i,j=0}^N p^2(i, j) \text{-----Equation 2- 4}$$

Homogeneity - Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. A high homogeneity value indicates that the Gray-level values of adjacent pixels within the image are similar, while a low homogeneity value indicates that the Gray-level values of neighbouring pixels are different.

$$Homogeneity = \sum_{i=0}^N \sum_{j=0}^N \frac{p(i, j)}{1+|i-j|} \text{-----Equation 2- 5}$$

Where, $P_{i,j}$ = is the Gray level matrix element that represents the probability of having neighbouring pixels with intensities i and j in the image and N = denotes the number of Gray levels.

Local Binary Pattern: The Local Binary Pattern (LBP) technique encodes the local texture information of an image by comparing the intensity values of a central pixel with its surrounding pixels. The resulting binary pattern is then used to extract texture features like uniformity, entropy, and histogram patterns A. Ramola *et al.*[50], H. Alshazly et al. [67]. It has drawn the attention of many academics due to its simplicity, speed, and computational efficiency, capacity to capture both global and local texture information, and robustness to variations in illumination F. Marpaung *et al.*[49]. In LBP, the study of local structures (similar to structural approaches) and the analysis of occurrences (similar to statistical methods) are combined [9].

The term "local structure" of Ethiopian injera describes the specific, small-scale physical features and texture patterns seen in injera bread. Eyes/holes, which are tiny pores or holes dispersed across the injera surface, cell walls, and other important elements of the regional structure. The structure between the eyes is made up of the thin cell walls, which also give it its spongy texture and colour.

The LBP operator, which produces 2^p numbers of output values, is denoted as $LBP(L, r)$, Here, L denotes the number of nearby points and r the radius of the circle created by neighboring points. The LBP code is calculated using Equation (2-6).

$$LBP_{P,R}(q_c) = \sum_{p=0}^{P-1} S(q_c - q_p) 2^p \text{-----Equation 2- 6}$$

q_c is the Gray level value of the central pixel and " q_p " is the gray level value of the neighboring pixel. P is the number of neighbouring points and R is the radius of the circle formed by neighbouring points) that produced 2^p a number of output values.

$S(q_c - q_p)$ is a threshold function.

$S(q_c - q_p) = 1, z \geq 0 ; 0, z < 0$ The advantage of LBP techniques is that they combine structural and statistical methodologies, which improves texture analysis performance. In order to produce more discriminant features and greater accuracy, several texture operators are frequently combined [14].

2.2.4 High level Image processing

High level processing includes image recognition and image interpretation. This step could include matching each individually segmented region with the features determined in the

preceding stage, which might aid in the interpretation of the information extracted from the image using techniques like neural networks L. Zhu *et al.*[38].

B. Feature Extraction with Deep Learning

While the technology of acquiring images by computer has advanced recently, classical computer vision is still ineffective for processing vast amounts of image information. Traditional computer vision techniques, for instance, are not very accurate in spotting both low-level and high-level characteristics in images Q. Lv *et al.*[68], P. Song *et al.* [69]. Therefore, it is vital to the proposed model, which aims to offer the basis for image classification and recognition by combining standard approaches for classifying images with feature extraction from deep learning.

2.3 Neural Networks

It is a collection of machine learning algorithms that imitates how the human brain evaluates and processes data to attempt to identify correlations in a set of data R. Dastres *et al.*[58]. A neural network consists of an arbitrary number of neurons that connect the input set to the output. The term "neural" is inspired by the structure of biological neurons. In a neural network, a hidden layer can be seen as an array of parallel-operating units, each acting as a vector-to-scalar function. Each unit receives inputs from multiple preceding units and computes a single output value, which becomes an input for the next layer.

For example, the defined function f is a composition of several functions:

$$f_i = a(\sum_{i=1}^n w_i x_i + w_0) \text{ -----Equation 2- 7}$$

Where, x_i =inputs to the neuron

n = total number of inputs w_i = weight associated with each input

w_0 = a bias term a = is the activation function

Each of these intermediate functions are called layers, so f_1 is the first layer or input layer, f_2 is the second layer and f_n is the last layer or the output layer. The layers that are not the first nor the last, are called hidden layers because we do not interact directly with them and the middle-outputs and weights are hidden to us. Each of the layers has its own parameters array w_i , also known as weights and biases. A neural network with more than one hidden layer is called deep neural network.

There are two types of neural networks basic neural networks and deep learning neural networks as shown in Figure 2- 2. A basic neural network typically includes an input layer, a hidden layer, and an output layer. Deep learning neural networks, however, are characterized by having multiple hidden layers stacked on top of each other. These networks are used in our research because they can learn hierarchical representations of data by gradually extracting more abstract properties as the input moves through the layers. Deep learning neural networks have gained significant attention and popularity due to their ability to handle complex problems and their superior performance in various domains such as image recognition, natural language processing, and speech recognition Q. Lv *et al.*[56]. Two adjacent layers of neurons are connected to each other, as shown in Figure 2-2 a.

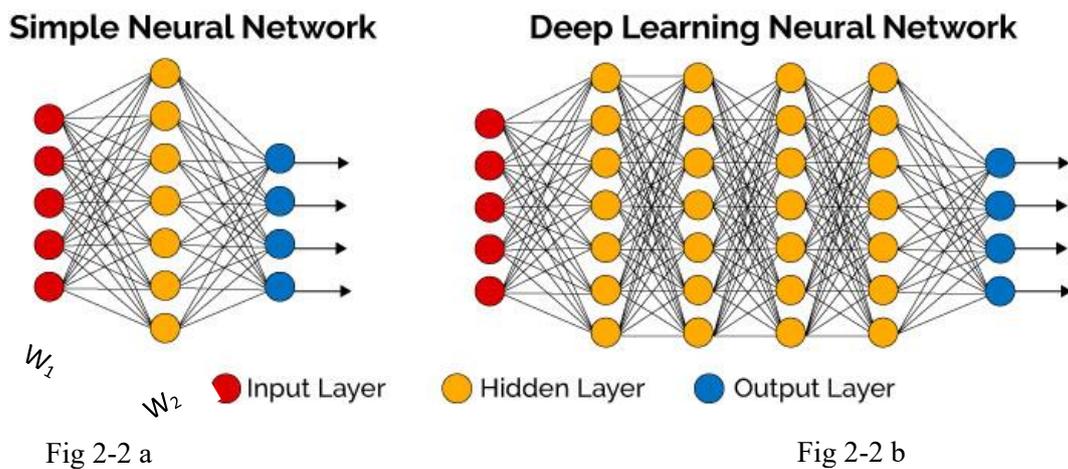


Figure 2- 2 The architecture of Deep learning technologies R. Dastres et al. [59]

2.4 Deep Learning and Food Image Classification

As presented in figure 2-3 deep learning is a subfield of artificial intelligence (AI) and machine learning that involves the use of neural networks which are a type of algorithm loosely based on the structure and function of the human brain. Deep learning algorithms attempt to learn high-level features from mass data, which make deep learning beyond traditional machine learning [59]. Without the involvement of the people, deep learning uses networks of artificial neurons to examine enormous datasets in order to automatically find underlying patterns (meaningful representations) R. Dastres *et al.*[58], [60].

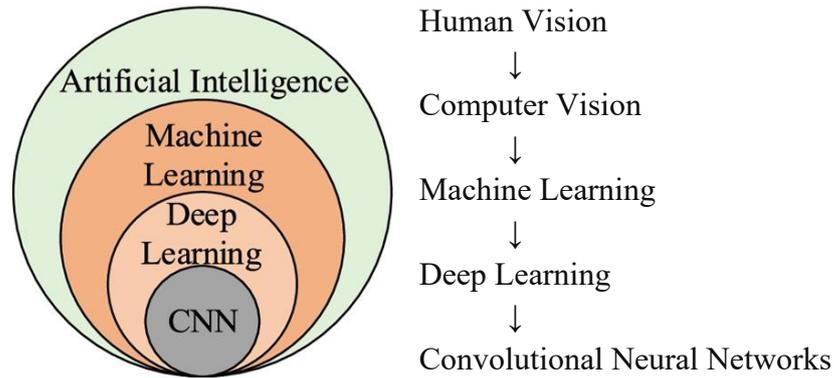


Figure 2- 3 Relationship between Human vision, CV, ML, DL, and CNN

Deep Learning (DL) techniques can be broadly categorized into three main groups: supervised learning, unsupervised learning, and hybrid learning, which combine both approaches [61]. We will describe the highlight of each of those methods and show how they may be used to solve practical problems as follows.

I. Supervised Learning

In classification applications, this class of deep learning algorithms is employed to perform a discriminative function. Overall, discriminative deep architectures use observed input data to estimate the probability distributions of various classes R. Wu *et al.*[74], R. K. Mishra *et al.* [75], Y. Guo *et al.* [76], A. Mosavi *et al.* [77]. Multi-Layer Perceptron (MLP), Convolutional Neural Networks, Recurrent Neural Networks (RNN), and their variations are the major types of discriminative architectures. These methods are briefly discussed as follows.

A. Multi-Layer Perceptron

The basic building block of deep neural networks (DNNs), also known as deep learning, is the feedforward artificial neural network known as the MLP. Typically, an MLP has an input layer that receives data, an output layer that processes the data and generates predictions or decisions, and one or more hidden layers that learn intricate representations and extract relevant characteristics from the input data [61]. One common supervised learning method used to train the MLP is the "Backpropagation" algorithm, which is an important component of neural networks [61], Y. Guo *et al.*[64]. When dealing with large models, optimizing the hyperparameters of the MLP like the number of hidden layers, neurons, and iterations can pose significant computational challenges [61], Y. Guo *et al.*[64].

B. Recurrent Neural Network

Another well-known deep learning architecture is the recurrent neural network, which uses sequential or time-series data and feeds the result of the preceding stage as input to the current

one[66]. Like CNNs and MLPs, recurrent neural networks (RNNs) learn from training data, but they are distinguished by their "memory," which allows them to use information from previous inputs to influence current input and output. However, RNNs face the "vanishing gradients" problem when dealing with long sequences, as they struggle to retain long-term information and primarily focus on the most recent data. This issue was addressed with the development of Long Short-Term Memory (LSTMs) and Gated Recurrent Units (GRUs), which have internal gates that control information flow [73], A. Mosavi et al. [77],[79].

Long short-term memory: is a well-known RNN architecture which was presented by Juergen Schmid Huber and Sepp Hochreiter as a solution for the vanishing gradient problem that affect the training of standard RNNs [68]. The model is designed to solve the problem of long-term dependencies by using a specific kind of memory cell that allows it to store inputs for longer periods. These memory cells use a system of "gates" to regulate the flow of information into and out of the cell, enabling them to retain information for extended periods. An LSTM (Long Short-Term Memory) unit has three main components: the input gate, output gate, and forget gate. These gates control the flow of information into and out of the memory cell, giving the LSTM the capacity to retain data for extended periods [61]. The LSTM network is considered one of the most effective RNNs, as it overcomes the challenges associated with training recurrent networks. Among the common uses of LSTM, image classification is a prominent and widely adopted application [73],[80],[81].

Gated recurrent units: The more recent generation of RNNs, called Gated Recurrent Units (GRUs), is essentially a simpler version of LSTM networks. Unlike LSTMs, GRUs use the hidden state instead of the cell state to regulate and control information flow within the neural network [25]. Though the three gates used by LSTMs to govern the flow of information into the memory cell are replaced with a single gate called update gate in the GRU, it functions similarly to an LSTM. Having a single gate makes GRUs quicker to execute and easier to train than LSTMs, but it's possible that it is not as good at storing and retrieving long-term dependencies J. Chung *et al.*[70]. GRU and LSTM have both demonstrated their effectiveness in achieving the desired result, despite the fact that GRUs have been demonstrated to perform better on some smaller and less frequent datasets J. Chung *et al.*[70].

C. Convolutional Neural Network

A Convolutional Neural Network is another discriminative deep learning architecture that learns directly from the input without the need for human feature extraction. It is a popular

neural network widely used for various image applications, including object identification, face recognition, image recognition, and image categorization A. Mosavi *et al.*[83],[84]. It is more powerful than a conventional network since it can automatically extract important information from the input without the need for human interaction [73], R. K. Mishra *et al.* [75], Y. Guo *et al.* [76],[78].

The convolutional neural networks architecture employs a hierarchical approach, with early layers focusing on extracting low-level features and later layers extracting high-level features from images F. Marpaung *et al.*[49]. Convolutional Neural Networks are particularly well-suited for image processing and time series data, especially for tasks such as image classification and language recognition, unlike traditional machine learning approaches Q. Lv *et al.*[56]. In this thesis, convolutional neural networks are utilized for classifying impurities in injera. Figure 2-4 describes the fundamental structure of a convolutional neural network comprising six distinct types of layers, input, convolutional, pooling, fully connected, and output layers. These layers are partitioned into feature extraction and classification parts. The feature extraction part encompasses the input, convolutional, pooling, and activation layers, while the classification part comprises the fully connected and output layers.

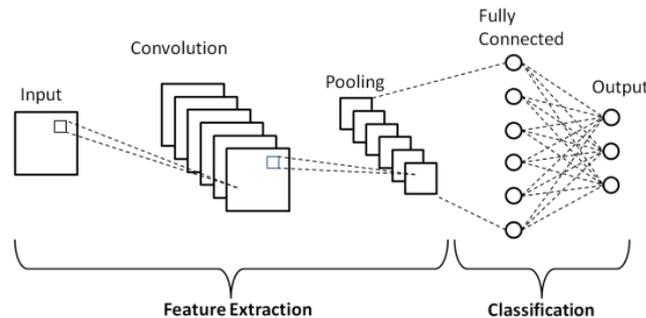


Figure 2- 4 Schematic diagram of a basic CNN architecture [73]

Input Layer: The neural network's input layer serves as its foundation. It takes in the raw input data such as images, text, or audio and sends it to the following layers for processing.

Convolutional Layer: The second and most crucial part of the convolutional neural network architecture is the convolution layer. This layer typically combines both linear and nonlinear operations, specifically the convolution operation and the activation function [10]. When performing a mathematical convolution operation, a small filter is moved over the input to extract useful information from the image. Convolutional layers use a series of filters that they have learned to identify patterns in the input, such as edges, textures, or shapes [10]. Initially,

the filter traverses each pixel of the input image sequentially. At each step, an element-wise multiplication is performed between the filter pixels and the corresponding pixels of the image. The results of these multiplications are then summed up to produce a single output pixel. By repeating this procedure for every pixel, an output value is generated and positioned accordingly in the output tensor, known as a feature map [13], Purwono et al. [86].

Referencing the 5x5 input tensor in Figure 2-5, alongside a 3x3 filter matrix. To generate the convolved feature matrix illustrated in Figure 2-6 below, slide the filter matrix across the image and calculate the dot product at each position. Tensor size, number of kernels, stride (s), which is the filter step, and padding (p), which is a critical hyperparameter, are the four parameters that determine the convolution process [10].

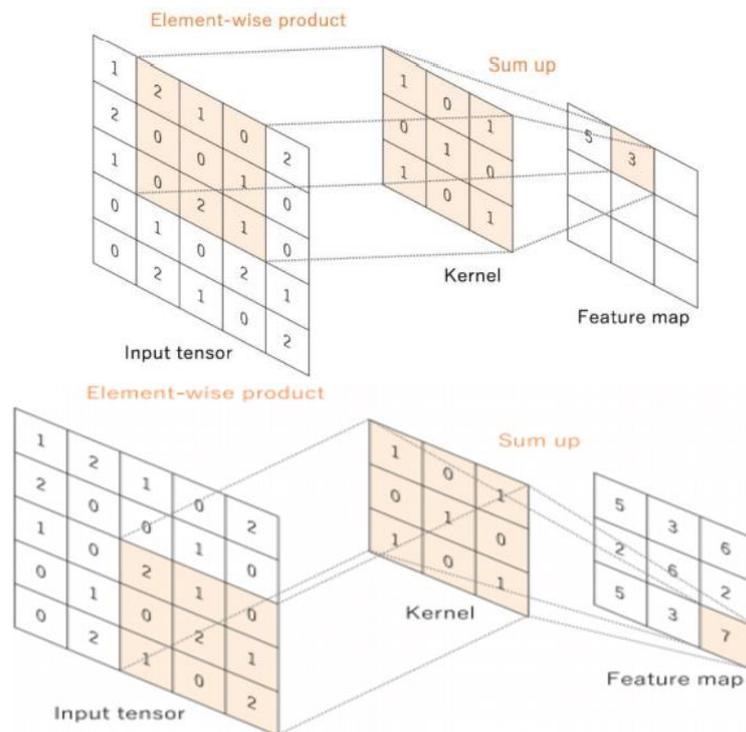


Figure 2- 5 Example of convolution operation

Pixels on the corners and edges of a convolutional layer are used far fewer times than those in the middle. As a result, information on image edges is not as well retained as information in the middle. Zero padding, which increases the input image with rows and columns of zeros, is an effective and straightforward solution to this issue. Padding, typically zero padding, is used to preserve both the information in the middle of images and the information on the edges during the convolution operation. This involves adding rows and columns of zeros on either side of the input image, as shown in Figure 2-6.

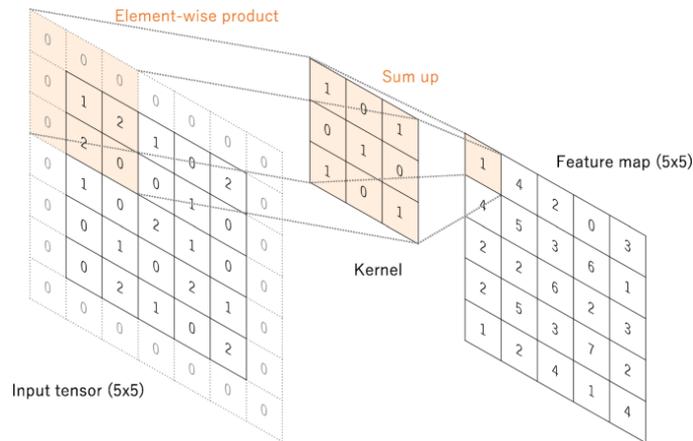


Figure 2- 6 A convolution operation with one zero padding

To compute the output size of a convolutional layer, we consider the following input parameters: an image with dimensions ($W_{ip} \times H_{ip}$), a filter with dimensions ($F \times F$), a stride value (s), and padding (p). The resulting feature map will have the following dimensions:

$$W_{op} = \frac{W_{ip} - F + 2P}{s} + 1 \text{-----Equation 2- 8}$$

$$H_{op} = \frac{H_{ip} - F + 2P}{s} + 1 \text{-----Equation 2- 9}$$

Pooling Layer: The pooling layer, also known as the subsampling layer, follows the convolutional layer in a convolutional neural network. It is used to reduce the computational power needed to process the data and to extract dominant features that are rotationally and positionally invariant, helping to control overfitting [72], Y. Ruf et al. [87], [72], Purwono *et al.* [86], Y. Ruf *et al.*[75], [88]. Common techniques used in this layer include max-pooling, average pooling, and spatial pyramid pooling [76], [77]. Max-pooling selects the maximum value within non-overlapping regions of the input feature map to create a down-sampled feature map, highlighting prominent features. Average pooling computes the average value within these regions, smoothing out features and reducing noise in the down-sampled feature map.

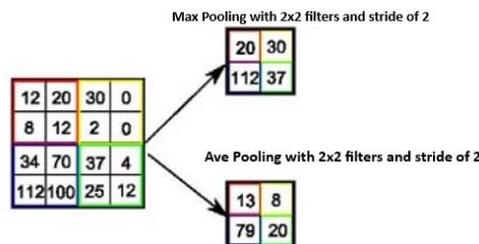


Figure 2- 7 Schematic diagram of pool layer operation type

Fully connected Layer: As the name suggests, each neuron in a fully connected layer is connected to every neuron in the previous layer. Fully connected layers are typically used towards the end of convolutional neural networks to take the features learned by the convolutional and pooling layers and use them for classifying the input into a label [78].

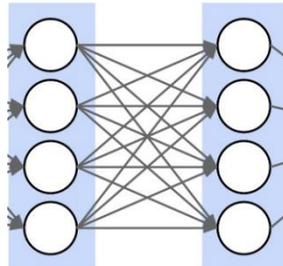


Figure 2- 8 Fully Connected Layer [67]

Activation Layer: These layer uses an activation function to determine the final value of a neuron. These functions enable the model to create a complex mapping between the network's input and output, essential for complex learning and modelling systems [88], [90], Z. Zahisham et al. [91]. Depending on the input, some neurons will be more important than others, so it is beneficial to enhance the output of significant neurons and neglect the output of less important ones. This is achieved by using an activation function after a layer, which involves calculating a weighted sum and adding bias. Understanding typical activation functions in convolutional neural networks and other neural networks helps in selecting the best one to speed up the learning process of the model Purwono *et al.*[74], [79].

The Rectified Linear Unit (ReLU): This function sets negative input values to zero while preserving positive values. As a result, any pixel values in an image that are less than zero are transformed to zero R. Wu *et al.*[74], A. Mosavi et al. [77],[81]. it has a greater classification accuracy than other classifiers' activation functions, and its computing cost is generally lower than that of other functions and it is widely employed in convolutional neural networks [60].

Sigmoid: This activation function restricts the output between zero and one and accepts real numbers as inputs. This implies that a value more than one in the input is changed to one, while a value less than zero is instantly set to zero Purwono *et al.*[74].

SoftMax: One of the deep learning activation functions that is frequently used for decision-making tasks, similar to sigmoid activation function tasks. It normalizes an input vector of arbitrary real values into a probability distribution across the classes.

Output Layer: The output layer is the final layer of the neural network, where each node represents the probability of an input belonging to a specific class. In this scenario, the output layer is configured to classify different types of injera: pure white teff injera, white teff with sawdust injera, pure red teff injera, and red teff with sorghum (zengada) flour injera.

2.4.1 CNN For Food Image Detection and Classification

The efficiency of food image classification has significantly improved with the advent of Convolutional Neural Networks (CNNs). Early methods relied on traditional AI techniques like image processing and machine learning, which required handcrafted features and domain-specific knowledge, limiting their generalization across different food categories. The introduction of deep learning, especially CNNs, revolutionized the field by enabling automatic learning of hierarchical representations from raw pixel data, making them more effective for image analysis tasks.

The researchers D. J. Attokaren *et al.*[81] employed the Food101 dataset with 101,000 images distributed across 101 food categories for training CNN models in food classification. Each category consisted of 750 training and 250 test images, all standardized to 299x299 pixels. Image augmentations such as 45-degree rotations, 0.2 width and height shifts, horizontal flips, and "reflect" fill mode were used to enhance pattern diversity and prediction accuracy. This approach yielded an accuracy score of 86.97%.

The paper of [82], explored two categorization approaches: creating a CNN model from scratch and employing transfer learning with the Inception v3 model. They utilized the Food-11 dataset, comprising 16,643 images across 11 food categories, split into training (9,866 images), validation (3,430 images), and testing sets (3,347 images). The CNN achieved 74.70% accuracy, while transfer learning with Inception v3 achieved 92.86%. Image sizes were 224x224 for the CNN and 299x299 for Inception v3.

The study primarily used CNN for significant information extraction and food image identification, utilizing a dataset of 35,000 images across 35 categories. The dataset was split into 70% for training and 30% for testing. Various feature extraction methods included HSV for color features, Canny for shape features, and GLCM for texture data. Evaluation metrics such as accuracy, precision, recall, and F1-score were employed, with reported ranges of 0.50-0.94 for accuracy, 0.66-1.00 for recall, and 0.59-0.94 for F1-score. The study achieved an overall classification accuracy of 76% F. Marpaung *et al.*[49].

In study of A. Chaitanya *et al.*[83], was proposed CNN as part of the classification scheme. The researchers used the Food-101 dataset, which includes 250 human-reviewed test images and 750 training images. The original 512-pixel-long images were cropped to 299 pixels to reduce noise. The model was trained and evaluated using 20 classes from the testing and training sets, achieving an accuracy of 97.00% for 20 classes and 96.52% for 25 classes.

The study compared CNN, the Google Inception v3 model, and a traditional method using Histogram of Oriented Gradients (HOG) and GLCM with Support Vector Machines (SVM) Y. Li *et al.*[84]. The dataset included 3,600 images of "Yantai Red Fuji" apples and 300 additional test images. Apples were categorized into low, midrange, and premium quality. The CNN model achieved high training and validation accuracies of 99% and 98.98%, respectively, while the traditional HOG/GLCM + SVM method achieved a 78.14% validation accuracy. Original images (3120×4160 pixels) were resized to 208×208 pixels for the models.

II. Unsupervised Learning

Unsupervised learning or generative deep architectures is a deep learning architecture that trains a model on unlabelled data without explicit target or label information. It is used for tasks like clustering, dimensionality reduction, and generative modelling [61].

2.4.2 Transfer Learning

Transfer Learning is a technique that utilizes features from a large dataset neural network model that has already been trained to solve a different but related task using a smaller training dataset [73], C. Tan *et al.*[99],[100]. Transfer learning has become an important deep learning approach to overcome the problem of limited training data. The model can learn effectively from a smaller dataset with fewer training epochs, less training time and resources, and improve the model performance and accuracy since it already has some knowledge relevant to the new job F. M. Shiri *et al.*[85].

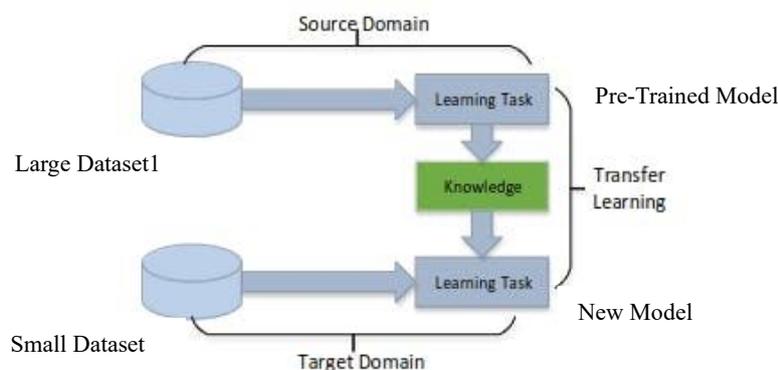


Figure 2- 9 Learning Process of Transfer Learning

The process described in Figure 2-9 above involves transferring knowledge from Dataset 1 to Dataset 2 to improve performance on a related but distinct task C. Tan *et al.*[86]. Initially, the model is pretrained on Dataset 1 to capture general feature representations. These learned features are then adapted for Task 2 by adjusting specific layers of the model using Dataset 2. This approach allows the model to retain knowledge from Task 1 while refining its characteristics to better suit the requirements of Task 2.

AlexNet: After Alex Net’s convolutional neural network architecture won the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) with a top-5 error rate of 15.3 %, it attracted a lot of attention C. Tan *et al.*[86], [87]. The ImageNet dataset, which has millions of labelled images divided into thousands of classes, is used to train the pre-trained AlexNet model. This shows that general image classification tasks are a good fit for these model [59].

ResNet50: ResNet50 (Residual Neural Network) architecture, introduced in 2015, significantly improved the training of deep neural networks. It is trained on large-scale image classification datasets like ImageNet. ResNet50, typically consisting of tens or hundreds of layers, excels in various image classification tasks. Variants such as ResNet-50, ResNet-101, and ResNet-152 are widely used as pre-trained models in computer vision applications. ResNet-50, with its deep layers and residual connections, is particularly effective at learning complex representations and has shown impressive accuracy in image recognition problems. In 2015, it won the ILSVRC with a top-5 error rate of 3.57%[87].

YOLO (You Only Look Once): it is a pre-trained model for object detection, which is the task of locating and identifying objects in an image which is introduced in 2016. Unlike traditional approaches, which perform detection in multiple stages, YOLO performs detection in a single pass and directly predicts bounding boxes and class probabilities. YOLO models are known for their real-time performance and high accuracy in object detection tasks [87].

2.5 Evaluation Techniques

In computational problems such as classification and detection, the goal is to predict the class of an instance. Common evaluation metrics include accuracy, precision, recall, and F1-score. These metrics are derived from the confusion matrix, which compares the model's predictions to the actual classifications F. M. Shiri *et al.*[98], D. Krstinić et al. [101]. Table 2-1 provides a general differentiation of accuracy, precision, F1-score, and recall.

Table 2- 1 Confusion Matrix

		Actual Value	
		Positive	Negative
Predicted Value	Positive	TP	FP
	Negative	FN	TN

True Positives (TP): the model predicted as positive, and the observation are indeed positive.

True Negatives (TN): The model predicted as negative, and observation is indeed negative.

False Positives (FP): The model predicted as positive, but observation is actually negative.

False Negatives (FN): The model predicted as negative, but observation is actually positive.

These four metrics in the confusion matrix are defined and calculated as follow.

Accuracy: The accuracy metric indicates the overall correctness of the model's predictions across all classes of injera. A high accuracy score suggests that the model is making correct classifications for the majority of the injera samples. It helps answer the question: How well does the model classify the different types of injera overall?

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \text{-----Equation 2- 10}$$

Precision: precision evaluates the proportion of correctly predicted positive instances.

$$\text{Precision} = \frac{TP}{TP+FP} \text{-----Equation 2- 11}$$

Recall (Sensitivity): It explains how many of the actual positive cases we were able to predict correctly with our model. It measures the proportion of correctly predicted instances for each class out of all actual instances of that class.

$$\text{Recall} = \frac{TP}{TP+FN} \text{-----Equation 2- 12}$$

F1-score: The F1-score combines precision and recall into a single value, providing a balanced measure of the model's performance of precision and recall for each class.

$$\text{F1-score} = 2 \cdot \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \text{-----Equation 2- 13}$$

Through such comparative evaluation allows us to glean insights into the strengths and weaknesses of each deep learning model to specific datasets and applications.

2.6 Related Works

The section provides a summary of relevant international studies on food detection techniques and local studies specifically addressing injera impurity detection and classification.

2.6.1 Global Studies on Food Detection and Classification

In the study of J. L. Joseph *et al.*[89], utilized a Convolutional Neural Network (CNN) for classification, training it on a dataset of 855 images distributed across 12 classes, with each class containing approximately 70 images. For testing, 20% of the total images were randomly allocated, while the remaining 80% (710 images) were used for training. The CNN achieved an 84% accuracy on the training dataset and demonstrated optimal performance on unseen datasets. Notably, the study did not employ any image pre-processing techniques to enhance or prepare the images before inputting them into the deep learning models.

The study W. M. Thu *et al.*[53] presents a food detection and classification system utilizing image processing techniques, including image pre-processing, segmentation, feature extraction, and classification. Features such as Local Ternary Patterns (LTP), Gray-Level Co-occurrence Matrix (GLCM), and their combination were extracted from the images. These features were then classified using a Support Vector Machine (SVM) classifier, achieving over 90% accuracy. Processing times varied depending on the feature extraction method, with LTP being the fastest. The study standardized input images to 224x224 pixels, resulting in different dimensions for the GLCM and LTP feature vectors. The combined descriptor was more influenced by the larger descriptor, necessitating dimensionality reduction techniques to mitigate the disparity and optimize feature effectiveness in the classification algorithm [90].

The work in R. Wu *et al.*[62] again employs the SURF-Local and Global Colour (SLGC) model to classify images of Asian foods using the UEC FOOD 100 dataset, which includes 14,366 images across 100 categories. The primary focus is on enhancing food identification by integrating both local and global image features. Local texture and gradient information are captured through SURF descriptors, while a global Hue Saturation Value (HSV) color histogram provides broader color information compared to RGB, making it less sensitive to lighting changes and more effective for food classification. However, relying solely on SURF and color features like in SLGC may have limitations, particularly in distinguishing visually similar foods. To address this, future approaches could benefit from capturing a wider range of

features across different modalities. Nonetheless, the study underscores the effectiveness of combining complementary local and global features to improve image classification performance, offering valuable insights for enhancing food identification systems.

In their study, [91] investigated the use of deep convolutional neural networks for both calorie estimation and food recognition. By downloading images from the internet, they compiled a collection of 1316 food images in 13 different categories. This indicates that the images were probably inconsistent and noisy, with different sizes, angles, illumination, etc. The authors used CNN to directly extract visual features from the original raw images rather than performing pre-processing methods. Then they combined features from color histograms, Gabor filters, Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), Bag of SIFT (Scale-Invariant Feature Transform). Their CNN-based technique surpassed conventional food classification algorithms like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) in terms of precision, recall, and accuracy measures.

The study suggests that enhancing the CNN's capacity to detect and accurately estimate calories from food images could be achieved by incorporating appropriate preprocessing steps. Implementing these preprocessing techniques could improve feature extraction, normalize the dataset, and enhance overall model performance in complex food recognition tasks.

In the study of S. Jiang *et al.*[4], proposed a method called multi-scale multi-view feature aggregation (MSMVFA), which combines three types of features extracted from food images: deep visual features, mid-level attribute features, and high-level features. These features are derived from three datasets: VireoFood-172 (110,241 images, 172 categories), ChineseFoodNet (16,000 images, 80 classes), and Food-101 (101,000 images, 101 categories). The MSMVFA method aggregates these features to form a comprehensive representation, which is then classified using a SoftMax classifier. The study finds that combining two or three types of features generally improves food recognition performance compared to using a single type of feature. However, the study highlights potential issues due to the lack of pre-processing methods for mid-level attribute and high-level features, which could lead to noisy data, redundancy, and overfitting.

Furthermore, in the paper titled "Food Image Classification Using Convolutional Neural Network" [92], The authors conducting experiments on the Food-11 dataset, which contains 16,643 images across 11 food categories. Firstly, they developed a CNN model from scratch, achieving an accuracy of 74.7%. This model was trained solely on the Food-11 dataset without

utilizing any pre-trained weights or architectures. Secondly, they employed transfer learning with the Inception V3 model, which was pretrained on a large-scale dataset. This transfer learning approach significantly boosted the accuracy to 92.86% in classifying food images on the same dataset. Notably, the authors did not include any image enhancement methods during the pre-processing phase of their work. Techniques such as contrast enhancement and denoising are typically used to improve image quality and clarity before feeding them into the model.

In another most recently, that investigated the use of deep learning for food classification was conducted by F. Marpaung *et al.*[49]. In this study, researchers utilized CNN to identify food items using color, shape, and texture features. Their dataset comprised 35,000 images across 35 food groups, with 100 images per groups. By combining GLCM with CNN, they achieved 67% accuracy rate. Despite the limited dataset size, the study did not incorporate image enhancement or data augmentation techniques.

2.6.2 Local Studies on Injera Impurities Detection and Classification

The research conducted at Bahrdar University [93] aimed to distinguish between pure teff injera, mixtures of corn with teff injera, and mixtures of Jasso with teff injera. The study utilized the Gray-Level Co-occurrence Matrix as the sole technique for handcrafted feature extraction. The dataset consisted of 600 images, with each food group comprising 100 images from the front side and 100 images from the back side of the injera. The research achieved promising accuracy results, ranging from 87% to 98%, using a combination of 10:90 and 20:80 ratios with Random Forest (RF) and SVM traditional classifiers.

The literature review highlighted that combining multiple handcrafted features in a strategic manner can significantly improve image classification performance compared to using a single handcrafted feature type for machine learning classifiers type. LBP approaches offer an advantage as they combine both structural and statistical methods, leading to enhanced performance in texture analysis [14]. Therefore, in order to obtain more discriminative features and improve accuracy, the combination of LBP with GLCM was employed and compared their performance with deep learning model alone.

Table 2- 2 Summary of related works

Reference	Title	Approach	Dataset	Accuracy	Gap Identified
[89]	Food Classification Using Deep Learning	researchers utilized CNN as their primary approach for classification.	dataset contained 12 classes, each containing 70 images.	84%	did not utilize any image pre-processing techniques
[4]	Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition	deep visual, mid-level attribute and high-level features are extracted and aggregated and feed to SoftMax.	1000 per class		does not use any pre-processing techniques
[62]	An SLGC Model for Asian Food Image Classification	local and global features using SURF and colour histogram respectively and both	143 per class	64%	relying solely on SURF and colour only has limited discrimination between visually similar foods
[49]	Selection of Food Identification System Using CNN Method	GLCM + CNN feature extraction methods	100 per class	67%	did not use pre-processing like enhancement, augmentation techniques
[53]	GLCM and LTP Based Classification of Food Types	GLCM, LTP and GLCM +LTP for features extraction and SVM for classifier	N/A	90%	Not balancing the feature vector's dimension of the descriptor using

[94]	Food Recognition using combined SURF and Gist Feature	Use Gist and SURF for features extraction and SVM as classifier	300 per class	93.3%	Gist and SURF may not capture enough features handle highly similar images.
[95]	A Deep Convolutional Neural Network for Recognizing Foods	Convolutional Neural Network	100 per class	76%	authors directly extracted visual features from the original images using CNN
[73]	Food Image Classification Using Convolutional Neural Network	transfer learning with the Inception V3 model and CNN model developed from scratch	1500 per class	74.7%	the report accuracy shows the model's performance without benefits of image enhancement

2.7 Summary

In this chapter, the focus is on reviewing current studies related to techniques for recognizing and classifying foods, assessing their effectiveness and limitations. The research identifies gaps, especially in integrating handcrafted features with convolutional neural networks (CNNs) and exploring deep learning methods for food categorization.

To provide a comprehensive understanding of the topic, we include information on essential methods of digital image processing. Additionally, we explore into the fundamentals of various deep learning methods, particularly convolutional neural networks.

The main contribution lies in addressing limitations of previous injera classification research by combining handcrafted features like GLCM and LBP with features extracted from CNNs. This approach includes building a model for classifying variants of injera such as pure white teff flour, white teff flour with sawdust, pure red teff, and mixed red teff with sorghum (zengada) flour injera.

CHAPTER THREE

3. DESIGN AND METHODOLOGY

3.1 Introduction

The research approach used to answer the research questions posed in Chapter One Section 1.3, is the focus of this chapter. Our aim is to develop improved impurity detection systems for Ethiopian injera by combining deep learning and traditional machine learning techniques to extract both deep-learned and hand-crafted features.

There are various processes involved in constructing the model. Initially, we prepare and categorize injera images, classifying them into four different types such as pure white teff flour injera, white teff flour with sawdust injera, pure red teff injera, and mixed red teff flour with sorghum (zengada). Following that, we partition the data into training, validation, and test. To ensure the dataset is ready for feature extraction, we apply pre-processing techniques after acquiring the images. This includes segmenting the images, standardizing their size, and normalizing their pixel values to enhance image quality and eliminate noise.

Next, we extract various features using the Gray-Level Co-occurrence Matrix and Local Binary Patterns (LBP) techniques. GLCM features such as contrast, correlation, homogeneity, and energy allow us to capture important texture information and characteristics. By combining the GLCM and LBP feature vectors, we obtain complementary texture information. Furthermore, we extract deep learning features from a convolutional neural network model, capturing both high-level and detailed visual features.

The fused feature vectors, comprising GLCM, LBP, and CNN features, are subsequently utilized to train various classifiers, including fully connected, long short-term memory (LSTM), AlexNet, ResNet50, and YOLO, for image classification. This involves concatenating and normalizing the feature vectors to construct a fused representation that incorporates information from multiple scales.

Our objective is to develop a robust model that leverages both low-level textures and high-level abstractions, resulting in an enhanced feature representation and compare the accuracy in identifying impurities pure CNN-based approach. To accomplish this, we constructed concatenated features and build a model using CNN and Handcrafted Features Fusion Approach and, we also built a pure CNN-based approach utilizing the same dataset for comparison.

This chapter cover the Proposed System Architecture, Data Preparation, Data Pre-processing (including Image Filtering and Resizing), Image Data Augmentation, Image Enhancement, Image Segmentation, Image Representation and Descriptions (such as Feature Fusion, Deep Feature Extraction, Dimensionality Reduction, and Handcrafted Feature Extraction), and finally, a Summary of the chapter.

3.2 Proposed System Architecture for Detection of Impurities in Injera

This section explains the overall system architecture of injera impurity detection model, which employed a three-phase approach, as shown in the proposed general System architecture of Figure 3-1. The initial phase focused on image data preparation, which involved collecting the images, pre-processing them, and applying image segmentation techniques.

In the second phase, features were extracted through a combination of convolutional neural networks, GLCM, and LBP techniques. The bigger descriptor's dimensions are decreased using the Principal Component Analysis (PCA) technique to guarantee that the impacts of the GLCM, LBP, and convolutional neural networks descriptors on the final image feature descriptor are equal. The intended system is developed using the right tools and processes in the third phase. After that, a specific training dataset is used to train the created model. The model is tested using a different testing dataset once the training procedure is completed to evaluate its performance.

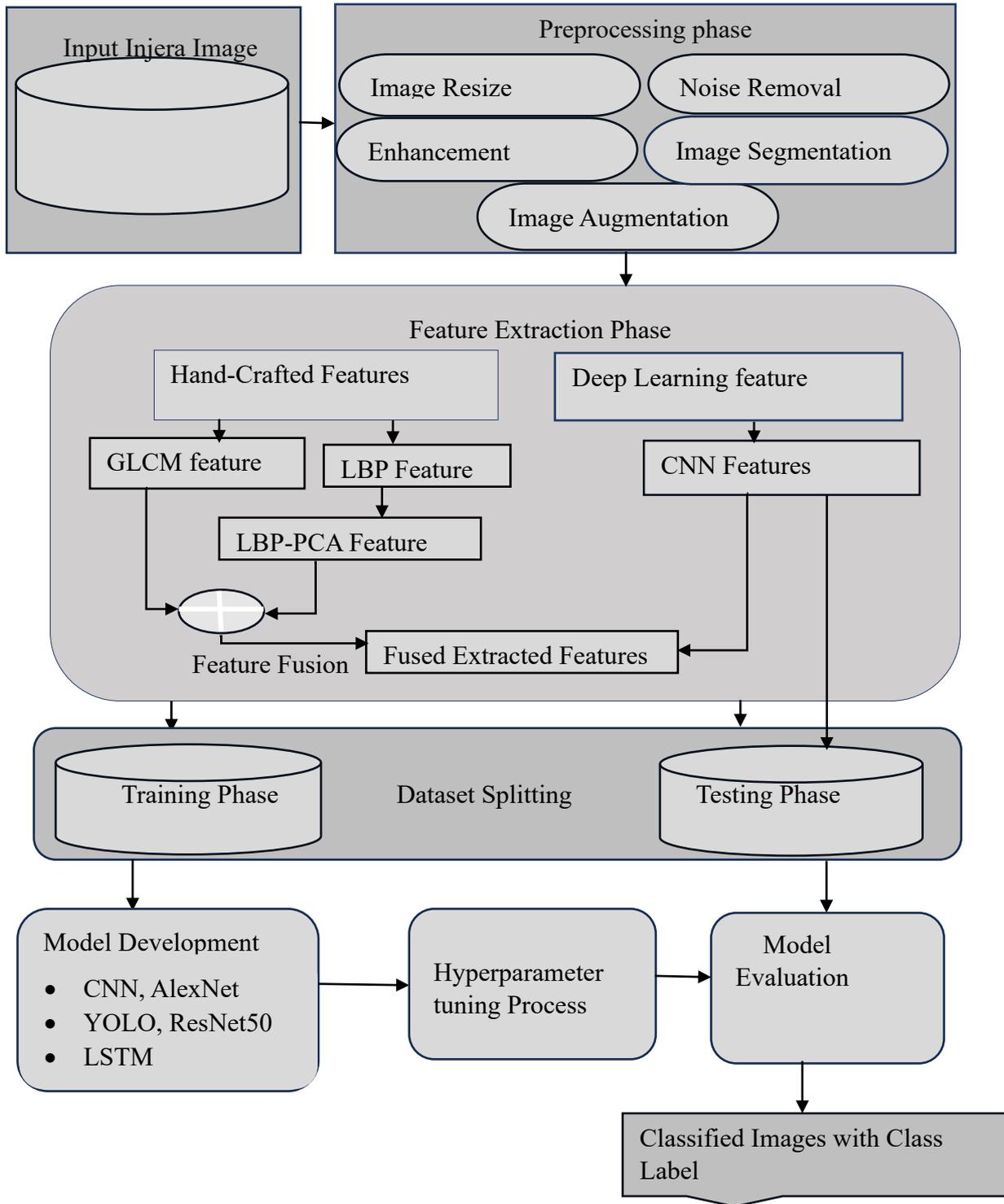


Figure 3- 1 Proposed general System architecture

3.2.1 Data Collection and Preparation

The most crucial step in using machine learning or deep learning algorithms is to collect the necessary data to train the models to recognise, predict, and detect objects. Therefore, it's essential to have adequate data to train the convolutional neural networks model. To classify

Ethiopian injera based on the impurity the required dataset is an image, images of injera prepared at home in the same way as done traditionally in every household. Preparing a large number of impurities injera images can be both costly and challenging. Due to this limitation, 135 images per category for white class and about 102 images per category for red class was captured using handheld mobile device cameras as indicated in figure 3 - 2.

The process involved mixing red or white teff flour with clean water and kneading it by hand in a traditional bowl. We prepared fermented dough for each class with the mixture ratio of 15% to 85 %, then this fermented dough was added to the corresponding flour-water mixture as a starter, and thorough mixing was performed until a homogeneous slurry was achieved [96]. The resulting dough underwent two stages of natural fermentation, lasting approximately 3 days at room temperature.

The surface water that developed on top of the dough was thrown away after primary fermentation. Then, around 10% of the fermenting dough liquid layer is combined with water and cooked for about 15 minutes. It is then combined with the remaining dough and fermented for 2 hours, allowing for the production of more eyes that are bigger [97]. Finally, when the fermenting dough is thin enough to pour onto the hot flat pan, locally known as mitad for steam-baking into injera [97].

White teff flour with sawdust injera and mixed red teff flour with sorghum (zengada) flour injera, which represent the impurity data category, were made using the identical fermentation process as the above. For the sawdust and sorghum (zengada) flour as well as the pure red and white teff flour, we use a mixture ratio of 15% to 85 %, respectively. Finally, we are going to capture the prepared injera for each category by using hand-held mobile devices camera. Since the total amount of data is small the data augmentation was taken to increase the amount of data via different augmentation parameters.

Generating a large number of impurity-laden injera images can be both expensive and challenging. To address this limitation, a dataset of 135 images per category for two white class and about 102 images per category for two red class was captured using handheld mobile device cameras, and data augmentation techniques were employed. The dataset was divided into four distinct groups: "pure red teff injera," "mixed red teff flour with sorghum (zengada) flour injera," "white teff flour with sawdust injera," and "pure white teff injera."

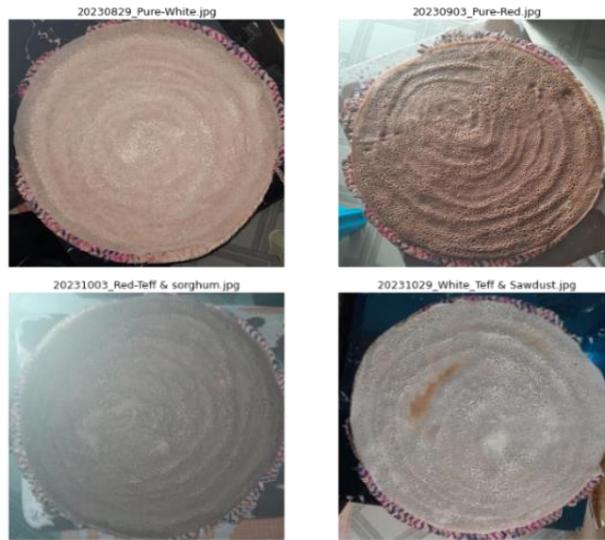


Figure 3- 2 Sample collected ‘injera’ images

3.2.2 Data Pre-processing

Pre-processing is a fundamental step in the development of a deep learning model for image classification since it increases the accuracy of the model and image features by suppressing undesirable distortions and strengthening some essential image features [98], A. H. Havelaar et al. [99]. The first and most important step in every pattern detection and recognition is data collecting. Pre-processing techniques aid in raising contrast, reducing undesired variations, and improving image quality to make images better suited for further analysis and categorization. Pre-processing techniques make it possible to extract significant features like shape, texture, and colour, which are essential for classifying between different categories from the image’s [100].

3.2.2.1 Image Resizing

Normalizing the size of the image keeps the input representation constant because different image sizes have varied amounts of features. As a result, standardizing all image sizes through high quality scaling is an essential pre-processing step for effectively successfully feeding image data into the models. Resizing forces every image to the standard medium dimension because smaller images can lose quality when enlarged and larger images can be compressed without losing information. the collected image dataset for each four class has a dimension of 3056 x 3056 pixels. For the purposes of such as reducing storage and transmission size, improving computational efficiency we were resized our image to 224x224 by preserving its aspect ratio. Bicubic interpolation is an effective technique for scaling images with high quality by efficiently smoothing pixel information. While bilinear interpolation is faster, it tends to

produce more artifacts [31], [32]. The majority of researchers downsized the food images to 299x299 or 224x224 pixels, as described in Section 2.3.3.3. We conducted an experiment using both of these image sizes and achieved nearly the same results. Due to the limited computational resources available to us, we ultimately resized the images to 224x224 pixels using the bicubic interpolation scaling method. This bicubic interpolation resampling algorithm can help minimize quality loss. The resulting resized image is shown in figure 3-3.



Figure 3- 3 Resized sample Image

3.2.2.2 Image Filtering

Images captured in real-world situations are often noisy due to factors like poor illumination or camera shake. Various noise reduction techniques can help minimize the adverse impact of noise on the performance of deep learning models. Noise typically lowers image quality and can be introduced during image capture, transmission, and other processes. Reducing noise is crucial in image processing, as the effectiveness of noise removal significantly influences the quality of image processing techniques. In our study, we used a median filter, which is simple and fast, to preserve edges while reducing noise. The median filter effectively eliminates the impact of noise with large magnitude values [31], S. Sridevy et al. [36], [37].

3.2.2.3 Image Data Augmentation

Data augmentation is one of the most crucial preprocessing methods for deep learning approach which need lots of data to train. Increasing the amount and variety of a training dataset by generating fresh synthetic data samples is a technique employed in deep learning, particularly in the field of machine learning. This aids in overcoming the problems of insufficient training data and unbalanced class sizes [92]. For image datasets, there are numerous techniques for data augmentation. By applying most popular and efficient image transformations, including

rotations, horizontal, vertical and random flips, adding gaussian distortion, zooming to each class, about 1215 and 918 new images are created per each white and red class respectively.

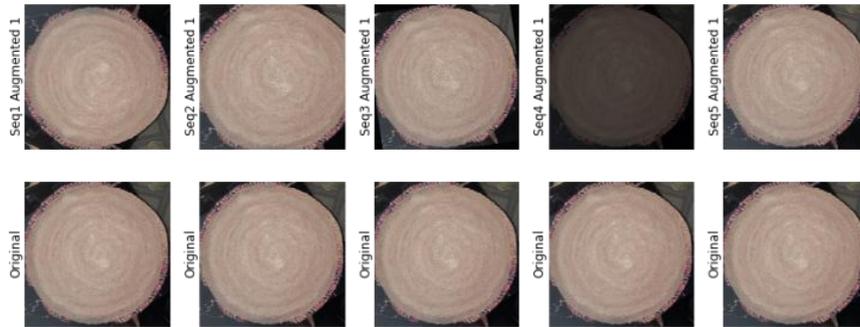


Figure 3- 4 Augmented sample Image

3.2.2.4 Image Enhancement

In our study, we utilized images taken with smartphone cameras, which may have been affected by distortions and noise. To address these issues, we employed image enhancement techniques aimed at increasing image intensity and reducing distortions that could have occurred during image capture or transfer. These variations in the images encompassed factors such as lighting, emotions, noise, pose, blur, and corruption [34]. By enhancing the images, we aimed to improve their overall quality, providing a better foundation for effective feature extraction. Our literature review Section 2.3 highlighted the effectiveness of contrast-limited adaptive histogram equalization, an advanced technique known for its adaptive nature, minimal artifact generation, and reduced amplification of noise. Given these advantages, we chose to utilize this technique for image enhancement in our study, successfully enhancing local details and ultimately improving the quality of the images for more accurate feature extraction.

3.2.2.5 Image Segmentation

One of the most crucial processes in image processing is image segmentation since it greatly affects whether image analysis concentrates on the target sample. It is the process of segmenting or partitioning a digital image into multiple segments or partitions S. Saxena *et al.*[43]. As stated in Section 2.2.3.1 of the literature review, region-based method is more accurate than thresholding method, but it is also slower and more complex. Thresholding method is generally faster, but they may not be able to accurately identify the foreground and background of objects when dealing with varying pixel intensities across objects. Since injera by itself is a circle, we only keep the middle portions of the image with equal radius from the center to the outermost sections of the image in order to enhance segmentation results.

Compared to thresholding techniques, this produces better segmentation results, shown in Figure 3-5 (a) and (b).



Figure 3- 5 Segmented Image using thresholding (a) and Masked out most parts (b)

3.2.2.6 Dataset Splitting

The dataset is divided into training, validation, and testing, each playing distinct roles in model development and evaluation. This partitioning ensures effective learning, optimization, and unbiased assessment of the models. The training dataset is used to train the model by providing important image features such as texture, and colour, which are crucial for accurately classifying different categories. The validation dataset helps in monitoring the model's progress, detecting overfitting (when the model becomes too specialized to the training data), and determining appropriate hyperparameter settings. Finally, the testing dataset serves as an independent evaluation set to assess the model's performance on entirely new data. Thus, the data is divided into three categories: 20% for testing, 80% for training, and 20% for validation from the training dataset. Figure 3-6 depicts the process of data splitting. The total dataset consisted of 4,239 images. Initially, 80% of the data, which is 2,712 images, was allocated for the training set. The remaining 20%, which equated to 848 images, was reserved for testing the model. Additionally, 679 of the training images were set aside for the validation set.

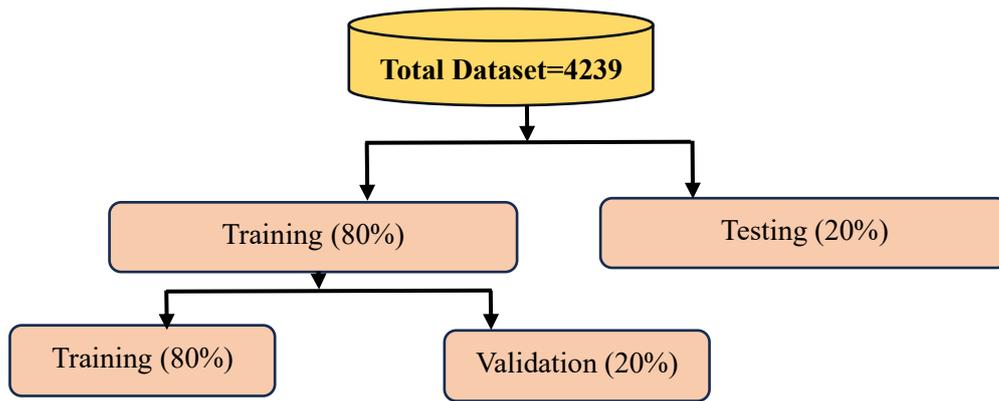


Figure 3- 6 data splitting process

3.2.3 Deep and Handcrafted Feature Extraction

According to the explanation in Chapter two, Section 2.1.1.2, image feature extraction has always been a key component in raising the effectiveness of image classification models. Both the GLCM and LBP operators, which are used in this study, are excellent texture descriptors that are often used in a variety of applications. LBP offers effectiveness, and GLCM raises robustness. The combined feature vector contains more information from the two techniques.

CNNs, a type of Deep Neural Networks, are highly effective in identifying and categorizing specific features in images. The feature extraction component is responsible for extracting relevant features from the input data, while the classification component classifies the input data based on the extracted features M. Khoshdeli *et al.*[101]. The CNN architecture primarily consists of two components: feature extraction, which includes the input, convolution, and pooling layers, and classification, which comprises the fully connected layer and output layer, as depicted in Figure 2-4 of Chapter two.

When an input image is passed through the convolution layer, filters (also known as kernels) slide over the image, generating multiple feature maps that reflect the extraction of distinct features. During the extraction process, neurons within each feature map share a set of weights (or convolution kernel), and the number of feature maps corresponds to the number of convolution kernels utilized. The convolutions in this study are performed in three dimensions (3D) to account for the representation of colour images as 3D matrices, incorporating width, height, and depth dimensions. The depth dimension specifically represents one of the three colours (channels).

After the convolutional layer, the pooling layer diminishes the dimensionality of the convolution output or feature map, along with reducing the number of learnable parameters,

consequently improving computational efficiency. Convolutional neural networks commonly employ three main types of pooling layers: Max pooling, Average pooling, and Sum pooling. Among these, Max pooling is the most popular choice due to its superior performance in various convolutional neural network architectures. Max pooling captures more information about image texture in each region, mitigates overfitting, exhibits resistance to noise, and generally outperforms average or sum pooling Y. Boureau *et al.*[110], M. Cogswell et al.[104]. Considering this, the approach of using max pooling is favoured in this study. By utilizing max pooling, the objective is to preserve a greater amount of texture information, thereby enabling more precise image classification.

After the input image is fed into the convolutional neural network's algorithm, it undergoes convolution and pooling layers. These layers break down the image into features, which are then analysed independently. The fully connected layer plays a crucial role by connecting all the extracted features from the preceding layer, reducing them to their one-dimensional forms, and transmitting the output value to the output layer, where the final classification occurs.

Any deep learning neural network research requires careful attention for the selecting of the activation function, as stated in chapter two section 3 in detail. There are perhaps three activation functions might be considered for the hidden layer; they are ReLU, logistic (Sigmoid) and Softmax Activation Function. ReLU activation function is now the most used activation function for image classification because it provides higher classification accuracy than other activation functions and has a comparatively low computing cost [60]. And the final output layer serves as the final layer of the neural network and its composition varies based on the number of classes in the problem.

3.2.4 Feature Fusion

In this paper, we introduced a multi-feature fusion learning framework for Ethiopian injera impurity detection and classification that combines features manually generated with high-level features obtained via convolutional neural networks. Figure 3-1 illustrates the flowchart of the proposed detection and classification model. Initially, convolutional neural networks which used in this paper are employed to capture high-level features. Subsequently, GLCM and LBP features are utilized to extract handcrafted features, of injera images. These high-level and handcrafted features are then concatenated to attain a more discriminative representation.

High-level features and handcrafted features are combined to create a more discriminative representation distinguishing between pure injera and non-pure injera images. After the feature

extraction process, the three types of features extracted were merged into a composite vector within the final fully connected layer. Each type of feature was given equal weight, resulting in a 1:1:1 weighting scheme. This composite vector was then passed through the SoftMax layer in the output layer for the final classification of injera images. The multi-feature fusion strategy, as depicted in Figure 3-7, illustrates the procedure employed for this purpose.

3.2.5 Dimensionality Reduction

Reduce the number of dimensions in data while retaining the most important information is the core objective of dimensionality reduction. By decreasing the features of operators with the highest features, we use it in our study to balance the effects of both GLCM and LBP features. Here, the feature vector's LBP dimension is greater than the GLCM dimension for the same dimension of the input image, therefore we perform the dimensionality reduction operation before fusing both features. Among the common dimensionality reduction methods Principal Component Analysis is a widely used method that can capture the most important patterns in a dataset. It can reduce the dimensionality of large datasets and is computationally efficient. PCA is noise-resistant and effective at removing duplicate features L. J. P. Van Der Maaten *et al.*[104]. After fusing the GLCM and LBP features, again we need to apply Principal Component Analysis (PCA) to overcome the dominance of the larger number of features.

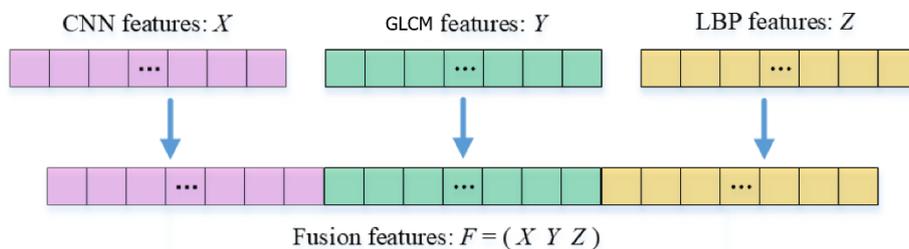


Figure 3- 7 multi-feature fusion strategy

3.3 Experimental Design

Five models are built in the study's experimental design to identify impurities in Ethiopian injera. In the first scenario, two models are created from scratch using CNN and LSTM, In the second scenario, we employ three pre-trained models, which are the YOLO, ResNet50, and AlexNet models. For every model in both approaches, a range of values for different hyperparameters were tested.

3.3.1 Hyperparameter Tuning

Hyperparameter tuning is a crucial step in building a Convolutional Neural Network (CNN) as it has a big influence on how well the model works Z. Li *et al.*[105],F. F. Firdaus et al.[107]. In order to get the best results, hyperparameter tuning involves choosing the appropriate values for a variety of hyperparameters, including the learning rate, optimizer, loss function, epochs, and batch size. There is no fixed set of rules for setting a model's hyperparameters that can always ensure the best possible outcome, as a result, several experiments are to find the best values Z. Li *et al.*[105], F. F. Firdaus et al.[107]. Prior to training the CNN models, hyperparameter tuning is carried out with the goal of determining the ideal values that will improve the accuracy and efficiency of the model. It includes systematically testing with various hyperparameter value combinations and selecting the one that performs the best. The selected hyperparameters for this study are described as follows.

Learning algorithm (Optimizer): To achieve the minimum loss function, Optimizer is responsible for updating the neuronal weights and learning rate within a neural network Z. Li *et al.*[105]. Despite the fact that there are many various kinds of learning algorithms, Adaptive Moment Estimation (Adam) is often used effectively on a broad variety of issues, and compatible with several CNN architectures Z. Li *et al.*[105], L. Alzubaidi et al. [108], [109]. Adam is a method of learning that was created especially for deep learning networks. Adam optimizer combines the benefits of both Momentum and RMSprop kinds of optimizers, offering two basic advantages: reduced computing power and more memory efficiency L. Alzubaidi *et al.*[108].

Loss function: utilized to evaluate model or used to minimizing the error between actual and predicted output I. Cardoza *et al.*[110].

Learning Rate: It determines the size of the step taken at each iteration prior to changing the network's weights Z. Li *et al.*[105]. The model learns more quickly when a higher learning rate is defined than when a lower learning rate is used, however there is a chance that the minimal loss function will be overshoot. Conversely, a lower learning rate requires more processing effort and epochs but increases the likelihood of achieving a minimum loss function. It is an adjustable parameter with a range of 0 to 1.0 that is crucial to the speed at which a model adjusts to an issue. For all accessible optimizers, the hyperparameter learning rate has a default value of 0.001 F. F. Firdaus *et al.*[107], I. Cardoza et al. [110].

Number of Epochs: Describe the number of times with which all training data is displayed to the network throughout training F. F. Firdaus *et al.*[106]. One epoch is the number of times the training dataset is sent through the neural network both forward and backward E. Bochinski *et al.*[111]. If there are too small epochs, the neural network may not have learned enough to tackle the issue effectively, leading to an underfitting model. On the other hand, if several epochs are specified overfitting could a problem I. Cardoza *et al.*[110].

Batch size: The number of samples (images) utilized to train a model during each epoch is described as the batch size. A larger batch size will accelerate the training process, depending upon the available computer resources. However, it results in poor generalization, which keeps the model from typically achieving high accuracy I. Cardoza *et al.*[110]. There is a direct relationship between learning rate and batch size; for high learning rates, large batch sizes outperform than the small ones, and [112] recommends selecting low learning rates for small batches size. In order to determine the ideal batch size, [112] recommend first experimenting with lower batch sizes usually 32 or 64 and the number of batch sizes should be a power of 2.

Activation Function: As elaborated in depth in section c of chapter two, the fundamental role of all forms of activation functions in all types of neural networks is mapping the input to the output. The activation function generates the appropriate output in response to a given input, so determining whether or not to fire a neuron L. Alzubaidi *et al.*[108].

3.3.1.1 Hyperparameter tuning techniques

Although there are more methods for doing hyperparameter tuning, these two methods, grid search and random search are usually used.

Grid Search: All possible hyperparameter combinations from the given hyperparameter space are attempted by the grid search.

Random Search: Random search looks for randomized hyperparameter combinations inside the hyperparameter space. It is made clear how many different combinations of hyperparameters are attempted using random search.

Grid search is a popular method for hyperparameter optimization, where all possible combinations of hyperparameters and their values are examined. When working with even a small number of hyperparameters, though, this strategy becomes computationally costly. To overcome these challenges by concentrating on the most important hyperparameters rather than

thoroughly trying every potential combination, the random search strategy reduces the computing load [108].

3.4 Summary

The architecture of the proposed system for injera impurity classification begins with the acquisition of raw RGB images of injera samples containing various impurities using a smartphone. These images then undergo pre-processing steps, such as resizing to standard dimensions and applying enhancement techniques like contrast adjustment and noise filtering.

Next, feature extraction is performed using two approaches. First, hand-crafted features are extracted using Gray Level Co-occurrence Matrices (GLCM) to encode texture properties and LBP to capture localized texture patterns. Simultaneously, a convolutional neural network is employed to capture higher-level visual characteristics.

The feature vectors obtained from GLCM, LBP, and convolutional neural networks are combined and fed into a convolutional neural networks-based classifier. The classifier consists of convolutional layers that learn feature representations and fully-connected layers that categorize the injera samples into four classes.

CHAPTER FOUR

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Introduction

In this chapter, we discuss the results obtained from our experiments. This study evaluates the performance of various deep learning models, including CNN, LSTM, AlexNet, ResNet50, and YOLO, in classifying injera images into pure red teff injera, pure white teff injera, white teff flour with sawdust flour injera, and red teff flour with sorghum flour injera categories. The previous chapter provided a detailed overview of this topic. This chapter presents comprehensive experimental details, including the outcomes of each experiment and a concise analysis of the findings. The experimental results are visually represented through a variety of graphs, tables, and numerical data. The experimental results demonstrate the performance of deep learning algorithms in two ways. First, the combination of deep learning algorithms and handcrafted features approaches, Second, the performance of a deep learning algorithm alone.

4.2 Data preparation and Pre-processing

Prior to training the model, the images captured on the smartphone were pre-processed to enhance the resolution quality and improve the model's classification effectiveness. First, all images were resized. Then, a median filter was applied, followed by segmentation, enhancement, and data augmentation techniques. Figure 4-1(b) shows the augmented and original images. Next, both manual feature extraction methods, such as GLCM and LBP, as well as a convolutional neural network-based approach, were used to extract features from the images. Finally, the dataset was split into training, testing, and validation sets, as detailed in Section 1.7. A validation set comprising 20% of the training set was created, as shown in Figure 4-1(a).

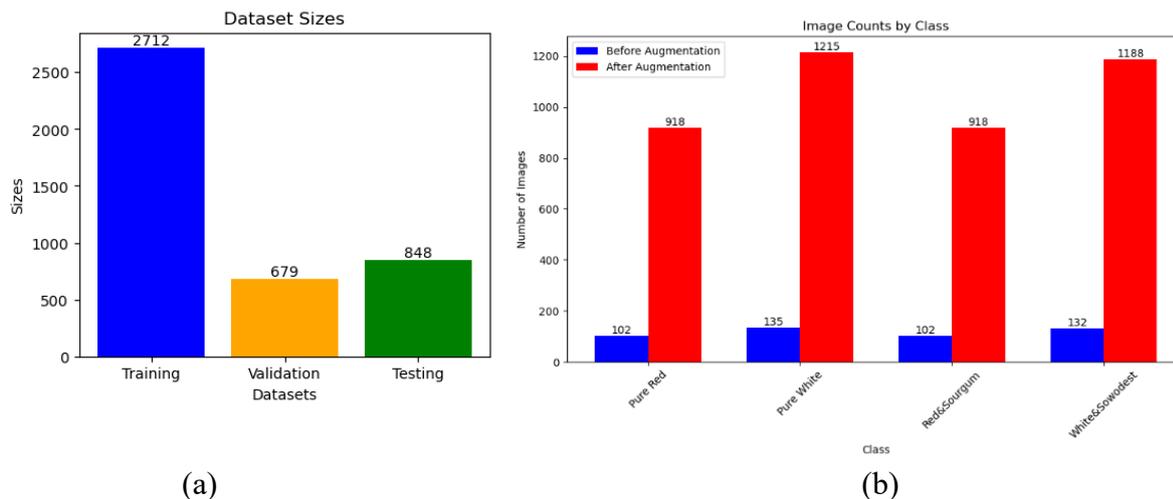


Figure 4- 1 (a)Training, Testing and Validation data size, (b) Original and augmented images

4.3 Hyper parameter Tuning

The model development process involved tuning several hyperparameters, such as the activation function, learning rate, optimization algorithm, loss function, batch size, epochs, dense units, hidden layers, and dropout rate. To automate the hyperparameter tuning, we first specified the search space for these hyperparameters, as detailed in Table 4-1. We then used an automated technique to explore different combinations of these hyperparameter values, evaluating them on the training and validation data to find the optimal configuration for the model.

Table 4- 1 hyperparameter search space

Model Type	Hyperparameter	Range
CNN,	filters	[32, 64, 128]
	Pool size	[2, 3]
LSTM,	Dense units	[32, 64]
	Hidden layers	[2, 3, 4, 5, 6, 7, 8]
AlexNet,	dropout	[0.1, 0.2, 0.3, 0.4, 0.5]
ResNet50	activation	[tanh, softmax, relu]
	optimizer	[adam, adamax]
YOLO	Learning rate	[0.001, 0.01, 0.1]
	Batch size	[32, 64, 128]
	epochs	[10 -150]

The hyperparameter tuning was performed using random search techniques. The tuner was set up with key parameters, including the model building function, validation loss as the objective metric, a maximum of 3 consecutive failed trials for all models, and a maximum of 40 trials

for CNN and LSTM models, and 25 trials for other models. Following an extensive search and evaluation of numerous hyperparameter combinations using the training and validation datasets, the tuner identified the optimal model architecture. The best hyperparameters were then extracted using the `get_best_hyperparameters` function, which retrieves the top hyperparameter values found during the search process. After completing the hyperparameter tuning phase, the optimal hyperparameter values for each model were compiled in Table 4-2.

Table 4- 2 Optimal Hyper parameters value

Model	Hyperparameter and corresponding Best Value							
	Epochs	Batch size	Dense Units	Hidden Layers	Dropout	Activation	Optimizer	Learning Rate
YOLO	50	64	32	3	0.2	softmax	adamax	0.01
CNN	100	32	32	4	0.3	softmax	adam	0.001
LSTM	100	128	32	8	0.2	softmax	adamax	0.001
AlexNet	150	64	64	2	0.1	softmax	adam	0.001
ResNet50	50	64	64	7	0.1	relu	adamax	0.01

4.4 Model Building and Experimental Results

Several experiments were conducted to build a model capable of classifying injera based on impurity mixtures. After completing the preprocessing and feature extraction tasks, and identifying the optimal hyperparameters through tuning, the models were constructed using these optimal parameter values.

To begin the model building process, the 'best_hyperparameters' dictionary, containing the optimal hyperparameters for each model type, was loaded from a file using the 'pickle' module. Additionally, empty lists were initialized to store the training and validation histories during the tuning process. Several empty dictionaries were also created to store performance summaries, evaluation results, predictions, and the trained models. Finally, the findings were presented, including plots of the loss vs. epochs, accuracy vs. epochs, and the confusion matrix for the best model identified during the model building process. The loss and accuracy trends for both the training and validation data during the training phase are shown in Figure 4.2 (a-h).

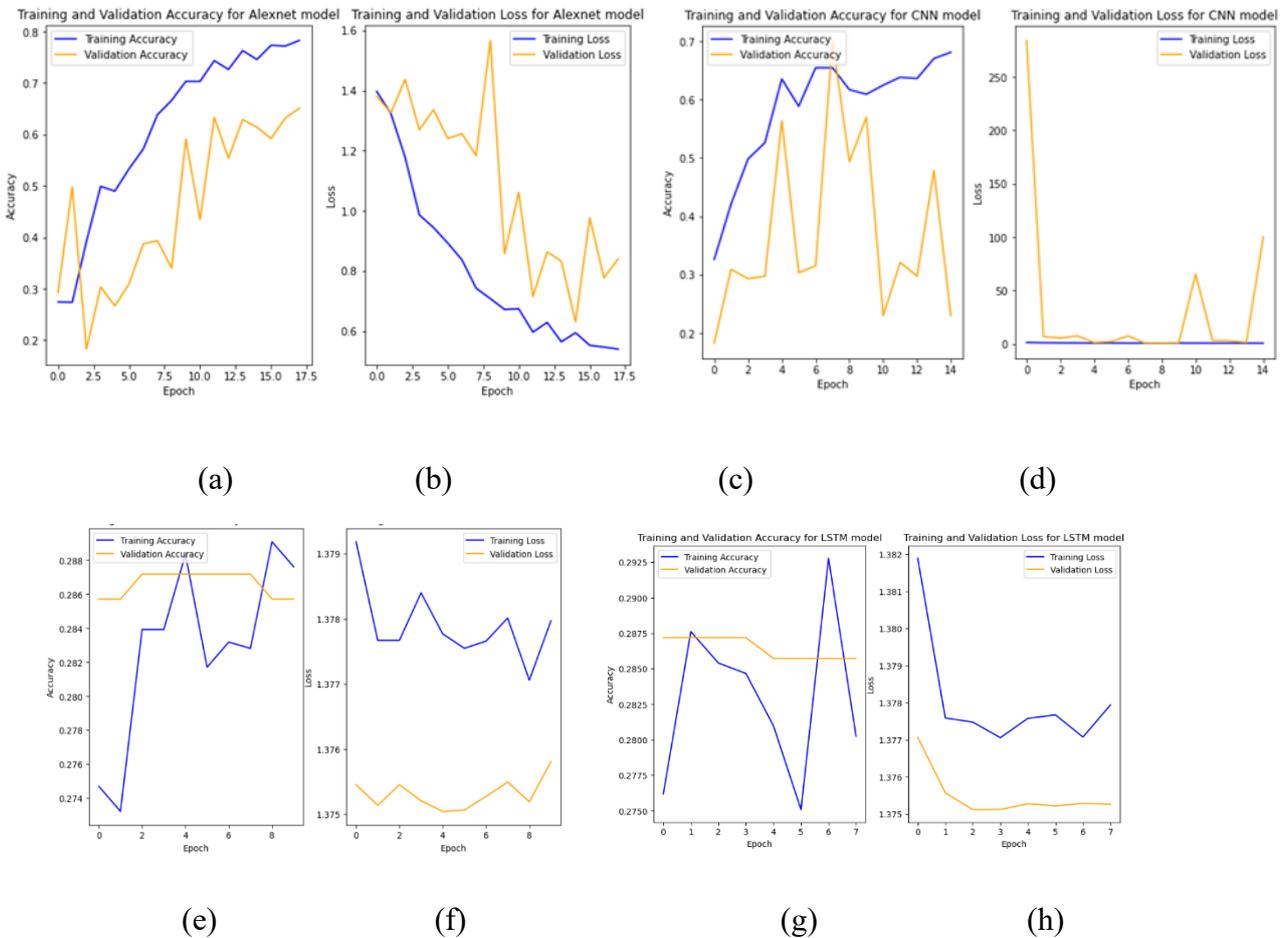


Figure 4- 2 a) Accuracy vs Epoch for AlexNet b) Loss vs Epoch for AlexNet
 c) Accuracy vs Epoch for ResNet50 d) Loss vs Epoch for ResNet50
 e) Accuracy vs Epoch for Yolo f) Loss vs Epoch for Yolo
 g) Accuracy vs Epoch for LSTM h) Loss vs Epoch for LSTM

The analysis of Figure 4.2 (a-h) reveals the following insights:

For Figure 4.2 (a-b), the training accuracy and loss remain constant, while the validation accuracy and loss exhibit variability, indicating this model is unsuitable for the given datasets. In contrast, the graphs in Figure 4.2 (c-h) for ResNet50, AlexNet, YOLO, and LSTM models show significant variations in both the training and validation accuracy and loss over multiple epochs, suggesting these models are also not suitable for the datasets.

Based on the experimental results, it was concluded that CNN-type models are the best choice for the problem at hand. This is due to superior performance in terms of accuracy, precision, recall, and F1 score, observed in the classification results from the combination of deep learning algorithms and handcrafted features approaches and the performance of a deep learning algorithm alone.

Table 4- 3 Result Analysis of pure CNN-based approach and CNN and Handcrafted Features Fusion Approach

Metric	Pure CNN-based approach	CNN and Handcrafted Features Fusion Approach
Training Accuracy	0.79	0.77
Training Loss	0.45	0.55
Test Accuracy	0.77	0.71
Test Loss	0.57	0.60

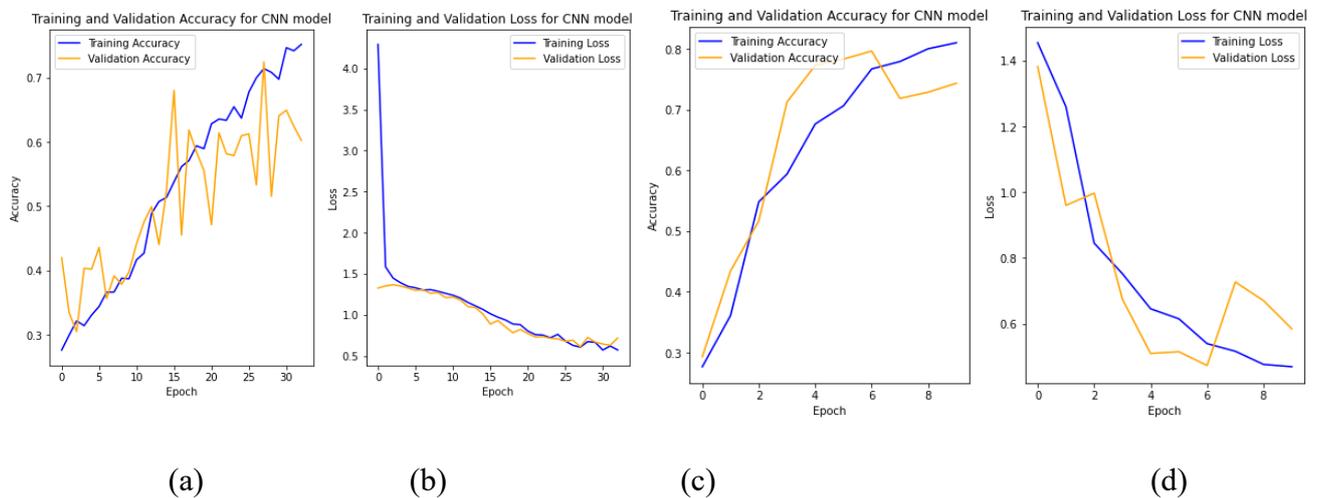


Figure 4- 3 Accuracy vs Epoch for CNN and Handcrafted Features Fusion Approach
 b) Loss vs Epoch for CNN and Handcrafted Features Fusion Approach
 c) Accuracy vs Epoch for Pure CNN-based approach
 d) Loss vs Epoch for Pure CNN-based approach

For the hybrid based approach, both the validation loss and training loss function gradually decreases with the increase in the training epochs and also the training accuracy becomes increase as the training epochs increase with a bit fluctuation , indicating the model is learning over epochs, but the validation accuracy is variable and unstable through the sample analysis process starting from the beginning to the end which is unpredictable as shown in Figure 4- 1 (a-b), these makes this approach is unsuitable for our problem.

For the Pure CNN-based approach, the training and test accuracy are 79% and 77% as tabulated in Table 4- 4 which provide better accuracy compared with the CNN + handcrafted features approach, and it produces the result with fewer iterations. and in addition, Figure 4- 2

(c-d) shows a rapid initial increase, also starts low and steadily improves and the trend is smoother compared to the hybrid approach. The training loss and validation loss decreases sharply initially, then continues to decline at a slower rate. The training loss of 0.47 and test loss of 0.57 further highlight the model's effective learning to generalize beyond the training data in contrast to CNN and Handcrafted Features Fusion Approach which struggle to generalize.

The divergence in performance between the two models suggests that the Pure CNN-based approach is a more robust and well-generalized model, capable of effectively learning and extracting relevant features from the data. In contrast, the CNN and Handcrafted Features Fusion Approach may be relying more on specific patterns in the training data, leading to reduced generalization to the test set.

Overall, the Pure CNN-based approach model demonstrates superior performance and generalization compared to the CNN and Handcrafted Features Fusion Approach, making it the more robust and reliable model for our research.

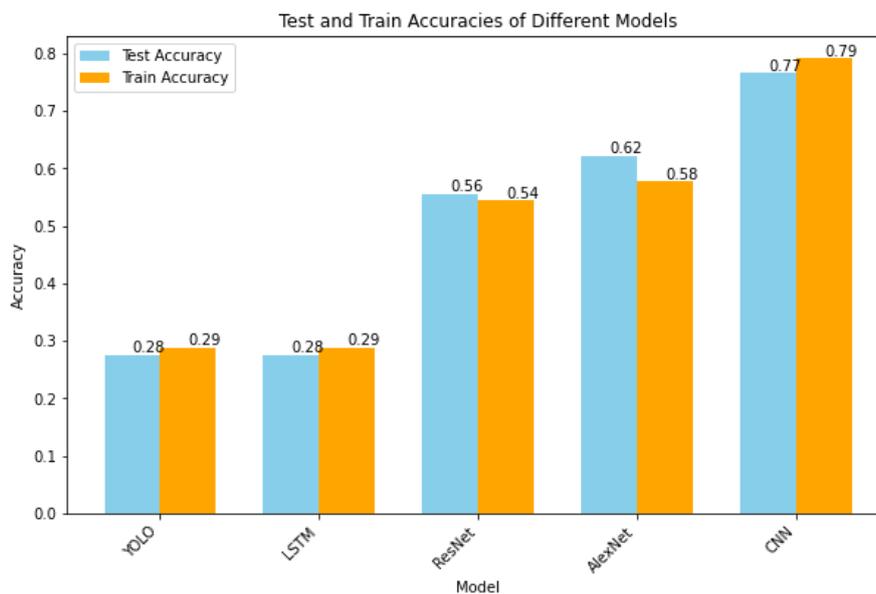


Figure 4- 4 Comparison of Accuracy vs. Model

The graph in Figure 4-4 compares the test and training accuracies of the YOLO, LSTM, Alex Net, ResNet50, and CNN deep learning models. Relatively, the ResNet50 and CNN models exhibit the highest test and training accuracies, indicating better performance and stronger generalization capabilities. This is also corroborated by the results shown in Figure 4-1.

4.5 Evaluation of Model

The evaluation of model classification performance is based on several key metrics, including accuracy, precision, recall, and F1-score. These metrics provide insights into the effectiveness and reliability of each classification model in accurately identifying impurities in teff injera images.

Table 4- 5 Accuracy, F1 score, Precision and Recall for CNN model in Pure CNN-based approach model

Metrix	Pure Red	Pure White	Red&Sourgum	White&Sowodest
Precision	0.85	0.70	0.88	0.72
Recall	0.92	0.62	0.54	0.96
F1-score	0.88	0.66	0.66	0.82
Support	183	234	185	246
Overall Accuracy	77%			

This table represents a high-performing model with an overall test accuracy of 77%. The model demonstrates strong precision, recall, and F1-score across all the classes, indicating that it is able to accurately classify the different types of images with a high degree of confidence.

Table 4- 6 Precision, recall, F1 score and Accuracy values for CNN and Handcrafted Features Fusion Approach

Metric	Pure Red	Pure White	Red&Sourgum	White&Sowodest
Precision	0.64	0.74	0.98	0.71
Recall	0.99	0.67	0.59	0.73
F1-score	0.77	0.71	0.56	0.72
Support	210	253	165	220
Overall Accuracy	71%			

In contrast to Pure CNN-based approach, CNN and Handcrafted Features Fusion Approach shows less performance with an overall accuracy of 71% for CNN model. The dataset was divided into training, validation, and evaluation parts, as explained in Section 5.2. The training and validation data were used to train the model, while the evaluation data was used to assess the model's performance. The validation data was used to monitor the model's response to new similar data, but was not used to directly train the network. Finally, the evaluation data was used to assess the model's overall performance. The output images shown

in Figure 4-7 and 4-8 are examples of the classification results from both approach on our dataset. Figure 4.7 displays examples of incorrectly classified injera images, while Figure 4.8 shows examples of correctly classified images. All of the displayed images were classified with 100% certainty.

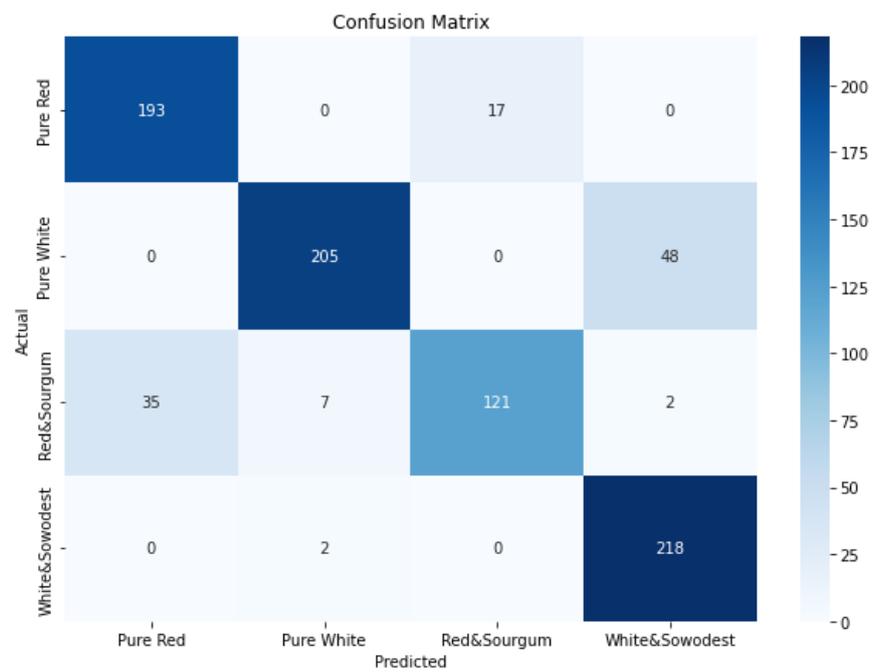


Figure 4- 5 Confusion matrix of CNN Model in pure CNN-based approach model

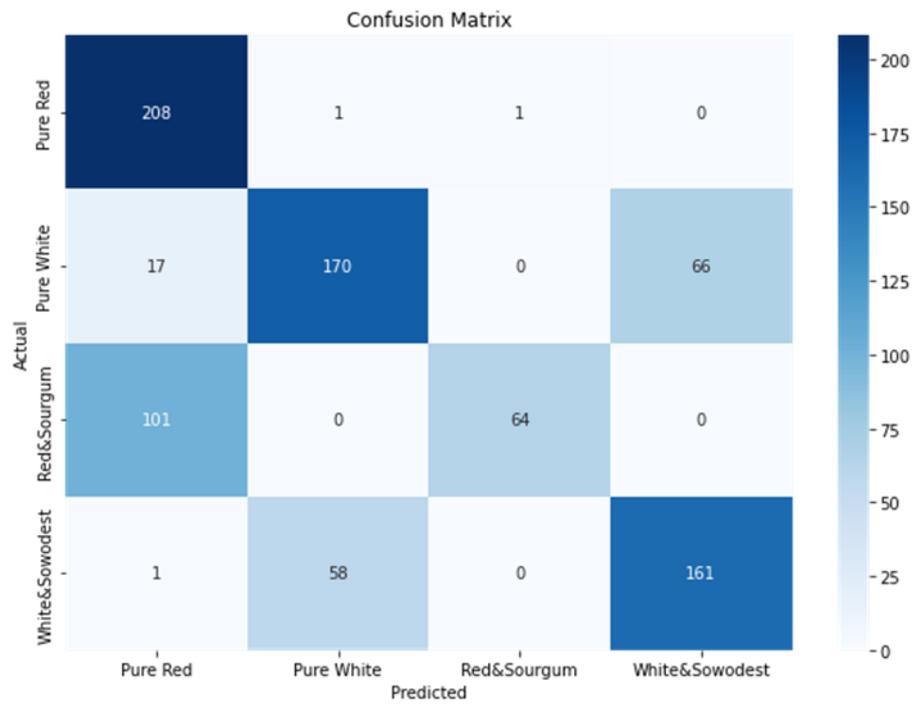


Figure 4- 6 Confusion matrix of CNN model for CNN and Handcrafted Features Fusion Approach



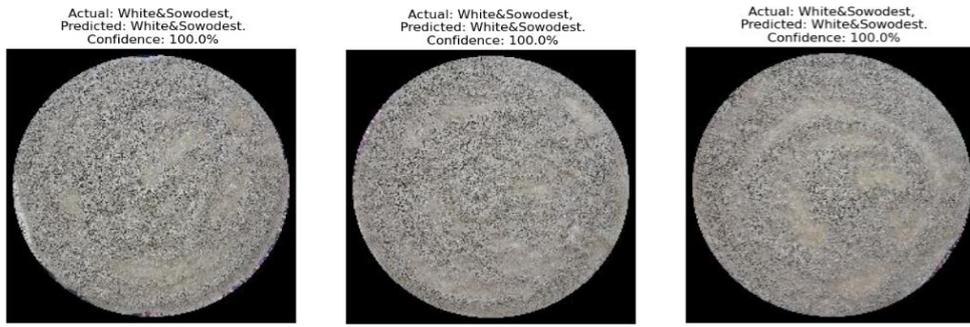


Figure 4- 7 Some correctly classified images

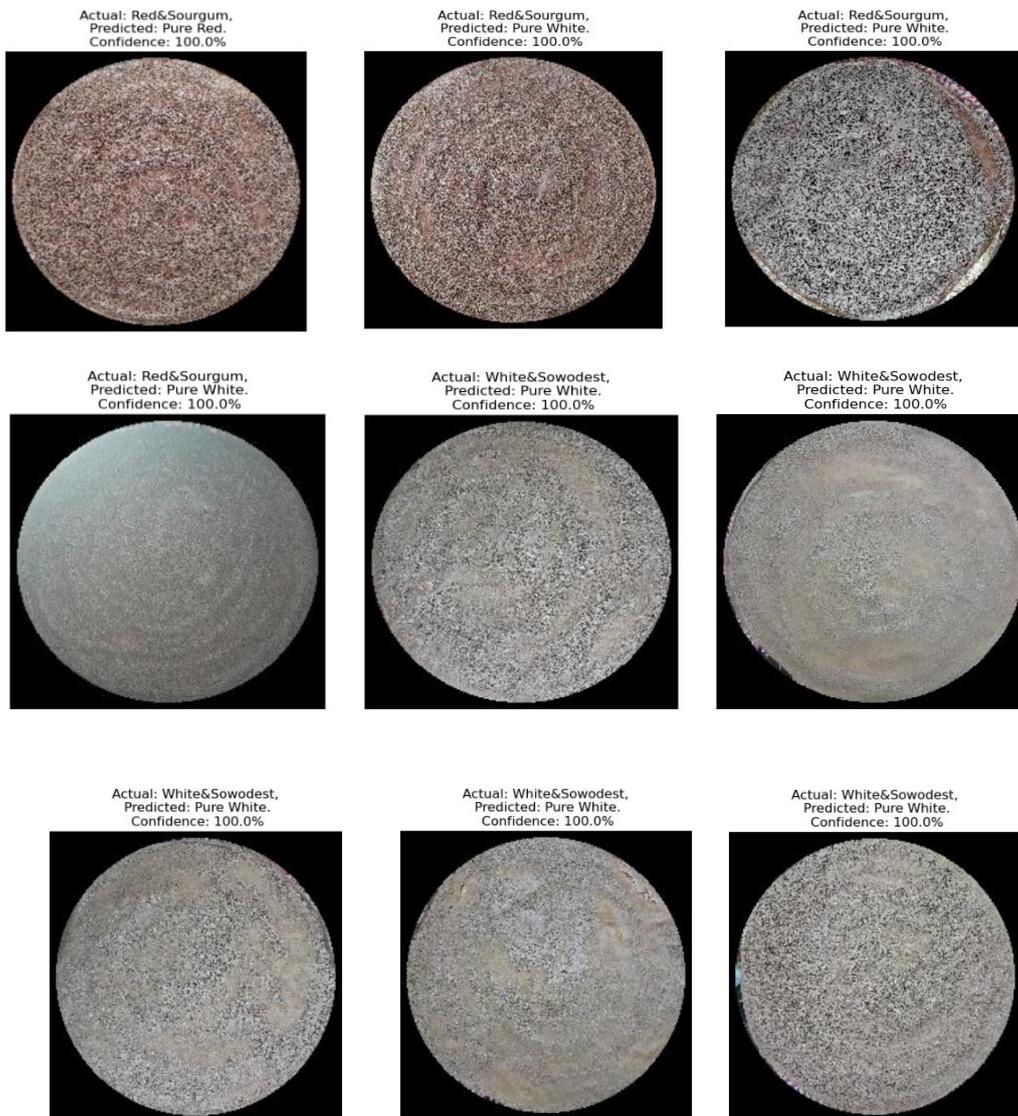


Figure 4- 8 Some incorrectly classified images

To determine the test cases that show failed classifications, we need to consider the various limitations and challenges present in the image dataset. The model's performance can be hindered by the characteristics of the input images themselves. Since the images were captured from different locations at different times with varying lighting conditions, the dataset may contain a varying range of backgrounds, these edge case deviate leads to be misclassified. Another factor to consider is the visual similarity between the classes, the model is trying to classify. From four different classes the two classes contain visually similar injera images, as a result the model is more likely to confuse and misclassify examples between these similar classes. By carefully examining these aspects of the image data, the input quality and class similarity we can identify the test cases that are most likely to result in failed classifications. This analysis can help pinpoint the specific challenges and limitations of the model, enabling targeted improvements to address these issues.

Over all, various algorithms, such as YOLO, LSTM, AlexNet, ResNet50, and CNN deep learning models, are tested for the comparative analysis in both Pure CNN-based approach model and CNN and Handcrafted Features Fusion Approach. The results are obtained for four parameters such as precision, recall, F1-score and accuracy. From those all models we obtained promising performance for CNN model in both approaches. The detail comparisons are plotted and tabulated for both approaches in a graph and table in section 4.4 and 4.5. furthermore, from the result, CNN model in CNN-based approach provides the highest results in all evaluation parameters with the fewest epochs 15 epochs. Considering the second approach for CNN model, the overall test samples were run and provided an accuracy of around 77%. However, here, the loss function is very maximal which is 0.60 and with the highest epochs 33, compared to with the first approach. Compared with the other models in both approaches, YOLO and LSTM offered the worst performance.

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATION

5.1 Conclusions

In conclusion, this study investigated the ability of different deep learning algorithms using two different approaches those are CNN and Handcrafted Features Fusion Approach and Pure CNN Approach. Using these two approaches we build the models from scratch, including CNN and LSTM, and pre-trained models such as YOLO, ResNet50, and AlexNet. CNN and Handcrafted Features Fusion Approach used the concatenations of GLCM with LBP handcrafted features and CNN features whereas the Pure CNN Approach used only the feature extracted from CNN in the classification of injera.

We prepared a custom dataset with a total of 471 injera images in four different categories of Injera. The dataset of 135 images per category for two different white class and about 102 images per category for two different red class was captured using handheld mobile device cameras, and data augmentation techniques were employed. We performed pre-processing steps including noise removal, resizing, enhancement, and segmentation to prepare the images for the model. The dataset was then augmented using eight augmentation factors, resulting in approximately 4239 total augmented images. The performance of all models was evaluated for pure CNN-based approach and achieved an accuracy of 79%, 58%, 54% for CNN, AlexNet, ResNet50 respectively and the lowest recorded for both YOLO and LSTM models in both approaches.

CNN models are the most widely used neural networks for image classification tasks, and from our experimental results it was concluded that CNN type networks are the best choice for our problems due to the accuracy, precision, recall, and F1 score of the classification result and the smallest loss function than ResNet50, YOLO and LSTM, and AlexNet models.

Experimental results demonstrated the highest training accuracy of 79% and test accuracy 77% for pure CNN-based approach and 77% training accuracy with 71% test accuracy for the CNN and Handcrafted Features Fusion Approach. Our results demonstrated that the model using pure CNN-based approach outperformed than the model using CNN and Handcrafted Features Fusion Approach. It is well known that CNN have a proven capability to learn complex features directly from raw image data, which might be more representative and discriminative than handcrafted features. During feature extraction the dimensions of hand-crafted feature is very

small as compared to feature dimension extracted by CNN. Reducing the dimensionality of CNN features to match the lower-dimensional handcrafted features could have led to a loss of some important information. Additionally, the fusion technique employed may not have effectively integrated the complementary strengths of the two feature sets. Even if it needs further research, this may be the possible reason of the lowest results of models using CNN and Handcrafted Features Fusion Approach.

5.2 Contributions of the study

The primary objective of this study was to build a deep learning model capable of classifying injera into four categories: pure white teff injera, teff with sawdust injera, pure red teff injera, and red teff flour mixed with sorghum flour injera. To the best of our knowledge, this research is the first investigation of injera impurity identification. This objective was achieved by experimenting with five different models for the impurity detection task, using both CNN and Handcrafted Features Fusion Approach as well as pure CNN approach. We achieved the best results using the CNN model in both approaches.

Another key contribution is the injera dataset used in this study, which will be publicly available to facilitate further advancements in this area. Deep learning approaches often face challenges due to limited data availability, so this dataset contribution is an important output of the work. Overall, the study presents deep learning-based solution for classifying injera purity, as well as a valuable new dataset to support continued research in this domain.

5.3 Recommendations

This study primarily focused on classifying injera made from four different flour compositions: pure white teff flour, white teff flour mixed with sawdust, pure red teff flour, and red teff flour mixed with sorghum at a 15:85 ratio.

- One of the challenges was that the required data preparation resulted in limited data size. To address this challenge, future work should focus on collecting more data with a similar composition to improve the performance of the deep learning models.
- Future research should explore a wider range of combination ratios of teff and other flour types, as well as mixtures containing various impurities such as mold, rice or wheat.
- The image prepared for this study was conducted in an uncontrolled environment. To further strengthen the research, future work should expand the dataset and evaluate the models using injera samples captured in a more controlled environment.

Reference

- [1] A. Choudhary, N. Gupta, F. Hameed, and S. Choton, "An overview of food adulteration: Concept, sources, impact, challenges and detection," *Int J Chem Stud*, vol. 8, no. 1, pp. 2564–2573, 2020, doi: 10.22271/chemi.2020.v8.i1.am.8655.
- [2] M. L. Habte, E. A. Beyene, T. O. F. A. Tilahun, G. C. Diribsa, and F. T. Admasu, "Nutritional Values of Teff (*Eragrostis tef*) in Diabetic Patients: Narrative Review," *Diabetes, Metabolic Syndrome and Obesity*, vol. 15, no. August, pp. 2599–2606, 2022, doi: 10.2147/DMSO.S366958.
- [3] E. Mulugeta and A. Mohammed, "Determination of the levels of iron from red, mixed and white teff (*Eragrostis*) flour by using UV/visible spectrophotometry," *Journal of Natural Sciences Research*, vol. 5, no. 9, pp. 34–42, 2015.
- [4] S. Jiang, W. Min, L. Liu, and Z. Luo, "Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 265–276, 2020, doi: 10.1109/TIP.2019.2929447.
- [5] X. Huang, Z. Li, M. Zhang, and S. Gao, "Fusing hand-crafted and deep-learning features in a convolutional neural network model to identify prostate cancer in pathology images," *Front Oncol*, vol. 12, no. September, pp. 1–17, 2022, doi: 10.3389/fonc.2022.994950.
- [6] D. Liu, Y. Liu, S. Li, W. Li, and L. Wang, "Fusion of Handcrafted and Deep Features for Medical Image Classification," *J Phys Conf Ser*, vol. 1345, no. 2, 2019, doi: 10.1088/1742-6596/1345/2/022052.
- [7] T. Chen and J. Yang, "A Novel Multi-Feature Fusion Method in Merging Information of Heterogenous-View Data for Oil Painting Image Feature Extraction and Recognition," *Front Neurorobot*, vol. 15, no. July, pp. 1–9, 2021, doi: 10.3389/fnbot.2021.709043.
- [8] J. Lu, C. X. Ma, Y. R. Zhou, M. X. Luo, and K. B. Zhang, "Multi-feature fusion for enhancing image similarity learning," *IEEE Access*, vol. 7, pp. 167547–167556, 2019, doi: 10.1109/ACCESS.2019.2953078.
- [9] A. Humeau-heurtier, "Texture Feature Extraction Methods : A Survey," *IEEE Access*, vol. PP, no. c, p. 1, 2018, doi: 10.1109/ACCESS.2018.2890743.
- [10] A. Patil and M. Rane, "Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition," *Smart Innovation, Systems and Technologies*, vol. 195, pp. 21–30, 2021, doi: 10.1007/978-981-15-7078-0_3.
- [11] A. H. Havelaar *et al.*, "Burden of foodborne disease due to bacterial hazards associated with beef, dairy, poultry meat, and vegetables in Ethiopia and Burkina Faso, 2017," *Frontiers in Sustainable Food Systems*, vol. 6, 2022, doi: 10.3389/fsufs.2022.1024560.
- [12] "ከጤፍ ዱቄት ጋር ሲጋቱራ በመቀላቀል ለሕብረተሰቡ ሊያከፋፍሉ የነበሩ ግለሰቦች በቁጥጥር ስር ዋሉ - Welcome to Fana Broadcasting Corporate S.C." Accessed: Sep. 09, 2023. [Online]. Available: <https://www.fanabc.com/archives/193401>

- [13] B. S. Advisor *et al.*, “COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES CENTER FOR FOOD SCIENCE AND NUTRITION Screening food test parameters to detect adulteration of teff (*Eragrostis tef* (Zucc .) Trotter) flour and injera with non-edible adulterants,” no. June, 2018.
- [14] S. K. Ardestani, “Combination of texture , color and shape operators to describe image content : a survey,” no. December, pp. 18–22, 2020, doi: 10.20944/preprints202012.0479.v1.
- [15] S. JiangW. MinL. Liuand Z. Luo, “Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition,” *IEEE Transactions on Image Processing*, vol. 29, no. XX, pp. 265–276, 2020, doi: 10.1109/TIP.2019.2929447.
- [16] А. Д. Кулик, “Интегративно-Модульный Подход И Его Реализация В Профессионально Ориентированном Обучении Иностранных Студентов-Нефилологов. Уровни А1, А2, В1,” *Интегративно-Модульный Подход И Его Реализация В Профессионально Ориентированном Обучении Иностранных Студентов-Нефилологов. Уровни А1, А2, В1*, vol. 36, no. 2, pp. 193–207, 2023, doi: 10.31862/9785426311961.
- [17] E. TsaleraA. PapadakisM. Samarakouand I. Voyiatzis, “Feature Extraction with Handcrafted Methods and Convolutional Neural Networks for Facial Emotion Recognition,” *Applied Sciences (Switzerland)*, vol. 12, no. 17, 2022, doi: 10.3390/app12178455.
- [18] A. Parvatand S. Dev, “A Survey of Deep-learning Frameworks,” pp. 1–7, 2017.
- [19] “Top 8 Deep Learning Frameworks You Should Know in 2023.” Accessed: Sep. 10, 2023. [Online]. Available: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-frameworks>
- [20] “Galaxy M13 | SM-M135FZGGMEA | Samsung Business Levant.” Accessed: Sep. 10, 2023. [Online]. Available: <https://www.samsung.com/levant/business/smartphones/galaxy-m/galaxy-m13-sm-m135fzggmea/>
- [21] A. M. HambalZ. Peiand F. Libent Ishabailu, “Image Noise Reduction and Filtering Techniques,” *International Journal of Science and Research*, vol. 6, no. 3, pp. 2319–7064, 2015, doi: 10.21275/25031706.
- [22] M. Maruand M. C., “Image Restoration Techniques: A Survey,” *Int J Comput Appl*, vol. 160, no. 6, pp. 15–19, 2017, doi: 10.5120/ijca2017913060.
- [23] S. S. Mohantyand S. Tripathy, “Application of different filtering techniques in digital image processing,” *J Phys Conf Ser*, vol. 2062, no. 1, 2021, doi: 10.1088/1742-6596/2062/1/012007.
- [24] C. SwathiB. K. AnoopD. A. S. Dhasand S. P. Sanker, “Comparison of different image preprocessing methods used for retinal fundus images,” *2017 Conference on Emerging*

- Devices and Smart Systems, ICEDSS 2017*, no. March, pp. 175–179, 2017, doi: 10.1109/ICEDSS.2017.8073677.
- [25] S. Sonawane, M. Awasthy, and N. Choubey, “A Literature Review on Image Processing and Classification Techniques for Agriculture Produce and Modeling of Quality Assessment system for Soybean industry Sample,” *International Journal of Innovative Research in Electronics and Communications*, vol. 6, no. 2, pp. 8–16, 2019, doi: 10.20431/2349-4050.0602002.
- [26] S. Sridevi, M. N. Devi, M. Sankar, and N. R. Moorthi, “Image pre-processing techniques utilized for the plant identification : A review,” vol. 12, no. 6, pp. 320–324, 2023.
- [27] R. Al-taie, B. Saleh, and H. Abu-Alsaad, “a Review Paper : Digital Image Filtering Processing,” *Technium*, vol. 3, no. 9, pp. 1–11, 2021.
- [28] N. Kumar and M. Nachamai, “Noise Removal and Filtering Techniques used in Medical Images,” *Oriental journal of computer science and technology*, vol. 10, no. 1, pp. 103–113, 2017, doi: 10.13005/ojcs/10.01.14.
- [29] C. Engineering, “Comprehensive Review of Denoising,” pp. 4597–4602, 2014.
- [30] B. Kanchanadevi and P. R. Tamilselvi, “PREPROCESSING USING IMAGE FILTERING METHOD AND TECHNIQUES FOR MEDICAL IMAGE COMPRESSION TECHNIQUES We begin with the Partial Differential Equation :,” pp. 2132–2135, 2020, doi: 10.21917/ijivp.2020.0304.
- [31] O. Rukundo, “Effects of Image Size on Deep Learning,” *Electronics (Switzerland)*, vol. 12, no. 4, pp. 1–21, 2023, doi: 10.3390/electronics12040985.
- [32] Mr. P. S. Parsania and Dr. P. V. Virparia, “A Comparative Analysis of Image Interpolation Algorithms,” *Ijarce*, vol. 5, no. 1, pp. 29–34, 2016, doi: 10.17148/ijarce.2016.5107.
- [33] Prof. P. S. Parsania and Dr. P. V. Virparia, “A Review: Image Interpolation Techniques for Image Scaling,” *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 02, no. 12, pp. 7409–7414, 2015, doi: 10.15680/ijrcce.2014.0212024.
- [34] P. B. Kutade and P. S. Arora Bhalotra, “A Survey on Various Approaches of Image Steganography,” *Int J Comput Appl*, vol. 109, no. 3, pp. 1–5, 2015, doi: 10.5120/19165-0620.
- [35] S. Karanwal, “Analysis on Image Enhancement Techniques,” *International Journal of Engineering and Manufacturing*, vol. 13, no. 2, pp. 9–21, 2023, doi: 10.5815/ijem.2023.02.02.
- [36] A. Kumar Pathak and M. P. Parsai, “A Study of Various Image Fusion Techniques,” *International Journal of Engineering Trends and Technology*, vol. 15, no. 2, pp. 59–62, 2014, doi: 10.14445/22315381/ijett-v15p213.

- [37] A. Dwivedi and S. Gupta, "Comparative study of different image reconstruction methods," *Proceedings - 2018 International Conference on Communication, Information and Computing Technology, ICCICT 2018*, vol. 2018-Janua, no. 1, pp. 1–4, 2018, doi: 10.1109/ICCICT.2018.8325875.
- [38] N. Verma and M. Dutta, "Contrast Enhancement Techniques : A Brief and Concise Review," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 7, pp. 1767–1771, 2017.
- [39] L. Zhu, P. Spachos, E. Pensini, and K. N. Plataniotis, "Deep learning and machine vision for food processing: A survey," *Curr Res Food Sci*, vol. 4, pp. 233–249, 2021, doi: 10.1016/j.crfs.2021.03.009.
- [40] R. Sarki *et al.*, "Image Preprocessing in Classification and Identification of Diabetic Eye Diseases," *Data Sci Eng*, vol. 6, no. 4, pp. 455–471, 2021, doi: 10.1007/s41019-021-00167-z.
- [41] R. Deshmukh and A. Vibhute, "a Review on Digital Image Processing: Applications, Techniques and Approaches in Various Fields," *Int J Adv Res (Indore)*, vol. 8, no. 6, pp. 726–734, 2020, doi: 10.21474/ijar01/11152.
- [42] Neetu Rani, "Image Processing Techniques: A Review," *Journal on Today's Ideas - Tomorrow's Technologies*, vol. 5, no. 1, pp. 40–49, 2017, doi: 10.15415/jotitt.2017.51003.
- [43] "TECHNIQUES-A REVIEW," vol. 6, no. 9, pp. 56–61, 2019.
- [44] S. Saxena, S. Jain, S. Tripathi, and K. Gupta, "Comparative Analysis of Image Segmentation Techniques," *Lecture Notes in Electrical Engineering*, vol. 668, no. 10, pp. 317–331, 2021, doi: 10.1007/978-981-15-5341-7_26.
- [45] M. Parmar, "Image Segmentation Techniques : A Comparative Study," vol. 3, no. 01, pp. 1–6, 2015.
- [46] G. A. Tahir and C. K. Loo, "A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment," *Healthcare (Switzerland)*, vol. 9, no. 12, 2021, doi: 10.3390/healthcare9121676.
- [47] C. A. and R. G. Shrabani Medhi^{1*}, "Open access Open access," *A Study on Feature Extraction Techniques in Image Processing*, vol. 2, no. 1, pp. 56–61, 2021.
- [48] S. Ahmed Medjahed, "A Comparative Study of Feature Extraction Methods in Images Classification," *International Journal of Image, Graphics and Signal Processing*, vol. 7, no. 3, pp. 16–23, 2015, doi: 10.5815/ijigsp.2015.03.03.
- [49] N. Bagri and P. K. Johari, "A Comparative Study on Feature Extraction using Texture and Shape for Content Based Image Retrieval," vol. 80, pp. 41–52, 2015.

- [50] F. Marpaung, Z. A. Koemadji, M. Hidayat, and A. Widiyanto, "Selection of Food Identification System Features Using Convolutional Neural Network (CNN) Method," vol. 10, no. 2, pp. 205–216, 2023, doi: 10.15294/sji.v10i2.Scientific.
- [51] A. Ramola, A. K. Shakya, and D. Van Pham, "Study of statistical methods for texture analysis and their modern evolutions," *Engineering Reports*, vol. 2, no. 4, pp. 1–24, 2020, doi: 10.1002/eng2.12149.
- [52] L. O. Osapoetra, W. Chan, W. Tran, M. C. Kolios, and G. J. Czarnota, "Comparison of methods for texture analysis of QUS parametric images in the characterization of breast lesions," *PLoS One*, vol. 15, no. 12 December, pp. 1–25, 2020, doi: 10.1371/journal.pone.0244965.
- [53] C. Reischauer, R. Patzwahl, D. M. Koh, J. M. Froehlich, and A. Gutzeit, "Texture analysis of apparent diffusion coefficient maps for treatment response assessment in prostate cancer bone metastases—A pilot study," *Eur J Radiol*, vol. 101, no. November 2017, pp. 184–190, 2018, doi: 10.1016/j.ejrad.2018.02.024.
- [54] W. M. Thu, Dr. P. E. Sanand, D. W. T. Tun, "GLCM and LTP Based Classification of Food Types," *International Journal of Science and Engineering Applications*, vol. 7, no. 08, pp. 218–222, 2023, doi: 10.7753/ijsea0708.1018.
- [55] M. Sikiö, *Textural Features in Medical Magnetic Resonance Image Analysis of the Brain and Thigh Muscles*, no. 2016. 2016.
- [56] H. Alshazly, C. Lin, S. E. Barth, and T. Martinetz, "Handcrafted versus CNN features for ear recognition," *Symmetry (Basel)*, vol. 11, no. 12, pp. 1–27, 2019, doi: 10.3390/SYM11121493.
- [57] Q. Lv, S. Zhang, and Y. Wang, "Deep Learning Model of Image Classification Using Machine Learning," *Advances in Multimedia*, vol. 2022, 2022, doi: 10.1155/2022/3351256.
- [58] P. Song *et al.*, "Feature Extraction and Target Recognition of Moving Image Sequences," *IEEE Access*, vol. 8, pp. 147148–147161, 2020, doi: 10.1109/ACCESS.2020.3015261.
- [59] R. Dastres, M. Soori, A. Neural, N. Systems, and I. Journal, "Artificial Neural Network Systems To cite this version : HAL IAd : hal-03349542," *International Journal of Imaging and Robotics (IJIR)*, vol. 21, no. 2, pp. 13–25, 2021.
- [60] K. Potter, and L. Karl, "TITLE : Transfer Learning and Pretrained Models AUTHOR," 2023.
- [61] C. U. L. Ganesan, "New Image Classification using Convolutional Neural Network (CNN) by Deep Learning," vol. 40, pp. 1719–1731, 2020.
- [62] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 6, pp. 1–20, 2021, doi: 10.1007/s42979-021-00815-1.

- [63] R. WuS. Zhaoand Z. Qu, “An SLGC Model for Asian Food Image Classification,” *Journal of Computer and Communications*, vol. 08, no. 04, pp. 26–43, 2020, doi: 10.4236/jcc.2020.84003.
- [64] R. K. MishraG. Y. S. Reddyand H. Pathak, “The Understanding of Deep Learning: A Comprehensive Review,” *Math Probl Eng*, vol. 2021, 2021, doi: 10.1155/2021/5548884.
- [65] Y. Guo *et al.*, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016, doi: 10.1016/j.neucom.2015.09.116.
- [66] A. MosaviS. Ardabiliand A. R. Várkonyi-Kóczy, “List of Deep Learning Models,” *Lecture Notes in Networks and Systems*, vol. 101, pp. 202–214, 2020, doi: 10.1007/978-3-030-36841-8_20.
- [67] I. H. Sarker, “Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective,” *SN Comput Sci*, vol. 2, no. 3, 2021, doi: 10.1007/s42979-021-00535-6.
- [68] A. C. Review, “Mechanisms and Health Aspects of Food Adulteration :,” 2023.
- [69] S. Hochreiterand J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [70] F. K. zeyrek, Celal Tugrul Arpaci, Ozlem Temiz Arisoy, Mustafa Onurdag, “Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ’ s public news and information.,” *Psychiatry Res*, vol. 14(4), no. January, p. 293, 2020.
- [71] J. ChungC. GulcehreK. Choand Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” pp. 1–9, 2014.
- [72] A. MosaviS. Ardabiliand A. R. Várkonyi-Kóczy, “List of Deep Learning Models,” *Lecture Notes in Networks and Systems*, vol. 101, no. August, pp. 202–214, 2020, doi: 10.1007/978-3-030-36841-8_20.
- [73] V. H. Phungand E. J. Rhee, “A High-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets,” *Applied Sciences (Switzerland)*, vol. 9, no. 21, 2019, doi: 10.3390/app9214500.
- [74] R. Vanutre, “Food Image Classification Using Various CNN Models,” vol. 9, no. 3, pp. 626–632, 2022.
- [75] Purwono *et al.*, “Understanding of Convolutional Neural Network (CNN): A Review,” *International Journal of Robotics and Control Systems*, vol. 2, no. 4, pp. 739–748, 2022, doi: 10.31763/ijrcs.v2i4.888.

- [76] Y. RufT. Hanneand R. Dornberger, “Classification of Brand Images Using Convolutional Neural Networks,” *Lecture Notes in Networks and Systems*, vol. 648 LNNS, pp. 528–539, 2023, doi: 10.1007/978-3-031-27524-1_50.
- [77] M. M. Shanthi Raniand P. Shanmugavadivu, “12. Deep learning based food image classification,” *Comput Intell*, pp. 197–208, 2020, doi: 10.1515/9783110671353-012.
- [78] D. Shuklaand A. Desai, “Recognition of fruits using hybrid features and machine learning,” *International Conference on Computing, Analytics and Security Trends, CAST 2016*, pp. 572–577, 2017, doi: 10.1109/CAST.2016.7915033.
- [79] L. Zhu, “Using deep learning for food recognition,” 2020.
- [80] U. De Barcelona, “A multimodal deep learning approach for food tray recognition Autor: Joan Peracaula Prat Directors: Marc Bolaños and Petia Radeva Realitzat a: Departament de Matemàtiques i Informàtica Barcelona, 13 de setembre de 2020,” 2020.
- [81] Z. ZahishamC. P. Leeand K. M. Lim, “Food Recognition with ResNet-50,” *IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET 2020*, pp. 0–4, 2020, doi: 10.1109/IICAIET49801.2020.9257825.
- [82] D. J. AttokarenI. G. FernandesA. SriramY. V. S. Murthyand S. G. Koolagudi, “Food classification from images using convolutional neural networks,” *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, vol. 2017-Decem, pp. 2801–2806, 2017, doi: 10.1109/TENCON.2017.8228338.
- [83] S. Rahman, “Food Image Classification Using Convolutional Neural Network,” 2018.
- [84] A. ChaitanyaJ. Shettyand P. Chiplunkar, “Food Image Classification and Data Extraction Using Convolutional Neural Network and Web Crawlers,” *Procedia Comput Sci*, vol. 218, pp. 143–152, 2022, doi: 10.1016/j.procs.2022.12.410.
- [85] Y. LiX. FengY. Liuand X. Han, “Apple quality identification and classification by image processing based on convolutional neural networks,” *Sci Rep*, vol. 11, no. 1, pp. 1–15, 2021, doi: 10.1038/s41598-021-96103-2.
- [86] F. M. ShiriT. PerumalN. Mustaphaand R. Mohamed, “A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU,” 2023.
- [87] C. Tan *et al.*, “A survey on deep transfer learning,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11141 LNCS, pp. 270–279, 2018, doi: 10.1007/978-3-030-01424-7_27.
- [88] “Architecture comparison of AlexNet, VGGNet, ResNet, Inception, DenseNet | by Khush Patel | Towards Data Science.” Accessed: Dec. 19, 2023. [Online]. Available: <https://towardsdatascience.com/architecture-comparison-of-alexnet-vggnet-resnet-inception-densenet-beb8b116866d>

- [89] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," pp. 01–14, 2020, doi: 10.5121/csit.2020.100801.
- [90] J. L. Joseph, V. A. Kumar, and S. P. Mathew, "Fruit classification using deep learning," *Lecture Notes in Electrical Engineering*, vol. 756 LNEE, no. 4, pp. 807–817, 2021, doi: 10.1007/978-981-16-0749-3_62.
- [91] R. Mustafa and P. Dhar, "A method to recognize food using GIST and SURF features," *2018 Joint 7th International Conference on Informatics, Electronics and Vision and 2nd International Conference on Imaging, Vision and Pattern Recognition, ICIEV-IVPR 2018*, pp. 127–130, 2019, doi: 10.1109/ICIEV.2018.8641072.
- [92] E. Jahani Heravi, H. Habibi Aghdam, and D. Puig, "A deep convolutional neural network for recognizing foods," *Eighth International Conference on Machine Vision (ICMV 2015)*, vol. 9875, no. Icmv, p. 98751D, 2015, doi: 10.1117/12.2228875.
- [93] A. Mikołajczyk and M. Grochowski, "2023 International Interdisciplinary PhD Workshop, IPhDW 2023," *2023 International Interdisciplinary PhD Workshop, IPhDW 2023*, pp. 117–122, 2023.
- [94] S. Alehegn, "IDENTIFICATION OF INJERA MIXTURE USING COMPUTER VISION," 2021.
- [95] R. Mustafa and P. Dhar, "A method to recognize food using GIST and SURF features," *2018 Joint 7th International Conference on Informatics, Electronics and Vision and 2nd International Conference on Imaging, Vision and Pattern Recognition, ICIEV-IVPR 2018*, no. 1, pp. 127–130, 2019, doi: 10.1109/ICIEV.2018.8641072.
- [96] E. Jahani Heravi, H. Habibi Aghdam, and D. Puig, "A deep convolutional neural network for recognizing foods," *Eighth International Conference on Machine Vision (ICMV 2015)*, vol. 9875, no. Icmv, p. 98751D, 2015, doi: 10.1117/12.2228875.
- [97] B. Kefale, "Evaluation of Injera Prepared from Composite Flour of Teff and Barley Variety," *Food Science and Quality Management*, vol. 97, pp. 1–4, 2020, doi: 10.7176/fsqm/97-01.
- [98] B. Engineering and P. Fulfillment, "Addis Ababa Institute of Technology School of Chemical and Bio Engineering Optimization of Drying Parameters for ' Moisture Reduced Teff Injera Making ' Using Response Surface Methodology By : Dessalegn Abit Advisor : Prof . Eduardo Ojito Cespedes Addis A," pp. 1–87, 2019.
- [99] E. Tsalera, A. Papadakis, M. Samarakou, and I. Voyiatzis, "Feature Extraction with Handcrafted Methods and Convolutional Neural Networks for Facial Emotion Recognition," *Applied Sciences (Switzerland)*, vol. 12, no. 17, 2022, doi: 10.3390/app12178455.
- [100] A. H. Havelaar *et al.*, "Burden of foodborne disease due to bacterial hazards associated with beef, dairy, poultry meat, and vegetables in Ethiopia and Burkina Faso, 2017," *Front Sustain Food Syst*, vol. 6, 2022, doi: 10.3389/fsufs.2022.1024560.

- [101] S. SridevyM. N. DeviM. Sankarand N. R. Moorthi, “Image pre-processing techniques utilized for the plant identification : A review,” vol. 12, no. 6, pp. 320–324, 2023.
- [102] M. KhoshdeliR. Congand B. Parvin, “Detection of nuclei in H&E stained sections using convolutional neural networks,” *2017 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017*, no. February, pp. 105–108, 2017, doi: 10.1109/BHI.2017.7897216.
- [103] Y. BoureauN. Le Rouxand F. Bach, “Ask the locals : multi-way local pooling for image recognition”.
- [104] M. CogswellF. AhmedR. GirshickL. Zitnickand D. Batra, “Reducing overfitting in deep networks by decorrelating representations,” *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, pp. 1–12, 2016.
- [105] L. J. P. Van Der MaatenE. O. Postmaand H. J. Van Den Herik, “Dimensionality Reduction: A Comparative Review,” *Journal of Machine Learning Research*, vol. 10, pp. 1–41, 2009, doi: 10.1080/13506280444000102.
- [106] Z. LiF. LiuW. YangS. Pengand J. Zhou, “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects,” *IEEE Trans Neural Netw Learn Syst*, vol. 33, no. 12, pp. 6999–7019, 2022, doi: 10.1109/TNNLS.2021.3084827.
- [107] F. F. FirdausH. A. Nugrohoand I. Soesanti, “Deep Neural Network with Hyperparameter Tuning for Detection of Heart Disease,” *Proceedings - 2021 IEEE Asia Pacific Conference on Wireless and Mobile, APWiMob 2021*, pp. 59–65, 2021, doi: 10.1109/APWiMob51111.2021.9435250.
- [108] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.
- [109] A. Aghaebrahimianand M. Cieliebak, “Hyperparameter tuning for deep learning in natural language processing,” *CEUR Workshop Proc*, vol. 2458, 2019.
- [110] I. CardozaJ. P. García-VázquezA. Díaz-Ramírezand V. Quintero-Rosas, “Convolutional Neural Networks Hyperparameter Tunning for Classifying Firearms on Images,” *Applied Artificial Intelligence*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2022.2058165.
- [111] I. CardozaJ. P. García-VázquezA. Díaz-Ramírezand V. Quintero-Rosas, “Convolutional Neural Networks Hyperparameter Tunning for Classifying Firearms on Images,” *Applied Artificial Intelligence*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2022.2058165.
- [112] E. BochinskiT. Senstand T. Sikora, “Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms,” *Proceedings - International Conference on Image Processing, ICIP*, vol. 2017-Septe, pp. 3924–3928, 2017, doi: 10.1109/ICIP.2017.8297018.

- [113] I. Kandeland M. Castelli, “The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset,” *ICT Express*, vol. 6, no. 4, pp. 312–315, 2020, doi: 10.1016/j.icte.2020.04.010.