# Machine learning in agriculture domain: A state-of-art survey

A SEMINAR REPORT

Submitted by

**ADITHYAN SURESH KUMAR**

**AJC21CS016**

to

the APJ Abdul Kalam Technological University
in partial fulfillment of the requirement for the award of the Degree

of

**Bachelor of Technology**

in

**Computer Science and Engineering**

Under the guidance of

**Ms. Resmipriya M G**

**Assistant Professor**

AMAL JYOTHI
COLLEGE OF ENGINEERING
AUTONOMOUS
KANJIRAPPALLY

**Department of Computer Science and Engineering**
**Amal Jyothi College of Engineering (Autonomous)**
**Kanjirappally-686518**

**November 2024**

**DEPT. OF COMPUTER SCIENCE AND ENGINEERING**

**AMAL JYOTHI COLLEGE OF ENGINEERING(AUTONOMOUS)**
**KANJIRAPPALLY**

# CERTIFICATE

*This is to certify that the seminar report entitled* ***" Machine learning in agriculture domain: A state-of-art survey "*** *submitted by* ***ADITHYAN SURESH KUMAR (AJC21CS016)*** *to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering during the year 2024-2025, is a bonafide work carried out by her under my guidance and supervision.*

Ms. Resmipriya M G
Internal Supervisor

Ms. Elisabeth Thomas    Ms. Shiney Thomas
Seminar Coordinators

Dr.Juby Mathew
Head of Department, CSE

# ACKNOWLEDGEMENT

First of all I sincerely thank the **Almighty GOD** who is most beneficent and merciful for giving me knowledge and courage to complete the seminar successfully.

I derive immense pleasure in expressing my sincere thanks to our Manager, **Rev. Fr. Boby Alex Mannamplackal**, to our Director(Administrator), **Fr. Dr. Roy Abraham Pazhaya-parampil** and to our Principal, **Dr. Lillykutty Jacob** for their kind co-operation in all aspects of my seminar.

I express my gratitude to **Dr. Juby Mathew**, HOD, Department of Computer Science and Engineering his support during the entire course of this seminar work.

I express my sincere thanks to **Ms. Resmipriya M G** , my internal guide and **Ms. Elisabeth Thomas** and **Ms. Shiney Thomas**, our Seminar Co-ordinators for their encouragement and motivation during the seminar.

I also express my gratitude to all the teaching and non teaching staff of the college especially to our department for their encouragement and help during my seminar.

Finally I appreciate the patience and solid support of my parents and enthusiastic friends for their encouragement and moral support for this effort.

**ADITHYAN SURESH KUMAR**

# ABSTRACT

Agriculture is one of the basic necessities of human beings' lives and also a prime source of employment across the world. In developing countries like India, agriculture forms the core of the economy, accounting for about 15.4of the nation's GDP. Agricultural processes can be basically divided into three phases: pre-harvesting, harvesting, and post-harvesting. Recent machine learning technology has re-birthed a revolution in this agricultural sector by increasing productivity and reducing losses in an agricultural cycle. It provides an in-depth survey of contemporary application of ML to help surmount challenges at all three stages of agriculture.

At the pre-harvesting stage, ML algorithms advise on the optimum crop selection, soil, and pest management practices for increased yields, thus reducing resource wastage. During harvesting, ML technologies allow for monitoring and automation of crop harvesting processes, increasing efficiency while reducing manual labor. The post-harvesting applications of ML hence review storage practices, quality control, and supply chain management to lessen losses, consequently improving market value.

Farmers, by applying machine learning, acquire advanced tools that not only make some recommendations to the farmer but also the insights to do more precise and efficient farming. The integration of ML in agriculture will support high-quality production while keeping a lesser demand for extensive human labor. The paper describes new developments in the applications of ML in agriculture, which have very great potential for use in farming practice to resolve significant challenges in each phase of the agricultural process. The paper, in this regard, goes on to take an in-depth review of the advancements made in an effort to underline the potential role of ML in increasing productivity and sustainability in agriculture.

# CONTENTS

# LIST OF TABLES

# ABBREVATIONS

1. AI - Artificial Intelligence
2. ML - Machine Learning
3. DL - Deep Learning
4. IoT - Internet of Things
5. CNN - Convolutional Neural Network
6. SVM - Support Vector Machine
7. YOLO - You Only Look Once

# CHAPTER 1

# INTRODUCTION

## 1.1 Overview

Agriculture is one of the world's oldest and most essential industries, providing food security, employment, and economic stability to billions of people. In many regions, particularly in developing countries, agriculture is the backbone of the economy, contributing significantly to the Gross Domestic Product (GDP) and supporting livelihoods. However, the agricultural sector faces numerous challenges that hinder its productivity and efficiency. These include climate variability, resource limitations, labor shortages, and inefficiencies in traditional farming practices. Moreover, farmers often lack access to advanced technologies and information, which limits their ability to make informed decisions, maximize crop yields, and minimize losses.

## 1.2 Motivation and Objectives

The motivation behind exploring machine learning in agriculture stems from the need to modernize and improve traditional farming practices to meet rising food demands. By leveraging ML, farmers can potentially increase productivity, reduce operational costs, and make better use of resources such as water, fertilizers, and pesticides. Additionally, ML provides the ability to anticipate environmental risks, such as pest infestations and extreme weather events, which can significantly impact crop yield and quality.

The primary objectives of this report are to review and categorize the current applications of machine learning in agriculture across the stages of pre-harvesting, harvesting, and post-harvesting. There is also a need to highlight the benefits of ML in optimizing agricultural practices, including predictive analytics, crop monitoring, disease detection, and automated harvesting. Identifying challenges and limitations associated with deploying ML in agriculture, such as data scarcity, high costs, and infrastructure requirements. Suggesting future research directions that could address the identified challenges, with a focus on developing scalable, cost-effective, and accessible ML solutions for smart farming. By examining the progress made in applying ML to agriculture, this report aims to underscore the transformative potential of these technologies in addressing both current and emerging challenges in global food production.

## 1.3 Report Organization

This section gives the organization of the report work. Overall, this report work gives a clear and thorough study about the existing study of the proposed system and gives a description about the implementation details. The second chapter gives a detailed description about the literature survey. Third chapter gives a brief idea about the implementation details including hardware and software requirements, outline of the proposed methodology. Fourth chapter depicts the Results and Discussions and fifth Chapter gives the Conclusion and Future Work.

# CHAPTER 2

# LITERATURE SURVEY

## 2.1 Introduction

The use of Machine Learning (ML) in agriculture has gained significant momentum, especially as the agricultural sector faces complex challenges, including labor shortages, climate change, and the need for enhanced productivity. ML techniques such as computer vision, deep learning, and IoT integrations have revolutionized traditional farming practices by offering data-driven insights for precision farming, which enables efficient resource management and minimizes environmental impact.

Early research in the field largely focused on enhancing crop yield and automating labor-intensive tasks such as soil preparation, seeding, and harvesting. However, as technology advanced, ML applications extended into predictive analytics for crop disease detection, yield forecasting, and automated grading and sorting of harvested products. These developments underscore a transformative approach within agriculture, transitioning from manual to highly automated processes that improve quality, efficiency, and sustainability.

Despite the advancements, deploying ML in agriculture presents unique challenges, especially in rural areas. These include limited access to high-quality datasets, infrastructure constraints, and the high cost of advanced technologies. Researchers have attempted to address these limitations through AutoML and transfer learning models, which reduce the complexity and resource intensity of implementing ML solutions in low-resource settings.

## 2.2 Pre-Harvesting Applications

The pre-harvesting stage includes activities such as soil preparation, seed selection, crop monitoring, irrigation management, and pest control. ML applications in this stage are instrumental in helping farmers make data-driven decisions about crop and resource management.

Soil quality is critical for crop productivity. Machine learning models, including Support Vector Machines (SVM), Extreme Learning Machines (ELM), and neural networks, have been applied to predict soil nutrient levels, pH balance, and moisture content. Studies show that using ML for soil classification and nutrient assessment can optimize fertilizer use, improve soil health, and enhance crop yield. For example, ELM models with Gaussian radial functions have achieved up to 80% accuracy in soil classification, and Least Squares Support Vector Machines (LS-SVM) models have been successful in predicting soil organic matter and pH.

Seed quality impacts germination rates and crop yield. Traditionally, seed quality assessment relies on manual methods prone to error. Machine vision and CNNs enable automated seed sorting and classification by analyzing shape, texture, and color features. For instance, researchers used CNNs to differentiate between high- and low-quality pepper seeds, achieving over 90% accuracy. Machine learning models also facilitate yield estimation by assessing seed viability and germination potential.

Early detection of pests and diseases is essential to minimize crop loss. Machine learning techniques, especially deep learning models like CNNs, detect diseases by analyzing leaf images. Real-time decision support systems combining IoT sensors with ML algorithms have been developed to monitor plant health and predict disease outbreaks, allowing targeted pesticide application. Studies comparing ML algorithms for pest detection have shown that DL models, such as AlexNet and VGG-19, significantly outperform traditional methods, achieving over 94% accuracy.

## 2.3 Harvesting Applications

Harvesting is a critical stage in agriculture where the accuracy and timing of crop collection directly impact profitability and quality. Machine learning (ML) applications in this stage primarily focus on ripeness detection, fruit quality assessment, and automation through robotic harvesting systems. These technologies help streamline the harvesting process, reduce labor costs, and improve crop quality by ensuring optimal timing.

Machine learning has enabled the development of automated harvesting robots that can identify ripe fruits and collect them efficiently[1]. Object detection models, such as You Only Look Once (YOLO) and Faster R-CNN, are commonly used in these robots to detect and locate fruits within the canopy in real time. For instance, in a study on automated fruit detection, YOLO was trained on thousands of apple and pear images, achieving over 90% accuracy in on-tree fruit detection. Such models allow robots to identify the precise location, size, and ripeness of fruits, making the harvesting process faster and more accurate.

Assessing ripeness is crucial for maximizing crop quality. Convolutional Neural Networks (CNNs) have proven highly effective in analyzing visual cues such as color, texture, and shape, which are indicators of ripeness. Models like VGG-16 and ResNet are widely used in detecting maturity levels in fruits such as tomatoes, bananas, and dates. In one study, VGG-16 achieved over 99% accuracy in classifying dates by maturity stage, helping farmers to determine the ideal harvest time[2]. This level of precision ensures that only ripe fruits are collected, enhancing overall marketability and reducing post-harvest waste.

Machine learning models, especially deep learning algorithms, have been integrated into fruit-picking robots to facilitate autonomous harvesting[3]. These robots use ML to differentiate between ripe and unripe produce, minimizing damage to crops during collection. For example, robots equipped with computer vision can capture images of fruit, process them using CNNs or feature extraction techniques, and then pick fruits that meet quality standards. In a study using robots with Single Shot Multibox Detectors (SSD) and stereo cameras, researchers achieved over 90% fruit detection accuracy, with the robot successfully harvesting fruits within a few seconds. These robots contribute to labor savings and offer a scalable solution for commercial orchards and farms[4].

Despite its potential, automated harvesting through ML faces challenges such as complex orchard structures, occluded fruits, and varying lighting conditions. Additionally, the cost and technical requirements of deploying ML models and robotic systems in rural and resource-constrained areas remain barriers. Further research is needed to develop more adaptable models that can work effectively in diverse environments and handle the variability in natural outdoor settings.

## 2.4 Post-Harvesting Applications

Post-harvesting tasks are essential for maintaining produce quality, optimizing storage, and ensuring that agricultural products meet market standards. ML applications in post-harvesting focus on grading, quality assessment, and shelf-life prediction to reduce waste and improve profitability. The use of ML in this stage allows for efficient sorting, grading, and monitoring, which are critical for both domestic and export markets.

Quality Assessment and Grading Systems One of the primary applications of ML in post-harvesting is quality grading, where fruits and vegetables are sorted based on attributes like size, shape, color, and texture. Convolutional Neural Networks (CNNs) are extensively used for image recognition in grading systems[5]. For instance, in a grading system for Cavendish bananas, CNNs achieved over 90% accuracy in classifying the bananas by size and quality. These automated grading systems ensure consistency, reduce human error, and save time, which is particularly beneficial in large-scale operations.

Shelf-Life Prediction ML models are also used to predict the shelf life of agricultural products by analyzing data from storage conditions such as temperature, humidity, and light exposure. For example, IoT sensors in storage facilities can provide real-time data on environmental factors, which is then processed by ML algorithms to predict spoilage risks. Predictive models for shelf-life estimation help farmers and distributors optimize storage conditions, reducing the likelihood of spoilage and extending the shelf life of produce.

Post-Harvest Disease Detection Detecting diseases and defects in fruits after harvest is crucial to prevent contamination and reduce losses. ML models, including CNNs and Support Vector Machines (SVM), are effective in identifying early signs of decay or disease. For example, some studies have employed CNNs to analyze individual spots or lesions on produce, detecting diseases with an accuracy improvement of 12% compared to traditional methods. Such precision in disease detection aids in removing affected produce before it impacts the quality of other items in storage.

Integrating ML with IoT has revolutionized storage management. By connecting IoT sensors to ML algorithms, real-time monitoring of temperature, humidity, and gas composition becomes possible. This information allows ML models to provide insights on when adjustments are needed to maintain optimal storage conditions. For example, ML-based storage management systems in fruit storage facilities monitor conditions to retain freshness and quality, resulting in fewer losses during transport and storage.

## 2.5 Gaps in Existing Research and Challenges

productivity and sustainability, various research gaps and practical challenges limit its widespread adoption. These challenges impact the ability of ML to effectively address real-world agricultural issues, particularly in rural and resource-constrained settings where agricultural technology adoption has traditionally lagged.

One of the primary challenges is the quality and availability of agricultural data. Machine learning models rely heavily on large and diverse datasets to capture the complexities of agricultural environments accurately. However, many regions lack structured, high-quality datasets for crop characteristics, soil conditions, weather patterns, and pest behavior. Even when data is available, it is often fragmented, noisy, or incomplete, resulting in models that struggle to generalize effectively across different locations and crop types. Additionally, data on crop diseases, soil nutrients, and weather conditions can vary greatly across regions, making it difficult to develop universally applicable models. The lack of accessible, region-specific datasets is a significant limitation that reduces the potential of ML applications to deliver reliable insights and recommendations.

Another critical issue is the high cost and technical complexity associated with deploying ML models in real-world agricultural environments. While ML algorithms perform well in controlled or laboratory settings, they often encounter difficulties when applied in open agricultural fields, which present unique challenges, such as varying lighting, occluded crops, and diverse backgrounds. These conditions can impact the performance of image-based models, like those used for crop and disease detection. Additionally, rural farming communities may lack the necessary infrastructure, such as reliable internet access, to support real-time data collection and ML model deployment. This creates a barrier to implementing IoT-enabled solutions that rely on connectivity to transmit and process data, limiting the reach of ML in agriculture.

The choice and implementation of ML algorithms present additional challenges. In many cases, agricultural tasks require customized ML models that are optimized for specific crops, regions, and environmental factors[6]. However, selecting and tuning the right model is complex and often requires specialized knowledge, which may not be readily available in agricultural communities. Furthermore, traditional machine learning models are computationally intensive, necessitating advanced hardware and technical expertise. These requirements are impractical for many farmers, especially small-scale farmers who operate on limited budgets. Although advancements in AutoML and lightweight models have begun to address some of these issues, further progress is needed to make ML models more accessible and efficient for broader agricultural use.

Deployment and scalability also pose significant obstacles. For ML models to be valuable on a large scale, they must perform reliably across diverse agricultural landscapes and crop varieties, adapting to regional differences in climate, soil type, and crop characteristics. However, existing models are often trained on limited datasets that do not capture these variations, leading to issues with model generalizability. Additionally, maintaining and updating ML models to reflect changes in agricultural practices or environmental conditions is challenging, especially in low-resource settings where model retraining and continuous updates are not feasible.

Finally, the integration of ML with existing agricultural practices and the usability of MLpowered tools remain areas that require more attention[9][10]. Many current ML applications are developed in isolation from the end-users—farmers and agricultural workers—who may have limited experience with digital technology. Consequently, tools based on complex ML models may not be user-friendly or intuitive, reducing their practical utility. Developing ML solutions that are easy to interpret, provide actionable insights, and integrate seamlessly into traditional farming practices will be crucial for encouraging adoption at the grassroots level.

Table 2.1: Summary of Earlier Works

| **Study** | **Method/Approach** | **Key Findings** |
| --- | --- | --- |
| Siqueira VS (2021) | Systematic review of AI in echocardiogram analysis. | Explored AI for automatic medical decision support. |
| Karie NM (2019) | Cognitive computing in cyber forensics. | Used deep learning for improved cyber forensics methodologies. |
| Meshram VA (2021) | Assistive devices for visually impaired. | Developed device for object recognition and mobility support. |
| Arah IK (2015) | Mini-review on pre/post-harvest factors for tomatoes. | Identified factors affecting tomato quality and shelf life. |
| Sujatha R (2021) | Comparison of ML vs. DL in citrus disease detection. | Found DL methods more effective for citrus disease classification. |

# CHAPTER 3

# METHODOLOGY

## 3.1 Overview

The methodology in this study is structured into three main stages: pre-harvesting, harvesting, and post-harvesting. The focus is on leveraging precision agriculture techniques and automated systems to enhance crop yield and quality at each stage. By applying machine learning (ML), deep learning (DL), and computer vision techniques, the study seeks to optimize agricultural processes and improve the accuracy of tasks such as soil analysis, crop monitoring, and automated fruit grading. These data-driven systems are designed to aid farmers in minimizing resource use and maximizing output through timely insights and efficient operations.

## 3.2 Pre-Harvesting Stage

### 3.2.1 Soil Analysis

Soil analysis begins with evaluating essential properties such as soil reaction (pH), organic carbon (OC), boron (B), phosphorus (P), and potassium (K), which are critical for determining fertility and crop suitability[11]. The methodology uses Extreme Learning Machines (ELM) with Gaussian radial basis functions to classify soil properties effectively. This approach achieved an accuracy rate of 80% in assessing soil fertility, enabling targeted fertilization strategies based on soil conditions.

### 3.2.2 Seed Quality Assessment

The assessment of seed quality focuses on analyzing visual features like color, shape, and texture to evaluate seed viability and potential yield. Algorithms including K-Nearest Neighbor (KNN), support vector machine (SVM), logistic regression, and GoogLeNet are applied to distinguish high-quality seeds. For maize seeds, GoogLeNet demonstrated a classification accuracy of 95%, indicating its effectiveness in selecting seeds with the highest growth potential[15].

### 3.2.3 Fertilizer Application and Irrigation

To optimize growth, the study employs predictive models for fertilizer application and irrigation based on nitrogen and moisture levels in the soil. This method ensures that the necessary resources are delivered in optimal amounts, reducing waste and supporting sustainable growth practices. The nitrogen and moisture data gathered through soil samples enable accurate adjustments in fertilizer and irrigation levels.

## 3.3 Harvesting Stage

### 3.3.1 Fruit Detection

Fruit detection employs advanced computer vision models such as YOLO (You Only Look Once) and RetinaNet to facilitate the real-time localization of harvest-ready fruits directly on the trees. These models are particularly advantageous in detecting and differentiating between ripe and unripe fruits by analyzing features like color, shape, and spatial location within the orchard. The YOLO model, a popular object detection algorithm known for its speed and accuracy, enables effective scanning of large areas. In this study, it achieved over 90% accuracy for detecting apples and pears, which makes it highly effective for large-scale applications in orchards. The high accuracy and speed of YOLO and RetinaNet allow for continuous and reliable detection, which reduces manual oversight and significantly accelerates the harvesting process[1].

### 3.3.2 Automated Harvesting

The automated harvesting process combines deep learning with robotics to further improve efficiency[2]. Specifically, a Single Shot MultiBox Detector (SSD) model is paired with a robotic arm, creating an integrated system that performs both fruit localization and physical harvesting. The SSD model enables the robotic arm to detect the precise location of the fruit, after which the robotic arm performs the picking motion. This system achieves a 90% detection rate for apples, and the robotic arm completes each fruit harvest within 16 seconds. By automating the picking process, this model minimizes dependency on manual labor, reduces human error, and ensures uniformity in fruit handling. The robotic system is capable of adjusting to various types of fruit, optimizing for different sizes, textures, and heights within the orchard[4]. This adaptability is particularly valuable in minimizing bruising or damage to the fruit, ensuring that only high-quality produce reaches subsequent stages.

### 3.3.3 Quality Evaluation

The quality evaluation of fruits at harvest employs models like VGG-16, a convolutional neural network known for its image classification capabilities. VGG-16 analyzes fruit characteristics such as size, color, firmness, skin texture, and maturity level, classifying them into quality categories suitable for market distribution. By using deep learning models, the evaluation process becomes highly accurate, with the VGG-16 model achieving over 99% accuracy in grading apples[7]. This ensures that only top-quality fruits proceed to market, while sub-standard fruits are identified and separated for alternative uses. The model's high accuracy rate is particularly beneficial for producers who require consistent quality standards for different market segments. Additionally, the quality evaluation system reduces subjectivity and variability that can arise from manual inspections, providing more consistent results across different batches of fruits[12].

## 3.4 Post-Harvesting Stage

### 3.4.1 Grading and Sorting

The grading and sorting of fruits after harvest utilize machine learning models to classify produce by attributes such as ripeness, color, shape, and defect presence. This study applies models like linear-SVM, quadratic-SVM, and random forest classifiers to automate the grading process, achieving a 97% accuracy in categorizing tomatoes. Grading levels range from premium quality for direct market sale to lower grades for processing or secondary markets. The ML algorithms process large volumes of images, using deep learning to identify subtle differences in the fruit's appearance, such as color intensity and blemishes[13]. By implementing this automated grading system, the post-harvest process becomes more efficient and less labor-intensive, ensuring consistent quality standards across large volumes of produce. These quality checks also enable producers to command higher prices in the market by meeting strict grading standards.

### 3.4.2 Classification by Size and Defect Levels

In addition to grading, the methodology applies classification models to sort fruits by size and detect surface defects. The K-means clustering algorithm and C4.5 decision tree classifier are used to categorize apples in this study, achieving a sorting accuracy of 79% on a dataset of 183 images. This system groups fruits based on size and identifies minor to significant defects such as bruising, discoloration, or signs of early spoilage. By clustering fruits according to specific size categories, the system enhances uniformity in packaging, which is important for both presentation and pricing in retail[14]. The detection of surface defects further reduces waste by enabling immediate processing or alternative use for lower-grade fruits. This stage ensures that consumers receive only the highest quality produce, and it helps producers manage inventory based on the quality and condition of the harvest.

### 3.4.3 Storage Management

The storage management component of the methodology applies neural networks and image recognition techniques to monitor and predict fruit spoilage. This model assesses parameters such as color changes, texture degradation, and firmness to predict when fruits are likely to spoil. By analyzing these features, the system provides timely alerts for fruits nearing the end of their shelf life, allowing for optimal inventory rotation and storage conditions. This predictive approach to storage management not only reduces post-harvest waste but also maximizes the profitability of the stored produce by preventing losses associated with over-ripening or spoilage[8]. The automated monitoring ensures that fruits remain within acceptable quality standards for longer periods, supporting both short- and long-term storage solutions. By optimizing storage management, the methodology enables producers to reduce waste, maintain market standards, and extend the availability of fresh produce.

# CHAPTER 4

# FINDINGS AND ANALYSIS

1. **Pre-Harvesting Findings**

   The pre-harvesting analysis revealed that soil analysis and seed quality assessment, supported by models like Extreme Learning Machines (ELM) and GoogLeNet, effectively improved crop yield potential by optimizing early-stage agricultural inputs. The soil analysis models achieved an accuracy of 80% in determining soil fertility through the classification of essential properties like pH and organic carbon levels. Similarly, seed quality models demonstrated 95% accuracy in categorizing viable seeds. These findings highlight the importance of early data-driven interventions in predicting crop performance and guiding farmers in precise input applications. Fertilizer application and irrigation strategies further contributed to efficient resource use. By using predictive models based on soil nitrogen and moisture data, the system could provide tailored fertilizer and water levels that minimized excess use. This pre-harvest analysis underscores the potential of targeted interventions to reduce costs and support sustainable agriculture.

2. **Harvesting Findings**

   In the harvesting stage, real-time fruit detection models such as YOLO and RetinaNet showed a high detection accuracy of over 90% for on-tree fruits like apples and pears. This significantly reduced the need for manual labor and expedited the harvest process. Automated harvesting systems, combining Single Shot MultiBox Detector (SSD) models with robotic arms, achieved a 90% success rate in fruit localization and harvested fruits within 16 seconds each. This automation not only improved labor efficiency but also minimized damage to the produce, ensuring that high-quality fruits were picked. The quality evaluation of fruits using the VGG-16 model, with an accuracy rate exceeding 99%, further highlighted the system's ability to ensure consistent quality standards. This stage's findings emphasize the advantages of using DL models in supporting large-scale, automated harvesting that meets market quality requirements.

3. **Post-Harvesting Findings**

   In post-harvesting processes, the grading and sorting of produce by quality, size, and defect levels contributed significantly to quality assurance. Machine learning models like linear-SVM and random forest classifiers achieved a grading accuracy of 97% for tomatoes, providing effective quality categorization for market pricing. Classification models, including K-means clustering and decision trees, demonstrated an accuracy of 79% in categorizing fruits by size and defect levels, which streamlined inventory sorting. For storage management, the predictive models used for spoilage assessment helped optimize storage conditions by monitoring indicators such as color, texture, and firmness. The neural network-based storage management system provided early alerts for fruits nearing the end of their shelf life, which effectively reduced post-harvest losses. This finding underscores the role of predictive monitoring in extending the shelf life of produce and maximizing its profitability.

# CHAPTER 5

# CONCLUSION AND FUTURE WORK

This study explored the application of machine learning and deep learning methodologies across critical agricultural stages, including pre-harvesting, harvesting, and post-harvesting. The findings demonstrate the significant potential of these technologies in transforming traditional farming practices into data-driven, precision agriculture systems. By leveraging models like ELM, YOLO, and VGG-16, the study achieved high accuracy in tasks such as soil fertility classification, fruit detection, automated harvesting, and quality grading.

The pre-harvesting stage findings highlighted the value of early interventions in soil and seed analysis to optimize crop growth potential. In the harvesting stage, automated detection and harvesting systems reduced labor requirements while maintaining quality standards. Finally, the post-harvesting processes of grading, sorting, and storage management demonstrated the role of machine learning in enhancing product quality and extending shelf life. Future work could explore additional machine learning models and advanced deep learning architectures to improve accuracy and reduce computational requirements. Transfer learning and AutoML approaches could further support small-scale farmers by reducing the complexity and time required for model deployment.

Additionally, the integration of Internet of Things (IoT) devices with ML models presents an opportunity for real-time data collection, enhancing decision-making processes and providing farmers with instant insights into crop health, soil conditions, and market trends. Another area for future research is optimizing ML models for low-resource environments, such as mobile devices and embedded systems, to enable widespread adoption among farmers. As agricultural needs and technologies evolve, there will also be a need for improved datasets and collaboration across agricultural research platforms to ensure model accuracy and adaptability in diverse environments. Addressing these areas can further enhance the role of technology in achieving sustainable and resilient agricultural practices.

# REFERENCES

[1] **Ucat RC, Dela Cruz JC**, Postharvest grading classification of cavendish banana using deep learning and tensorflow. In: 2019 International Symposium on Multimedia and Communication Technology (ISMAC); 2019. p. 1–6. 10.1109/ISMAC.2019.8836129. 10 V. Meshram, K. Patil, V. Meshram et al. Artificial Intelligence in the Life Sciences 1 (2021) 100010

[2] **Hossain MS, Al-Hammadi M, Muhammad G**, Automatic fruit classification using deep learning for industrial applications. IEEE Trans Ind Inf 2019;15(2) 10.1109/TII.2018.2875149

[3] **Altaheri H, Alsulaiman M, Muhammad G**, Date fruit classification for robotic harvesting in a natural environment using deep learning. IEEE Access 2019;7:117115–33 10.1109/ACCESS.2019.2936536

[4] **Onishi Y, Yoshida T, Kurita H, Fukao T, Arihara H, Iwai A**, An automated fruit harvesting robot by using deep learning. Robomech J 2019;6:13 https://doi.org/10.1186/s40648-019-0141-2

[5] **X. Liu, S.W. Chen, S. Aditya, N. Sivakumar, S. Dcunha, C. Qu, C.J. Taylor, J. Das, and V. Kumar**, "Robust Fruit Counting: Combining Deep Learning, Tracking, and Structure from Motion", arXiv:1804.00307v2 [cs.CV] 2 Aug 2018.

[6] EU, Fruit and Vegetables: Marketing Standards, 2011, https://ec.europa.eu/agriculture/fruit-and-vegetables/marketing-standards_en (Accessed: July 2021). .

[7] **Esguerra Elda B**, 'Post-harvest management of mango for quality and safety assurance guidance for horticultural supply chain stakeholders. Rome: Food and Agriculture Organization of The United Nations; 2018.

[8] **Kirk R, Cielniak G, Mangan M**, LabFruits: a Rapid and Robust Outdoor Fruit Detection System Combining Bio-Inspired Features with One-Stage Deep Learning Networks. Sensors 2020;20:275 10.3390/s20010275.

[9] **Jr Piedad E, Larada JI, Pojas GJ, Vithalie L, Ferrer V**, Postharvest classification of banana (Musa acuminata) using tier-based machine learning. Postharvest Biol Technol 2018;145:93–100 https://doi.org/10.1016/j.postharvbio.2018.06.004.

[10] **Lia J, Chena L, Huanga W**, Detection of early bruises on peaches (Amygdalus persica L.) using hyperspectral imaging coupled with improved watershed segmentation Algorithm. Postharvest Biol Technol 2018;135:104–13 http://dx.doi.org/10.1016/j.postharvbio.2017.09.007.

[11] **Sofu MM, Er O, Kayacan MC, Cetisli B**, Design of an automatic apple sorting system using machine vision. Comput Electron Agric 2016;127:395–405 http://dx.doi.org/10.1016/j.compag.2016.06.030.

[12] **United states department of Agriculture (USDA) Grade standards for Fruits**, https://www.ams.usda.gov/grades-standards/fruits (Accessed: July 2021).

[13] **Ireri D, Belal E, Okinda C, Makange N, Ji C**, A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. Artif Intell Agricul 2019;2:28–37 https://doi.org/10.1016/j.aiia.2019.06.001.

[14] **Gaetano Paltrinieri**, Handling of fresh fruits, vegetables and root crops -a training manual- for grenadaAgricultural marketing improvement TCP/GRN/2901. Food and Agriculture Organization of the United Nations; 2021. http://www.fao.org/3/au186e/au186e.pdf. accessed:July.

[15] **Arsenovic M, Karanovic M, Sladojevic S, Anderla A, Stefanovic D**, Solving current limitations of deep learning based approaches for plant disease detection. Symmetry (Basel) 2019;11:939 10.3390/sym11070939.

# APPENDIX A

# PRESENTATION SLIDES

# Machine learning in agriculture domain

A STATE-OF-ART SURVEY

Guided by:
Ms. Resmipriya M G
Assistant Professor
Dept. of CSE

Adithyan Suresh Kumar
S7 CSE A
Roll no:16 (AJC21CS016)
25/09/2024

# Contents



# Contents



# ABSTRACT

Agriculture remains a fundamental sector globally, particularly in developing countries like India, where it contributes 15.4% to the GDP and serves as a key source of employment. The agricultural process is broadly divided into three phases: pre-harvesting, harvesting, and post-harvesting. Each phase faces its own challenges, ranging from resource inefficiencies to crop losses. However, recent advancements in machine learning (ML) have introduced transformative changes, helping farmers optimize production, reduce losses, and increase sustainability.

In the pre-harvesting stage, ML assists with crop selection, soil and pest management, and irrigation optimization, ensuring better use of resources and higher yields. During the harvesting stage, ML technologies are used to automate processes, allowing for real-time monitoring of crop maturity, which increases efficiency and reduces reliance on manual labor. Post-harvesting, ML improves quality control, storage practices, and supply chain management, which helps reduce spoilage and enhance product marketability.

By integrating ML, farmers gain access to advanced tools that offer predictive insights and recommendations, enabling more precise farming practices. This paper reviews the latest developments in ML applications for agriculture, underlining its potential to improve productivity, profitability, and sustainability. The findings emphasize the growing role of ML in transforming traditional farming into a more data-driven, efficient, and sustainable practice.

1

# Introduction

Agriculture is a vital sector, contributing 15.4% to India's GDP and providing essential employment. However, traditional farming methods face significant challenges, such as inefficient resource use, poor soil management, and unpredictable weather, which lead to lower yields and increased costs. Farmers often lack real-time data on soil conditions, market trends, and crop health, making decision-making difficult.

Machine learning (ML) offers solutions by providing precise, data-driven insights that can optimize every stage of farming. From selecting the best crops and monitoring soil conditions to predicting harvest times and grading produce, ML enables farmers to improve efficiency, reduce waste, and enhance profitability. With the growing demand for sustainable and scalable agricultural practices, ML is emerging as a crucial tool that helps address these challenges and transforms traditional farming into a more advanced, technology-driven process.

2

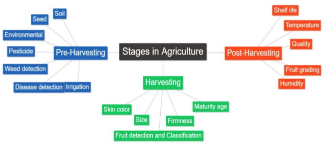
# Challenges in Traditional Farming

Traditional farming suffers from inefficiencies:
- **Soil Management**: Inadequate knowledge of soil properties.
- **Weather and Irrigation**: Poor irrigation scheduling and unpredictable weather patterns.
- **Pest Control**: Inefficient pesticide usage often results in overuse or underuse.
- **Market Awareness**: Lack of real-time market trends affects the timing of sales and crop choices. These challenges drive up costs and reduce productivity.

3

# Machine Learning in Agriculture

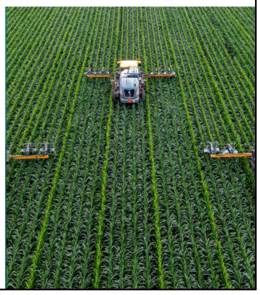
Machine learning enables smarter, data-driven farming by automating tasks and improving precision. Applications include:
- **Pre-harvesting**: Monitoring soil, optimizing irrigation, and recommending best crop practices.
- **Harvesting**: Predicting optimal harvest times based on crop maturity and quality.
- **Post-harvesting**: Grading and sorting produce to maximize value and reduce spoilage.




4

# Pre-harvesting Stage Overview

Pre-harvesting involves activities like soil preparation, seed selection, and irrigation management. Success in this stage directly impacts final crop yield. ML helps farmers:
- Monitor soil health.
- Predict weather patterns for better irrigation.
- Apply pesticides with precision, minimizing waste and crop damage.



# ML Applications in Pre-harvesting

In the pre-harvesting stage, ML is used to:

- **Soil Analysis**: ML models analyze soil fertility indices such as pH, organic carbon, and nutrient content.
- **Irrigation Optimization**: Predicting weather patterns allows for optimal irrigation, reducing water waste.
- **Pesticide Application**: ML helps in the precise application of pesticides, reducing environmental harm and crop damage.



# Soil Parameter Monitoring with ML

Machine learning plays a key role in monitoring soil conditions, including:

- **Soil Fertility**: Models like Extreme Learning Machine (ELM) predict soil fertility, helping farmers reduce fertilizer use while maintaining soil health.
- **Moisture and Temperature Monitoring**: ML estimates soil moisture and temperature, improving irrigation practices and crop growth.

| Soil Parameter | Machine Learning Model | Application | Accuracy Example |
|---|---|---|---|
| pH | Extreme Learning Machines (ELM) | Predicting soil fertility | ELM with Gaussian RBF: 80% accuracy |
| Organic Carbon | Cubist Regression | Estimating soil moisture | N/A |
| Moisture Content | Neural Networks | Determining soil temperature | N/A |



# Seed Quality Classification

ML is transforming seed quality assessment through automation:

- **Machine Vision and Neural Networks**: These models classify seeds based on factors like color, size, and texture, improving seed selection accuracy.
- **Germination Prediction**: ML predicts the success rate of seed germination, helping farmers choose the most viable seeds for planting, which improves overall yield.

1 Precision Farming
Utilizes data-driven insights to optimize planting and harvesting, maximizing yield.

2 Crop Monitoring
Employs sensors and drones to monitor crop health in real-time, enabling timely interventions.

3 Yield Prediction
Predicts future crop yields using historical data and algorithms, aiding in better planning.

4 Disease Detection
Identifies plant diseases early using image recognition, reducing crop loss and improving management.

5 Resource Management
Optimizes the use of water, fertilizers, and pesticides, promoting sustainability and reducing waste.



# Pesticides and Disease Detection

Accurate and timely disease detection is critical for preventing crop loss. ML offers:

- **Real-time Disease Detection**: By analyzing images of leaves, ML algorithms can detect early signs of disease.
- **Targeted Pesticide Use**: ML helps farmers apply pesticides only where necessary, reducing costs and environmental impact.

9

# Harvesting Stage Overview

Harvesting is time-sensitive, as crop maturity determines market quality. ML assists by:

- **Maturity Detection**: ML models assess the right time for harvesting based on crop size, color, and other maturity indicators.
- **Quality Control**: By integrating market data, ML ensures that harvesting aligns with market demands, improving profitability.

11

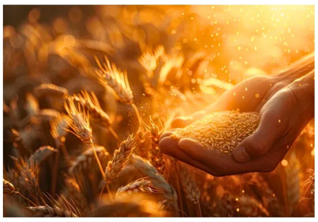
# Machine Learning in Harvesting

ML enhances the harvesting process by automating the detection of ripe crops:

- **Fruit Detection**: Using ML algorithms, farmers can identify ripe crops with high accuracy, optimizing harvest timing and reducing waste.
- **Labor Reduction**: ML reduces the reliance on human labor, making the harvesting process faster and more efficient.
- **Autonomous Crop Detection**: Robots equipped with ML can autonomously detect ripe crops and harvest them without human intervention.
- **Precision and Efficiency**: These robots are programmed to pick fruits without causing damage, ensuring that the highest quality produce is harvested.

12

# Post-harvesting Stage Overview



Post-harvesting activities are essential for maintaining crop quality. ML helps optimize:

- **Grading and Sorting:** ML systems categorize produce based on visual characteristics, such as size and color, ensuring consistent quality.
- **Storage Optimization:** By predicting the shelf-life of produce, ML helps farmers store crops under ideal conditions, minimizing spoilage.

13

# Challenges and Solutions in Agricultural ML

Implementing ML in agriculture faces several **challenges**:

- **Data Availability**: Many farms lack access to high-quality data needed to train ML models.
- **Long Training Times**: ML models often require high computational resources and time for training.
- **Deployment Challenges**: Limited internet connectivity and hardware constraints in rural areas make it difficult to implement ML effectively.

<u>**Solutions**</u>:

- **AutoML**: Automated Machine Learning (AutoML) simplifies the ML pipeline by automating data cleaning, model selection, and hyperparameter tuning. This allows farmers to adopt ML solutions without requiring expert knowledge.
- **Transfer Learning**: This technique can be used to reduce training times and improve model accuracy, especially when data is limited.

# Future of AI in Agriculture



The future of agriculture will be increasingly shaped by AI and ML. We can expect:

- **Increased Automation**: More robots and drones will handle tasks like planting, monitoring, and harvesting.
- **Sustainability Improvements**: AI will help optimize resource use, making farming more environmentally friendly.
- **Higher Yields**: With better insights into crop conditions and market demands, AI will help farmers improve yields while reducing costs.


# Conclusion



Machine learning is transforming agriculture by optimizing pre-harvesting, harvesting, and post-harvesting processes. It enables farmers to make data-driven decisions, improve yields, and reduce waste. By automating tasks like soil monitoring, crop quality control, and produce sorting, ML enhances efficiency and reduces labor costs. While challenges like data availability and deployment remain, solutions like AutoML simplify the adoption process. As ML continues to evolve, it will play a pivotal role in making agriculture more sustainable, productive, and profitable, ensuring better food security and resource management for the future.

# References


1. Siqueira VS, Borges MM, Furtado RG, Dourado CN, MCosta R. *Artificial Intelligence applied to support medical decisions for the automatic analysis of echocardiogram images: a Systematic Review.* Artif Intell Med 2021;120(5):102165 doi.org/10.1016/j.artmed.2021.102165.
2. Karie NM, Kebande VR, Venter HS. *Diverging deep learning cognitive computing techniques into cyber forensics.* Forensic Sci Int 2019;1:61–7 vol. 1, pp. 61-67doi.org/10.1016/j.fsisyn.2019.03.006.
3. Doowon J. *Artificial intelligence security threat, crime, and forensics: taxonomy and open issues.* IEEE Access 2020;8 184560-184574. 10.1109/ACCESS.2020.3029280.
4. Meshram VV, Patil K, Meshram VA, Shu FC. *An Astute Assistive Device for Mobility and Object Recognition for Visually Impaired People.* IEEE Trans Hum Mach Syst 2019;49(5):449–60 10.1109/THMS.2019.2931745.
5. Patil K, Jawadwala Q, Shu FC. *Design and construction of electronic aid for visually impaired people.* IEEE Trans Hum Mach Syst 2018;48(2):172–82 10.1109/THMS.2018.2799588.
6. Microsoft, *"What is automated machine learning (AutoML)?"* https://docs.microsoft.com/en-US/azure/machine-learning/concept-automated-ml,


# Acknowledgement

I would like to express my heartfelt appreciation to my teachers, whose guidance and support have been the cornerstone of this project. Your dedication to nurturing our intellectual growth and your willingness to share your extensive knowledge have been truly inspiring.

I would also like to extend a special thanks to my fellow students. Your role as a thoughtful and engaged audience has been invaluable.

Q & A

# APPENDIX B

# BASE PAPER

Artificial Intelligence in the Life Sciences 1 (2021) 100010

Contents lists available at ScienceDirect

Artificial Intelligence in the Life Sciences

journal homepage: www.elsevier.com/locate/ailsci

# Machine learning in agriculture domain: A state-of-art survey

Vishal Meshram[a], Kailas Patil[a,*], Vidula Meshram[a], Dinesh Hanchate[b], S.D. Ramkteke[c]

[a] Department of Computer Engineering, Vishwakarma University, Pune, India 411048
[b] Department of Computer Engineering, Pratishthan's Kamalnayan Bajaj Institute of Engineering & Technology, Baramati, Pune, Maharashtra, India
[c] ICAR-National Research Center for Grapes, Manjri Farm P.O., Pune-Solapur Road, Pune, India 412307

ARTICLE INFO

*Keywords:*
Deep learning
Harvesting
Machine learning
Post-harvesting
Pre-harvesting
Precision agriculture

ABSTRACT

Food is considered as a basic need of human being which can be satisfied through farming. Agriculture not only fulfills humans' basic needs, but also considered as source of employment worldwide. Agriculture is considered as a backbone of economy and source of employment in the developing countries like India. Agriculture contributes 15.4% in the GDP of India. Agriculture activities are broadly categorized into three major areas: pre-harvesting, harvesting and post harvesting. Advancement in area of machine learning has helped improving gains in agriculture. Machine learning is the current technology which is benefiting farmers to minimize the losses in the farming by providing rich recommendations and insights about the crops. This paper presents an extensive survey of latest machine learning application in agriculture to alleviate the problems in the three areas of pre-harvesting, harvesting and post-harvesting. Application of machine learning in agriculture allows more efficient and precise farming with less human manpower with high quality production.

## 1. Introduction

Agriculture is considered an important pillar of the world's economy and also satisfies one of the basic need of human being i.e. food. In most of the countries it is considered the major source of employment. Many countries like India still use the traditional way of farming, farmers are reluctant to use advanced technologies while farming because of either the lack of knowledge, heavy cost or because they are unaware about the advantages of these technologies. Lack of knowledge of soil types, yields, crops, weather, and improper use of pesticides, problems in irrigation, erroneous harvesting and lack of information about market trend led to the loss of farmers or adds to additional cost. Lack of knowledge in each stage of agriculture leads to new problems or increases the old problems and add the cost to farming. Growth in the population day by day also increases the pressure on the agriculture sector. Overall losses in the agriculture processes starting from crop selection to selling of products are very high. As per the famous saying "Information is the Power", keeping track of information about the crops, environment, and market, may help farmers to take better decisions and alleviate problems related to agriculture. Technologies like blockchain, IoT, machine learning, deep learning, cloud computing, edge computing can be used to get information and process it. Applications of computer vision, machine learning, IoT will help to raise the production, improves the quality, and ultimately increase the profitability of the farmers and associated domains. The Precision learning in the field of agriculture is very important to improve the overall yield of harvesting.

Blockchain technology, cloud computing, internet of things (IoT), machine learning (ML) and deep learning (DL) are the latest emerging trends in the computer field. It has been already used in different domains like healthcare, cybercrime, biochemistry, robotics, metrology, banking, medicine, food etc. to solve the complex problems by the researchers. Many applications of machine learning, IoT in different domains are presented [1–5]. Deep learning algorithms are making machine learning more powerful and accurate. By using automated machine learning (AutoML) one can cut the demand of ML experts, automate the ML pipeline with more accuracy.

While performing agriculture tasks the steps as below is generally followed by farmers.

Step 1: Selection of Crop
Step 2: Land Preparation
Step 3: Seed Sowing
Step 4: Irrigation & fertilizing
Step 5: Crop Maintenance [use of pesticides, crop pruning etc.]
Step 6: Harvesting
Step 7: Post-Harvesting activities

As per the above algorithm, the agriculture related tasks are categorized in the for major sub areas. Fig. 1 shows these four sub-domains of agriculture tasks.

* Corresponding author: Computer Science & Engineering, Vishwakarma University, Vishwakarma University, Kondhwa (Bk), Pune, Maharashtra 411048, India.
*E-mail address:* kailas.patil@vupune.ac.in (K. Patil).

https://doi.org/10.1016/j.ailsci.2021.100010
Received 21 August 2021; Received in revised form 28 September 2021; Accepted 30 September 2021
Available online 2 October 2021

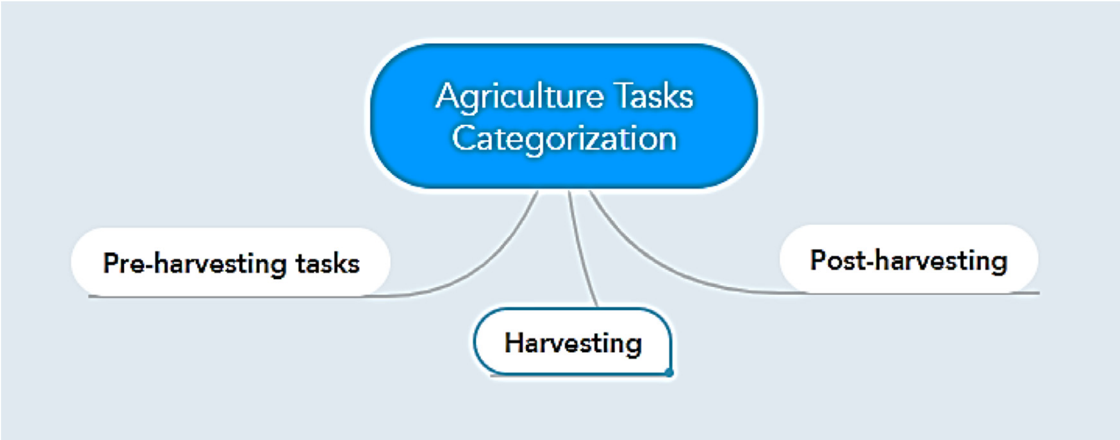



**Fig. 1.** General categorization of agriculture tasks.

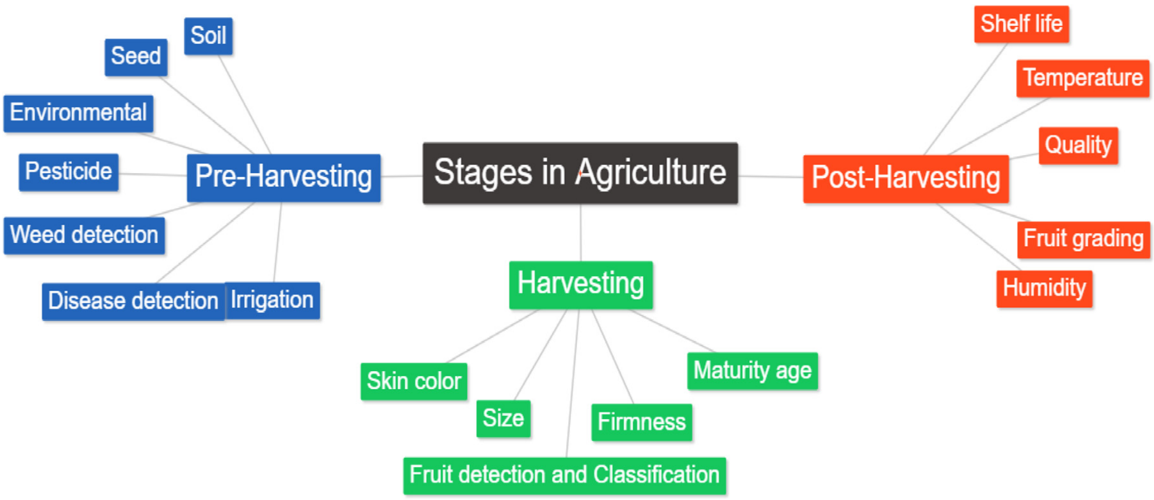



**Fig. 2.** Important parameters considered in each stage of farming.

**Table 1**
Important factors to be considered in each stage.

| S. No. | Stage | Activities / Factors | References |
|---|---|---|---|
| 1 | Pre-harvesting | Soil, seeds quality, fertiliser/pesticide application, pruning, cultivar selection, genetic and environmental conditions, irrigation, crop load, weed detection, disease detection. | [6, 7, 9] |
| 2 | Harvesting | Fruit/crop size, skin color, firmness, taste, quality, maturity stage, market window, fruit detection and classification. | [7] |
| 3 | Post-harvesting | Factors affecting the fruit shelf-life such as temperature, humidity, gasses used in fruit containers, usage of chemicals in postharvest and fruit handling processes to retain the quality, fruit grading as per quality. | [7] |

During pre-harvesting tasks farmers focuses on selection of crops, land preparation, seed sowing, irrigation, and crop maintenance which includes use pesticides, pruning etc. In yield estimation the farmers do the activities like yield mapping and counting the number of fruits so that they can predict the production and make the necessary arrangements required at the time of harvesting or post-harvesting. While harvesting farmers are focused on maturity of crops or fruits market need quality. Whereas in post-harvesting farmers are focused on post-harvest storage and processing systems. Fig. 2 shows the important factors that should be considered in each stage of farming. Table 1 summarizes few works in each stage of agriculture tasks.

The major branches of the agriculture are Agronomy, Horticulture, Forestry, Livestock, Fisheries, Agriculture Engineering and Economics. The scope of the paper is confined to use of machine learning in agriculture, specifically on fruits.

In the following sections, the review of the most recent techniques of machine vision systems used for classification and object detection in each stage of farming is presented. Section 2 explains the use of ML in the pre-harvesting stage. In Section 3, usage of ML in the stage of harvesting is explained and in Section 4 usage of ML in the post-harvesting stage is explained. Sections 5 and 6 focuses on discussion and challenges in use of the Artificial Intelligence (AI), ML, and DL.

## 2. Pre-harvesting

Pre-harvesting parameters play a key role in overall growth of crop/fruits. In pre-harvesting machine learning is used to capture the parameters of soil, seeds quality, fertilizer application, pruning, genetic and environmental conditions and irrigation. Focusing on each component it is important to minimize the overall losses in production. Here few important components in the pre-harvesting are considered and how neural networks and machine learning are used to capture the parameters of each component.

**Table 2**
Analysis of pre-harvesting parameter: Soil.

| S. No. | Property | Important features | Classes defined in the work | Dataset used (Public / Own) | Total number of images used for training | Models / Method / Algorithms compared | Best model / method/ algorithm | Results | Reference |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Soil | Village wise soil fertility indices of available Soil Reaction (pH), Organic Carbon (OC) and Boron (B), Phosphorus (P), and Potassium (K) | For P, K and OC three classes: Low, Medium, and High. For B six classes: Very Low, Low, Medium, Moderately High, High, and very High. For pH Four classes: Strongly Acidic (SA), Highly Acidic (HA), Moderately Acidic (MA), and Slightly Acidic (SLA). | public (reports available during the years 2014 to 2017) | NA | Extreme Learning Machine (ELM) with different activation functions like sine-squared, Gaussian radial basis, triangular basis, hyperbolic tangent, and hard limit | ELMs with Gaussian radial basis function | 80% of accuracy | [10] |
| 2 | | Soil Organic matter (SOM) and pH parameter | SOM and pH parameters | Own | 523 soil samples | four Machine Learning models Cubist regression model (Cubist), extreme learning machines (ELM), least squares-support vector machines (LS-SVM), and partial least squares regression (PLSR) | ELM | R2 = 0.81 | [11] |
| 3 | | Moisture content (MC), organic carbon (OC), and nitrogen (TN) | Estimating moisture content (MC), organic carbon (OC), and nitrogen (TN) | Own | 140 set | Cubist, partial least squares regression (PLSR), least squares support vector machines (LS-SVM), and principal component regression (PCR) | LS-SVM is best for MC and OC and TN is best by the Cubist | MC - RMSEP:0.457%, RPD:2.24 TN - RMSEP: 0.071 and RPD :1.96 | [12] |
| 4 | | soil moisture | Auto-regressive error function (AREF) combined with computational models | Own The soil moisture and density were determined by volumetric rings with 100 cm3 collected in eight positions along the plots, at depths from 25 mm to 75 mm | NA | One Neuro-Fuzzy model (ANFIS) and two artificial neural networks (a Multi-Layer Perceptron (MLP) and a Radial Basis Function (RBF)). Multiple linear regression (MLR) models with two and six independent variables | Neural Network with AREF | RMSE between 1.27% and 1.30%, R2 around 0.80, and APE between 3.77% and 3.75% | [13] |
| 5 | | Soil Temperature | soil temperature (ST) at 6 different depths of 5, 10, 20, 30, 50 and 100 cm | Public (For Bandar Abbas, 10 years measured data sets for the period of 1996–2005 and for Kerman, 7 years measured data sets for the period of 1998–2004) | NA | ELM, SaE-ELM, genetic programming (GP) and artificial neural network (ANN) | SaE-ELM | MABE - 0.8660–1.5338 C R - 0.9084–0.9893 | [14] |

### 2.1. Soil

Liakos, et al. [8] and Sharma, et al. [9] presented a soil management survey with the application of ML techniques for prediction or identification of soil properties (estimation of soil temperature, soil drying, and moisture content). The categorization and estimation of the soil attributes help farmers in minimizing extra cost on fertilizers, cut the demand of soil analysis experts, increase profitability, and improve health of soil, whereas Suchithra and Pai [10] presented pH values and soil fertility indices classification and predication model. Yang, et al. [11] observed that important indicators of soil fertility are pH values and Soil Organic matter (SOM) and thus the authors have done prediction of SOM and pH parameters in paddy soil. Morellos, et al. [12] has predicted organic carbon (OC), nitrogen (TN), and moisture content (MC) parameters of the soil. The aim of study was to compare machine learning algorithms and linear multivariate algorithms on basis of their performance of prediction. As soil moisture is frequently associated with variability in yield, Johann, et al. [13] have estimated the moisture content of soil using with Auto-regressive error function (AREF) along with machine learning algorithms. Nahvi, et al. [14] developed a new model by employing Self-adaptive evolutionary (SaE) agent in extreme machine learning (ELM) architecture. This new model is used for the assessment of daily soil temperature (ST) at 6 different depths of 5, 10, 20, 30, 50 and 100 cm. The detail summary of work done by different authors on soli parameter is mentioned in Table 2.

### 2.2. Seeds

Seed germination is a vital factor for quality of seed, which is an important determining factor of yield and quality of production. Seed germination rate calculation is still done manually with the help of trained persons which is not only a tiresome process but also prone to error. Thus, various machine leaning and image recognition techniques have been proposed by different authors to automate the process of seed sorting and calculation. Various computer vision, machine learning techniques, Convolution Neural Network (CNN) methods have been presented in D. Sivakumar, et al. [15], Huang, et al. [16], Zhu, et al. [17]. Image recognition technique for seed sorting with high accuracy is developed by Young, et al. [18]. Ke-ling, et al. [19] used a multilayer perceptron neural network model for improving the accuracy of the classification method to separate pepper seeds of high-quality from low-quality. Uzal, et al. [20] and Veeramani et al. [21] used the deep neural network (DNN) model using CNN for the assessment of the quantity of seeds per pod in soybean and for sorting of haploid seeds on basis of shape, phenotypic expression, and the embryo pose. Nkemelu, et al. [22], built a model using CNN for plant seedlings classification into 12 species. Medeiros, et al. [23] assessed the proficiency of computer vision as an alternative to routine vigor tests to expedite the process of accurate evolution of seed physiological potential. Amiryousefi, et al. [24] used image analysis technique, principal component analysis (PCA), to save time and cost of placing seeds in different clusters by reducing the features to be considered for clustering. Vlasov, et al. [25], Kurtulmuş, et al. [26] used machine learning (ML) techniques for efficient seed classification. The detail summary of work done by different authors is mentioned in Table 3.

### 2.3. Pesticides and disease detection

In-time disease detection is the most important task to save crops from major loss. Some farmers regularly analyze leaf or branches of tree while growing and identify the diseases or many times to avoid the diseases, they apply the pesticides on all the crops equally. Both the activities are based on human experience which is prone to errors and risky. Decision of which pesticide, when to apply and where to apply is totally dependent on type of disease, its stage and affected area. Application of unnecessary pesticide on all the crops may harm crops as well as farmer's health. Precision agriculture helps farmers for application of the right pesticide at right time at right place. Many works combined pesticides prediction with the detection of disease on plants. This section discusses bout disease detection using machine learning.

Alagumariappan, et al. [27], developed a real-time decision support system integrated with a camera sensor module for plant disease identification. In this work authors evaluated the performance of three machine learning algorithms namely, Extreme Learning Machine (ELM) and Support Vector Machine (SVM) with linear and polynomial kernels and observed that the performance of ELM is better when compared to other algorithms. Savary, et al. [28] studied how diseases cause the crop losses and their implications for global food production losses and food security. The objective of this work is to show that crop loss research is vital and should be consider as full branch of plant science.

Sujatha, et al. [29], compared the ML algorithms (SVM, RF, SGD) with DL algorithms (Inception-v3, VGG-16, VGG-19) in terms of citrus plant disease detection and observed that DL methods performed much better. Karadağ, et al. [30] studied detection of healthy and fusarium diseased peppers (capsicum annuum) from the reflections obtained from the pepper leaves with the help of spectroradiometer. Artificial Neural Networks (ANN), Naive Bayes (NB) and K-nearest Neighbor (KNN) machine learning algorithms were used for classification. Authors claimed that leaf reflections can be used in disease detection. Pandya [31], presented data about different types of pesticides, their applications and impact on environment. Arsenovic, et al. [32] discussed the shortcomings of available DL models used for plant disease detections. A novel model is built which consist of two-stage architecture Disease Net, for classification of plant disease, which achieved 93.67% training accuracy. Barbedo [33], explored the new approach by using DL to identify plant diseases from individual lesions and spots instead of considering entire leaf. This approach helps to detect multiple diseases on the same leaf with 12% higher accuracy. Saleem, et al. [34], presented a detail review of DL models used to envision different diseases of plant. Many research gaps have been enlisted in the plant disease detection and suggested that advanced DL algorithms should be used to increase the accuracy.

Liu, et al. [36], Kour, et al. [37] studied the apple leaf diseases and apple fruit diseases respectively. A CNN model was proposed to classify apple leaf diseases into Brown spot, Rust, Mosaic, and Alternaria leaf spot. A new dataset was created consisting of 13,689 images of diseased leaves which was used to train the novel architecture based on AlexNet in [34]. For apple disease detection and classification in Kashmir Valley, another model called Fuzzy Rule-Based Approach for Disease Detection (FRADD) was proposed in [35]. Though the accuracy of the model is good, it takes into account only one disease known as scab and limited numbers of fruit types. Xing, et al. [38] proposed a new model called Weakly DenseNet-16, to overcome the limitations of pre-trained models which are trained on ImageNet dataset. A dataset consisting of 17 species of citrus pests and seven types of citrus diseases (9051 images of citrus pests and 3510 images of citrus diseases) was created. Weakly DenseNet-16 performed well with the accuracy 93.33% as compared to MobileNet-v1 (85.04%), MobileNet-v2 (87.82%), ShuffleNet-v1 (83.44%), ShuffleNet-v2 (83.21%), NIN-16 (91.66%), SENet-16 (88.36%), and VGG-16 (92.93%). Doh, et al. [39] proposed a solution to detect the citrus fruit diseases using their physical attributes such as the texture, color, structure of holes on the fruit and morphology. The proposed solution composed of K-Means clustering technique, ANN and SVM algorithms. Results show that the use of SVM with ANN helps in increasing disease detection and classification rate. The detailed summary of the published works is presented in Table 4.

## 3. Harvesting

After taking care of parameters in pre-harvesting stage like soil, seeds, weeds etc., when the fruits/vegetables are ready then harvesting is the most important stage. The important parameters should be focused in this stage are fruit/crop size, skin color, firmness, taste, quality, maturity stage, market window, fruit detection and classification for harvesting. Careful and right harvesting of fruit is directly correlated with the profit. In the survey, we observed that auto-harvesting robots, machine learning, deep learning techniques are achieving better results and helping farmers in reducing the losses in harvesting stage. This section presents the application of ML, DL algorithms in the harvesting.

Hua, et al. [40] presented a detail survey on automated fruit harvesting systems for sweet pepper, tomato, apple and kiwifruit as an example to demonstrate the recent advances in intelligent automatic harvesting robots in horticulture. The use automatic robots in field helps to increase the production, saves the harvesting time which ultimately increase the profits of the farmers. Kushtrim, et al. [41] developed a CNN model based on single shot detector (YOLO) algorithm for on-tree fruit detection. A dataset consisting of real and synthetic images of apple and pear trees was created. For labeling the images, open-source labeling tool called as BBox-Label-Tool was used. More than 5000 images of pear and apple fruits were used while training the model. Amazon cloud platform was used to train the model. The model achieved more than 90% accuracy for on-tree fruit detection. Two deep neural network models were investigated in the proposed work, a small CNN model and a VGG-16

**Table 3**
Analysis of pre-harvesting parameter: Seed.

| Sr. No. | Property | Important features | Classes defined in the work | Dataset used (Public / Own) | Total no of images used for training | Models / Method / Algorithms compared | Best model / method / algorithm | Results | Model evaluation technique | Reference |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Seed | color, shape, and texture | maze seed | Own | 4000 | ensemble learning, K-nearest neighbor (KNN), logistic regression, support vector machine (SVM), and Speeded Up Robust Features (SURF) algorithm to classify the extracted features, GoogLeNet, VGG19 | GoogleNet | 95% | Confusion Table, Training loss. Testing loss. Training accuracy. Testing accuracy | [16] |
| 2 | Cotton Seed | | Jinxin5, Jinxi7, Shennongmian1, Xinjiangzaomian13, Xinluzao-mian29, Xinluzhong52 and Xinluzhong42 | own, dataset collected from Shihezi, Xinjiang Uyghur Autonomous Region, China | 13,160 | SVM, PLS-DA, and LR models based on deep features extracted by self-design CNN and ResNet models | self-design CNN | 80% | classification accuracy | [17] |
| 3 | pepper seeds | 15 features (ten color features: R, G, B, L*, a*, b*, hue, saturation, brightness, and Gray, three geometric features: width, length, and projected area, seed weight and density) | germinated seed (1) and un-germinated seed (0) | Own | 400 seeds | multilayer perceptron (MLP); BLR binary logistic regression, single feature models | multilayer perceptron and binary logistic regression | 90% | classification accuracy | [19] |
| 4 | soybean pods | 38 tailored features, geometrical characteristics (area, perimeter, major and minor axis length), shape features (density, elongation, ompactness, rugosity and axis ratio), first 4 Hu moments, and finally a 25 bins histogram of the profile of the pod straighten mask added along the short axis | 2-SPP, 3-SPP, and 4-SPP | Own | 18,178 | tailored features extraction (FE) followed by a Support Vector Machines (SVM), CNN | CNN | 86.20% | accuracy | [20] |
| 5 | haploid maize seeds | texture, morphology, color and shape | True-Diploid, True-Haploid | Own | 4021 | DeepSort, Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR) | DeepSort | 0.961 | 5-fold cross-validation | [21] |

**Table 4**
Analysis of pre-harvesting parameter: Pesticides and disease detection.

| Sr. No. | Property | Important features | Classes defined in the work | Dataset used (Public / Own) | Total no of images used for training | Models / Method / Algorithms compared | Best model / method / algorithm | Results | Model evaluation technique | Reference |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Disease detection | color, shape, and texture | 12 different species and 42 different classes (both healthy and diseased) | Own (PlantDisease) | 79,265 | AlexNet, VGG 19, Inception, DenseNet, ResNet, PlantDiseaseNet Object Detection: Two-Stage Methods - Faster R-CNN, Faster R-CNN with TDM, Faster R-CNN with FPN, One-Stage Methods - YOLOv3, SSD513, RetinaNet | PlantDiseaseNet | 94% | TOP-1 Accuracy | [32] |
| 2 | Plant disease | individual lesions and spots | Healthy, Mildly diseased, Moderately diseased, Severely diseased | Own (Plant-Disease) | PDDB - 1575 XDB - 46,409 | GoogLeNet CNN | GoogLeNet CNN | 12% higher | Confusion matrices | [33] |
| 3 | Plant disease and pest detection | deep features | 8 classes : 5 disease (Coryneum beijerinckii, Apricot monilia laxa, Peach monilia laxa, Cherry myzus cerasi, Xanthomonas arboricola); 3 pest (Walnut leaf mite ga, Peach sphaerolecanium prunastri, Erwinia amylovora) | Own | 1965 | extreme learning machine (ELM), support vector machine (SVM), and K-nearest neighbor (KNN), VGG16, VGG19, and AlexNet | ResNet50 model and SVM classifier | 98% | accuracy, sensitivity, specificity, and F1-score, confusion matrix | [35] |
| 4 | Apple Leaf Diseases | edge, corner, color, shape and object, | 4 classes: Brown spot, Rust, Mosaic, and Alternaria leaf spot | Own | 13,689 | AlexNet Precursor, VGG 19, Inception, DenseNet, ResNet, PlantDiseaseNet, SVM BP AlexNet GoogLeNet ResNet-20 VggNet-16 Our Work | AlexNet Precursor | 97.62% | confusion matrix | [36] |
| 5 | Apple Fruit Disease | background and foreground pixels | 4 classes: Poor, Average, Good, Excellent | Own (Two datasets) | NA | Fuzzy Rule-Based Approach for Disease Detection (FRADD) | FRADD | 91.66 | accuracy | [37] |

fine-tuned model to classify the fruits by Hossain, et al. [42]. The first model was built with six layers while the second was fine-tuned visual geometry group-16 pre-trained DL model. Two datasets were used to evaluate the performance of the proposed models. Dataset-1 is publicly available and it consists of 2633 color images whereas dataset-2 consists of total 5946 images, distributed among 10 classes. It was claimed VGG-16 fine-tuned model achieved excellent accuracy on both datasets. Kirk, et al. [43] studied on improving network performance on unseen data through a structured approach and analysis of the network input. Instead of modifying network architecture and increasing depth of neural network, the fusion of features was chosen. Result shows that the model complexity for more accuracy and generalization capabilities can be avoided by using bio-inspired features. It is claimed that for the color centric data classes this approach shows more promising results with the robust DL model in real world. For this the work author created dataset consists of 6189 images over 2 months, August and September 2018, and manually annotated 150 of them. Altaheri, et al. [44] proposed a machine vision system to categorize date fruit images according their maturity stages which help in harvesting decision. A dataset of 8072 images were created consisting of five date types: Naboot Saif, Khalas, Barhi, Meneifi, and Sullaj with different pre-maturity and maturity stages. The images were captured in various angles, scales, illumination conditions, and there were few occluded images. Transfer learning from two famous CNN models AlexNet and VGGNet were used to build the three classification models to classify date fruit according to their maturity stage, type, and whether they are harvestable or not. Result shows that VGG-16 model outperformed with the accuracy of 99.01% in 20.6 msec. Bauer, et al. [45] developed a platform that chains up-to-date ML techniques, modern computer vision, and integrated software engineering practices to measure yield-related phenotypes from ultra-large aerial imagery named as AirSurf. Author claims that this platform help to increase the yield and crop marketability before the harvest. Zhang, et al. [46] developed a harvesting robot for autonomous harvesting which consists of low priced gripper and ML technique for detection of cutting-point. The purpose of the study was to develop an autonomous harvester system which can harvest any crop with peduncle rather than damaging to its flesh. Onishi, et al. [47] proposed a new system (robot arm) consisting of Single Shot MultiBox Detector (SSD) and stereo camera for autonomous detection and harvesting of fruits. The system was tested on apple tee called “Fuji”. Robot arm detects the harvestable fruit position and harvest it by twisting the hand axis. An experimental result shows that system was able to detect 90% fruits and took only 16 s for harvesting. Liu, et al. [48] proposed a novel pipeline consisting of segmentation, 3D localization and frame to frame tracking for accurately counting the fruits from order of images. This model was evaluated on orange and apple fruits dataset. Table 5, presented the detail summary of harvesting techniques.

## 4. Post-harvesting

Post-harvesting is last and most crucial area in agriculture which require more attention. After successfully completing all stages starting from yield-estimation till harvesting, negligence in post-harvesting may spoil all the efforts and cause severe loss to farmers. The subtasks that

**Table 5**
Analysis of harvesting techniques.

| Sr. No. | Property | Important features | Classes defined in the work | Dataset used (Public / Own) | Total no of images used for training | Models / Method / Algorithms compared | Best model / Method /Algorithm | Results | Model evaluation technique | Reference |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Real-Time Fruit Detection within tree | fruit shapes, color and/or other attributes | apple and pear fruits | own | 5000 | Single-Shot Convolution Neural Network (YOLO) | YOLO | 90% | confusion matrix. | [41] |
| 2 | fruit classification | NA | 1st dataset: 15 classes, 2nd dataset: 10 classes | 1st dataset: Public, 2nd Dataset: own | 1st dataset: 2633, 2nd dataset:5946 | 2 deep learning Models : 1) light model of six CNN layers and 2)VGG-16 based architecture | VGG-16 based architecture | 99.75% | Confusion matrix | [42] |
| 3 | Outdoor Fruit Detection | Bio-Inspired Features, fusion of features | 3 classes: Ripe Strawberry, Unripe Strawberry, Both Classes | own (DeepFruit) | 4219 | Feature Pyramid Networks, Residual Neural Networks and RetinaNet | L*a*b*Fruits system | performance increase of 6.6 times | F1 score, the harmonic mean of precision and recall | [43] |
| 4 | Date Fruit Classification | local and spatial features and patterns | five date types in different pre-maturity and maturity stages: Naboot Saif, Khalas, Barhi, Menei, and Sullaj | own | 8000 | VGG-16, AlexNet | VGG-16 | 99.01% | Confusion matrix. | [44] |
| 5 | fruit harvesting robot | NA | apples Detected, Undetected | public | 169 | Single Shot MultiBox Detector (YOLO) | YOLO | 0.9 | precision, recall | [47] |

can be consider in this stage are shelf-life of fruits and vegetables, post-harvest grading and export. Every country has their own standard rules and regulations for grading the fruits [49–51].

In [52], an information manual with directions for “Post-harvest management of mango for quality and safety assurance” was presented. This is very insightful for all the stakeholders of horticultural supply chain. Study showed that wrong post-harvest handling methods can affect the quality and quantity of fruits which increases the overall losses. 31% losses which are identified at retail level were caused by decay only. The other practices which add losses are poor harvesting, careless handling, and improper packaging and carriage conditions.

The wrong disease management during production causes the decay at high-level of pre-harvest infections. The decays in the form of anthracnose and stem end rot are very commonly observed. A training manual for “handling fresh fruits, vegetables and root crops” for Grenada was presented in [53], as a part of the “Agricultural Marketing Improvement” Project TCP/GRN/2901 which was implemented by Grenada Government and FAO. The goal of this project was to increase the profits for horticulture products and root crop growers through a well-organized agricultural marketing system. This document provides in detail study about all post-harvest stages with how to minimize the losses in every stage. Ucat, et al. [54] explored the use of image processing with deep leaning algorithm to classify Cavendish banana as per their grades. Python, OpenCV and Tensorflow were used to build the model to classify the bananas into different categories such as Class A big-hand or small-hand, Class B big-hand or small-hand and Cluster class (part of hand). Result shows that the model achieved more than 90% classification accuracy. Ireri, et al. [55] proposed a machine vision system for post-harvest tomato grading. The system works on RGB images given as an input to the system. Dataset was created by manually labeling the tomato images into four categories according to their defect, healthy and ripeness parameters. Four different models were built to classify image into one of the category according to the matching features, total 15 features were considered while taking the decision Result shows that RBF-SVM performed well as compared to others for category 1 i.e. healthy or defected with 0.9709 detection accuracy. Piedad, et al. [56] developed a system for banana (Musa acuminata AA Group ’Lakatan’) classification using ML techniques based on tier-based. A non-invasive tier-based technique was used in this study. ANN, SVM and RF classifiers were used to classify bananas into extra class, class I, class II and rejected classes. Result shows that the random-forest algorithm outperformed as compared to others with the 94.2% accuracy. Lia, et al. [57] studied and compared two hyper-spectral imaging technologies namely long-wave near infrared (LW-NIR) and short-wave near infrared (SW-NIR) for early identification of Bruise of ‘Pinggu’ peaches. An improved watershed segmentation algorithm based on morphological gradient reconstruction and marker extraction was developed and tested on multispectral PC images in this study. Experimental result shows that a proposed algorithm accurately classified 96.5% of the bruised and 97.5% of sound peaches respectively. An automated real–time grading system with quality inspection for apple fruit was developed by Sofu, et al. [58]. The developed system comprises a roller, transporter and class conveyors joined with an enclosed cabin with camera, load cell and control panel units. System not only classifies the apples on the basis of color, size and weight parameters but also identifies defective apples. The proposed system took only 0.52 s to capture the apple image and process. Average 15 apples per seconds were sorted by the system. Author claims average sorting accuracy between 73 and 96% and the system can be used to sort different fruits like orange, potatoes and so on. A grading and sorting system based on machine vision for date fruit was developed by Ohali [59]. The system was able to categorize the date fruit into three classes (grade 1, 2 or 3) from the given RGB image as an input. A back-propagation algorithm was tested in the study which showed 80% accuracy. Fruits and vegetables quality depends on their parameters like shape, size, texture, color and defects. Different methods needs to apply in order to classify the fruits and vegetables according to their quality parameters like data collection, pre-processing of data, image segmentation, feature extraction, and finally classification. Bhargava, et al. [60] presented a detail survey to compare the various algorithms used in every stage of the fruits and vegetables quality inspection. Meshram, et al. [61] proposed a new framework called “MNet: Merged Net” to reduce the fruits misclassification problem. Author created his own dataset of top Indian fruits consists of 12,000 images with

**Table 6**
Analysis of post-harvesting works.

| Sr. No. | Property | Classes defined in the work | Dataset used (Public / Own) | Total no of images used for training | Models / Method / Algorithms compared | Results | Model evaluation technique | Reference |
|---|---|---|---|---|---|---|---|---|
| 1 | POSTHARVEST GRADING CLASSIFI-CATION OF CAVENDISH BANANA | 4 classes | own | 1116 | Python OpenCV and Tensorflow | 0.9 | accuracy | [54] |
| 2 | Defect discrimination and grading in tomatoes | 4 classes: category 1, 2, 3, and 4. depends upon defect, healthy, and ripeness (red color intensity) | own | 8000 | linear-SVM, quadratic-SVM, cubic-SVM, and radial basis function (RBF-SVM), ANN, decision tree, and random forest | 0.9709 | Confusion matrix | [55] |
| 3 | Postharvest classification of banana (Musa acuminata) | extra class, class I, class II and reject class | own | 1164 | artificial neural network, support vector machines and random forest | 0.942 | Classification Accuracy, F-Score, Confusion matrix | [56] |
| 4 | Automatic apple sorting system | small, normal, large, light and dark, defective and non-defective | own | 183 | K-means, C4.5 decision tree | 0.79 | statistical test | [58] |
| 5 | Date fruit grading | 3 classes: grades 1, 2 and 3 | own | 1860 | back propagation neural network (BPNN) | 0.8 | Confusion matrix | [59] |

six classes. Table 6, presented the detail summary of post-harvesting works.

## 5. Discussion

This paper has extensively reviewed the available literature on application of machine learning and deep learning in agriculture. Different state-of-the-art machine learning and deep learning models in different stages of agriculture, including pre-harvesting, harvesting and post-harvesting in different domains were reviewed. Deep learning technology is becoming mature day-by-day. This survey shows that use of CNN in agriculture is huge and it is also getting remarkable results. By exploiting depth, other structure and hardware support, the learning capacity and accuracy of the CNN is significantly improved. Still there are challenges like dataset creation, time required for training and testing, hardware support, deployment of big models on small devices like boards or android phones, user awareness etc.

A popular technique called "Transfer Learning" is often used to mitigate the problems of small dataset, time required for training and to improve the accuracy of the model. Internet of Things (IoT) systems combined with machine learning provides a beneficial solution to improve farming gains. Real time parameters of the farms are gathered using IoT, and the collected data is used by machine learning algorithms either to predict or for recommendations to farmers for improvements in farming. From the survey it is also observed that Single-Shot Convolution Neural YOLO (You only look once) is a state-of-the-art, real-time object detection system which must be used for detection and localization to increase the classification accuracy.

Automated machine learning (AutoML) is the latest approach which can be used to build highly efficient, more accurate, high quality ML models in a less time [62,63]. AutoML is used to automate the entire ML pipeline shown in the Fig. 3, starting from data cleaning to model selection and hyperparameters tuning. These are time-consuming and iterative tasks of machine learning model development. As compared to traditional ML model development which is time-consuming, resource-intensive, need domain expertize, AutoML can accelerate the complete process to get production-ready model in less time without requiring domain expertise. In depth surveys on automated machine learning (AutoML) is presented in [64–69].

## 6. Challenges and recommendations

From this survey one can comprehend the importance of machine learning in the agriculture domain. In each phase of agriculture starting from pre-harvesting to post-harvesting, researchers have applied machine learning algorithms to solve the complex problems. Today's need is to develop precise and customized machine learning models which can perform fast, automatically analyze bigger, more complex data and help to optimize the agriculture processes like classification, recommendations or predications.

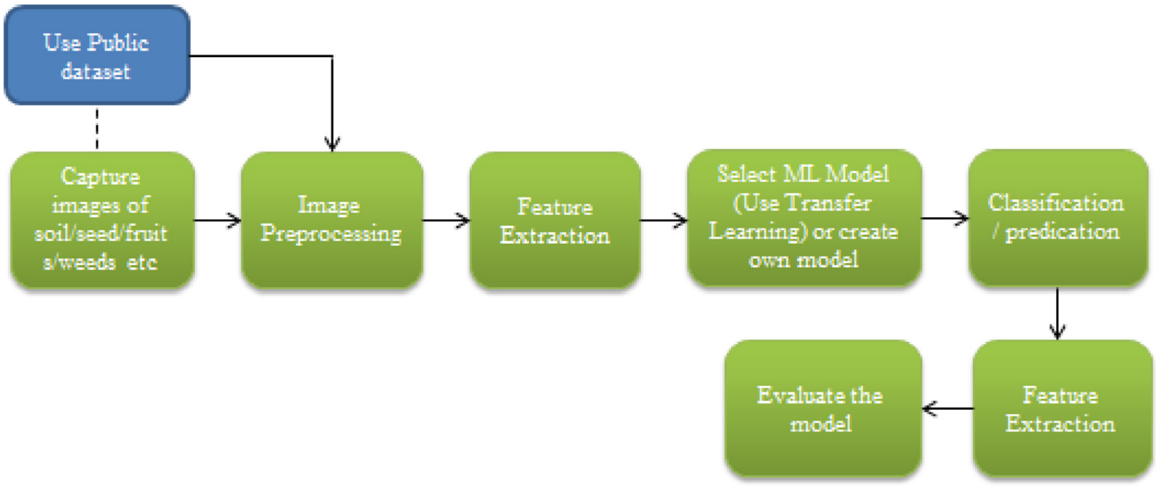



**Fig. 3.** Steps of Machine Learning used in literature.

The benefits of machine learning in agriculture domain are enormous. However, the benefits come with its challenges. Few such challenges while implementing machine learning algorithms in agriculture domain are listed as follows:

1) *Data:* Data is the most fundamental requirement to build the machine learning models. Many researchers faced the challenges regarding data like lack of data, unavailability of data in required format, poor quality of data, data may contain extraneous features etc. From this survey it is observed that, many researchers use data source sites like Kaggel, Meandly, IEEE Dataport etc. to get the data to build models. If the required data is not available then researchers need to build their own dataset [70–75].
2) *Pre-processing of the data:* As there are lot of problems associated with data, one has to apply the different pre-processing techniques to make the data suitable for training, testing, and validation testing the model. This might be time consuming process.
3) *Selection of machine learning algorithms:* Wide list of machine learning algorithm is available which make it difficult to find out more suitable algorithm to build the customize machine learning model. Many times, it is required to do random selection or after comparing results of multiple algorithms one can come to conclusion for best suitable algorithm. This trial-and error technique may delay the model deployment process.
4) *Training and testing of the machine learning model:* Building the accurate model needs huge data for training. Testing and validation are also important to check the accuracy of the model before its deployment. Building a model from scratch for best desired and possible outcomes needs long training and multiple time testing which are very time-consuming tasks. It needs high configuration hardware resources; domain knowledge programmers, testing tools etc. Overfitting and underfitting are the common challenges faced while building the models.
5) *Deployment of models:* This is the most challenging phase to bring the models in the production as there is absence of deployment skills, third party library dependencies, size of models, complex real-world scenarios, deployment platform hardware limitations, (like android phones, embedded boards) etc.

Some more challenges are important to make a note of:

1) Understanding the business need and identification of problem.
2) Understanding user and their interaction with technology
3) User friendly application design.
4) Performance of models in the real-word scenarios.
5) Power consumption by model and battery limitations to run the model on the devices.
6) For computer vision models camera configurations at user end.

The applications of machine learning and deep learning in the field of agriculture are huge with many challenges. After this in-depth survey following are a few recommendations to make the implementation process more fast, accurate, smooth and deployable.

1) Focus to build a machine learning model to solve specific problem like classification or recommendation.
2) For training the model try to create own dataset and make this available to other researchers through open platform like Kaggel, Meandly, IEEE Dataport etc.
3) For testing and validation of the models use publically available dataset.
4) To reduce the time required for training a model use the “Transfer Learning” techniques.
5) AutoML is the state-of-the-art approach which can be used to build more accurate, high quality ML models in a less time.
6) Deployment of the model in real-time application is recommended to help the intended users in their mundane work.

## 7. Conclusion

In this paper an in-depth survey of applications of machine learning algorithms in agriculture domain is presented. According to this review, agriculture activities are broadly categorized into three major areas as pre-harvesting, harvesting and post harvesting. Important parameters to be considered in each stage are shown in Fig. 2 and Table 1. Machine learning algorithms/techniques used in each stage are reviewed and presented in Tables 2,3,4,5 and 6 respectively. Machine learning is the state-of-art technology which is used to solve complex problem in the agriculture and helping farmers to reduce their losses. In this survey it is seen that machine learning algorithms have obtained remarkable outcomes to solve agriculture related problems.

Our study indicated that there is need to follow the machine learning pipeline with standard experimental methods. Researches should create their own dataset and make this available to others through different platforms, so that others can use it for testing and validation of their own models. This comprehensive survey of various machine learning algorithms used in different stages of agriculture will be more helpful to other researches who are working in this field.

## Financial and ethical disclosures

This work is not supported fully or partially by any funding organization or agency.

## Declaration of Competing Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References


[1] Siqueira VS, Borges MM, Furtado RG, Dourado CN, MCosta R. Artificial Intelligence applied to support medical decisions for the automatic analysis of echocardiogram images: a Systematic Review. Artif Intell Med 2021;120(5):102165 doi.org/10.1016/j.artmed.2021.102165.

[2] Karie NM, Kebande VR, Venter HS. Diverging deep learning cognitive computing techniques into cyber forensics. Forensic Sci Int 2019;1:61–7 vol. 1, pp. 61-67doi.org/10.1016/j.fsisyn.2019.03.006.

[3] Doowon J. Artificial intelligence security threat, crime, and forensics: taxonomy and open issues. IEEE Access 2020;8 184560-184574. 10.1109/ACCESS.2020.3029280.

[4] Meshram VV, Patil K, Meshram VA, Shu FC. An Astute Assistive Device for Mobility and Object Recognition for Visually Impaired People. IEEE Trans Hum Mach Syst 2019;49(5):449–60 10.1109/THMS.2019.2931745.

[5] Patil K, Jawadwala Q, Shu FC. Design and construction of electronic aid for visually impaired people. IEEE Trans Hum Mach Syst 2018;48(2):172–82 10.1109/THMS.2018.2799588.

[6] Arah IK, Amaglo H, Kumah EK, Ofori H. Preharvest and postharvest factors affecting the quality and shelf life of harvested tomatoes: a mini review. Int J Agron 2015;2015 6,Article ID 478041http://dx.doi.org/10.1155/2015/478041.

[7] Prange RK. Pre-harvest, harvest and post-harvest strategies for organic production of fruits and vegetables. Acta Hortic 2012;933:43–50 DOI10.17660/ActaHortic.2012.933.3.

[8] Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: a review. *Sensors* (Switzerland) 2018;18(8):1–29 10.3390/s18082674.

[9] Sharma A, Jain A, Gupta P, Chowdary V. Machine learning applications for precision agriculture: a comprehensive review. IEEE Access 2021;9:4843–73 10.1109/ACCESS.2020.3048415.

[10] Suchithra MS, Pai ML. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. Inf Process Agricul 2019;7(1):72–82.

[11] Yang M, Xu D, Chen S, Li H, Shi Z. Evaluation of machine learning approaches to predict soil organic matter and pH using vis-NIR spectra. *Sensors* (Switzerland) 2019;19(2):263–77 10.3390/s19020263.

[12] Morellos A, Pantazi X, Moshou D, Alexandridis T, Whetton R, Tziotzios G, Wiebensohn J, Bill R, Mouazen AM. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosyst Eng 2016;152:104–16 http://dx.doi.org/10.1016/j.biosystemseng. 2016.04.018.

[13] Johann AL, de Araújo AG, Delalibera HC, Hirakawa AR. Soil moisture modeling based on stochastic behavior of forceson a no-till chisel opener. Comput Electron Agricul 2016;121:420–8 http://dx.doi.org/10.1016/j.compag.2015 .12.020, 2016.

[14] Nahvi B, Habibi J, Mohammadi K, Shamshirband S, Saleh Al Razgan O. Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature. Comput Electron Agricul 2016;124:0168–1699 http://dx.doi.org/10.1016/j.compag.2016.03.025.

[15] Sivakumar D, SuriyaKrishnaan K, Akshaya P, Anuja GV, Devadharshini GT. Computerized growth analysis of seeds using deep learning method. Int J Recent Technol Eng 2019 Volume-7Issue-6S5.

[16] Huang S, Fan X, Sun L, Shen Y, Suo X. Research on classification method of maize seed defect based on machine vision. J Sens 2019;2019:9 Article ID 2716975https://doi.org/10.1155/2019/2716975.

[17] Zhu S, Zhou L, Gao P, Bao Y, He Y, Feng L. Near-infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties. Molecules 2019;24:3268 10.3390/molecules24183268.

[18] Young J, Se JK, Dayeon K, Keondo L, Wan CK. Super-high-purity seed sorter using low-latency image-recognition based on deep learning. IEEE Robot Autom Lett 2018:2377–3766 https://doi.org/10.1186/s12859-018-2267-2.

[19] Ke-ling TU, Lin-juan LI, Li-ming YANG, Jian-hua WANG, Qun SUN. Selection for high quality pepper seeds by machine vision and classifiers. J Integr Agric 2018;17(9):1999–2006 10.1016/S2095-3119(18)62031-3.

[20] Uzal LC, et al. Seed-per-pod estimation for plant breeding using deep learning. Comput Electron Agricul 2018;150:196–204 https://doi.org/10.1016/j.compag.2018.04.024.

[21] Veeramani B, Raymond JW, Chanda P. DeepSort: deep convolutional networks for sorting haploid maize seeds. BMC Bioinformatics 2018;19(9):289 Suppl-https://doi.org/10.1186/s12859-018-2267-2.

[22] D. Nkemelu, D. Omeiza, and N. Lubalo,"Deep convolutional neural network for plant seedlings classification", 2018, arXiv:1811.08404v1 [cs.CV].

[23] Medeiros Dantas de, Pereira MDias, Fernanda Santos Neri Soares T, Gomes Noronha B, Teixeira Pinheiro D. Computer vision as a complementary method to vigour analysis in maize seeds. J Exp Agricul Int 2018;25(5):1–8 Article no.JEAI.43464.

[24] Amiryousefi MR, Mohebbi M, Tehranifar A. Pomegranate seed clustering by machine vision. Food Sci Nutr 2017;6(1):18–26 10.1002/fsn3.475.

[25] Vlasov AV, Fadeev AS. A machine learning approach for grain crop's seed classification in purifying separation. IOP Conf Ser 2017;803(2017):012177 10.1088/1742-6596/803/1/012177.

[26] Kurtulmuş F, Alibaş İ, Kavdir I. Classification of pepper seeds using machine vision based on neural network. Int J Agricul Biol Eng 2016;9(1):51–62 10.3965/j.ijabe.20160901.1790.

[27] Alagumariappan P, Dewan NJ, Muthukrishnan GN, Bojji Raju BK, Bilal RAA, Sankaran V. Intelligent plant disease identification system usingmachine learning. Eng Proc 2020;2:49 10.3390/ecsa-7-08160.

[28] Savary S, Ficke A, Aubertot J-N, Hollier C. Crop losses due to diseases and their implications for global food production losses and food security. Food Secur 2012 DOI 10.1007/s12571-012-0200-5.

[29] Sujatha R, Chatterjee JM, Jhanjhi NZ, Brohi SN. Performance of deep learning vs machine learning in plant leaf disease detection. Microprocess Microsyst 2021;80 https://doi.org/10.1016/j.micpro.2020.103615.

[30] Karada˘g K, Tenekeci ME, Tasaltın R, Bilgilic A. Detection of pepper fusarium disease using machine learningalgorithms based on spectral reflectance. Sustain Comput 2018:8 https://doi.org/10.1016/j.suscom.2019.01.001.

[31] Pandya IY. Pesticides and their applications in agriculture. Asian J Appl Sci Technol (AJAST) 2018;2(2):894–900 ISSN: 2456-883X.

[32] Arsenovic M, Karanovic M, Sladojevic S, Anderla A, Stefanovic D. Solving current limitations of deep learning based approaches for plant disease detection. Symmetry (Basel) 2019;11:939 10.3390/sym11070939.

[33] Barbedo JGA. Plant disease identification from individual lesions and spots using deep learning. Biosyst Eng 2019;180:96–107 https://doi.org/10.1016/j.biosystemseng.2019.02.002.

[34] Saleem MH, Potgieter J, Arif KM. Plant disease detection and classification by deep learning. Plants 2019;8:468 10.3390/plants8110468.

[35] Türkoğlu M, Hanbay D. Plant disease and pest detection using deep learning-based features. Turk J Electr Eng Comput Sci 2019;27:1636–51 10.3906/elk-1809-181.

[36] Liu B, Zhang Y, He D, Li Y. Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry (Basel) 2018;10:11 10.3390/sym10010011.

[37] Kour V, Arora S. Fruit Disease Detection Using Rule-Based Classification. In: Proceedings of Smart Innovations in Communication and Computational Sciences, Advances in Intelligent Systems and Computing (ICSICCS-2018); 2019. p. 295–312. 10.1007/978-981-13-2414-7_28.

[38] Xing S, Lee M, Lee K. Citrus pests and diseases recognition model using weakly dense connected convolution network. Sensors 2019;19:3195 10.3390/s19143195.

[39] Doh B, Zhang D, Shen Y, Hussain F, Doh RF, Ayepah K. Automatic Citrus Fruit Disease Detection By Phenotyping Using Machine Learning. In: Proceedings of the 25th international conference on automation & computing. Lancaster UK: Lancaster University; 2019. 5-7 September.

[40] Hua Y, Zhang N, Yuan X, Quan L, Yang J, Nagasaka K, Zhou X. Recent advances in intelligent automated fruit harvesting robots. Open Agricul J 2019;13:101–6 10.2174/1874331501913010101.

[41] Kushtrim B, Demetrio P, Alexandra B, Brunella M, Grappa C. Single-shot convolution neural networks for real-time fruit detection within the tree. Front Plant Sci 2019;10:611 10.3389/fpls.2019.00611.

[42] Hossain MS, Al-Hammadi M, Muhammad G. Automatic fruit classification using deep learning for industrial applications. IEEE Trans Ind Inf 2019;15(2) 10.1109/TII.2018.2875149.

[43] Kirk R, Cielniak G, Mangan M. L*a*b*Fruits: a Rapid and Robust Outdoor Fruit Detection System Combining Bio-Inspired Features with One-Stage Deep Learning Networks. Sensors 2020;20:275 10.3390/s20010275.

[44] Altaheri H, Alsulaiman M, Muhammad G. Date fruit classification for robotic harvesting in a natural environment using deep learning. IEEE Access 2019;7:117115–33 10.1109/ACCESS.2019.2936536.

[45] Bauer A, Bostrom AG, Ball J, Applegate C, Cheng T, Laycock S, Rojas SM, Kirwan J, Zhou Ji. Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: a case study of lettuce production. Hortic Res 2019;6:70 https://doi.org/10.1038/s41438-019-0151-5.

[46] Zhang T, Huang Z, You W, Lin J, Tang X, Huang H. An autonomous fruit and vegetable harvester with a low-cost gripper using a 3D sensor. Sensors 2020;20:93 10.3390/s20010093.

[47] Onishi Y, Yoshida T, Kurita H, Fukao T, Arihara H, Iwai A. An automated fruit harvesting robot by using deep learning. Robomech J 2019;6:13 https://doi.org/10.1186/s40648-019-0141-2.

[48] X. Liu, S.W. Chen, S. Aditya, N. Sivakumar, S. Dcunha, C. Qu, C.J. Taylor, J. Das, and V. Kumar,"Robust Fruit Counting: Combining Deep Learning, Tracking, and Structure from Motion", arXiv:1804.00307v2 [cs.CV] 2 Aug 2018.

[49] United states department of Agriculture (USDA) Grade standards for Fruits, https://www.ams.usda.gov/grades-standards/fruits (Accessed: July 2021).

[50] EU, Fruit and Vegetables: Marketing Standards, 2011,https://ec.europa.eu/agriculture/fruit-and-vegetables/marketing-standards_en (Accessed: July 2021).

[51] Government Of India, AGMRK ("Agricultural Marketing Adviser"), 2004, https://upload.indiacode.nic.in/showfile?actid=AC_CEN_23_31_00011_193701_1535099362507&type=rule&filename=fruits_and_vegetables_grading_and_marking_rules,_2004.pdf (accessed:July 2021).

[52] Esguerra Elda B. Post-harvest management of mango for quality and safety assurance guidance for horticultural supply chain stakeholders. Rome: Food and Agriculture Organization of The United Nations; 2018.

[53] Gaetano Paltrinieri. Handling of fresh fruits, vegetables and root crops -a training manual- for grenadaAgricultural marketing improvement TCP/GRN/2901. Food and Agriculture Organization of the United Nations; 2021. http://www.fao.org/3/au186e/au186e.pdf. accessed:July.

[54] Ucat RC, Dela Cruz JC. Postharvest grading classification of cavendish banana using deep learning and tensorflow. In: 2019 International Symposium on Multimedia and Communication Technology (ISMAC); 2019. p. 1–6. 10.1109/ISMAC.2019.8836129.

[55] Ireri D, Belal E, Okinda C, Makange N, Ji C. A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. Artif Intell Agricul 2019;2:28–37 https://doi.org/10.1016/j.aiia.2019.06.001.
[56] Jr Piedad E, Larada JI, Pojas GJ, Vithalie L, Ferrer V. Postharvest classification of banana (Musa acuminata) using tier-based machine learning. Postharvest Biol Technol 2018;145:93–100 https://doi.org/10.1016/j.postharvbio.2018.06.004.
[57] Lia J, Chena L, Huanga W. Detection of early bruises on peaches (Amygdalus persica L.) using hyperspectral imaging coupled with improved watershed segmentation Algorithm. Postharvest Biol Technol 2018;135:104–13 http://dx.doi.org/10.1016/j.postharvbio.2017.09.007.
[58] Sofu MM, Er O, Kayacan MC, Cetisli B. Design of an automatic apple sorting system using machine vision. Comput Electron Agric 2016;127:395–405 http://dx.doi.org/10.1016/j.compag.2016.06.030.
[59] Ohali YA. Computer vision based date fruit grading system: design and implementation. J King Saud Univ – Comput Inf Sci 2011;23:29–36 10.1016/j.jksuci.2010.03.003.
[60] Bhargava A, Bansal A. Fruits and vegetables quality evaluation using computer vision: a review. J King Saud Univ – Comput Inf Sci 2018;33:243–57 https://doi.org/10.1016/j.jksuci.2018.06.002.
[61] Meshram VA, Patil K, Ramteke SD. MNet: a framework to reduce fruit image misclassification. Ing Syst Inf 2021;26(2):159–70 https://doi.org/10.18280/isi.260203.
[62] Microsoft, "What is automated machine learning (AutoML)?" https://docs.microsoft.com/en-US/azure/machine-learning/concept-automated-ml, (Accessed: July 2021).
[63] Hibayesian, "awesome-automl-papers", 2017, https://github.com/hibayesian/awesome-automl-papers (Accessed: July 2021).
[64] X. He, K. Zhao, X. Chu, "AutoML: A Survey of the State-of-the-Art", 2021, arXiv:1908.00709v6 [cs.LG] 16 Apr 2021.
[65] Zoller M, Huber MF. Benchmark and survey of automated machine learning frameworks. J Artif Intell Res 2021;70:409–74. arXiv:1904.12054v5 [cs.LG] 26 Jan 2021.
[66] R. Elshawi, M. Maher, S. Sakr, "Automated machine learning: state-of-the-art and open challenges", 2019, arXiv:1906.02287v2 [cs.LG] 11 Jun 2019.
[67] Q. Yao, M. Wang, Y. Chen, W. Dai, Y. Li, W. Tu, Q. Yang, Y. Yu, "Taking the human out of learning applications: a survey on automated machine learning", 2019, arXiv:1810.13306v4 [cs.AI] 16 Dec 2019.
[68] L. Yang, and A. Shami, "On hyperparameter optimization of machine learning algorithms: theory and practice", 2020, arXiv:2007.15745v2 [cs.LG] 7 Aug 2020.
[69] H.J. Escalante, "Automated Machine Learning - a brief review at the end of the early years", 2020, arXiv:2008.08516v3 [cs.LG] 24 Aug 2020.
[70] Mihai Oltean. Fruits 360 dataset. Mendeley Data 2018;V1 10.17632/rp73yg93n8.1.
[71] prabira Kumar sethy. Indian Fruits-40. Mendeley Data 2020;V1 10.17632/bg3js4z2xt.1.
[72] Meshram VA, Thanomliang K, Ruangkan S, Chumchu P, Patil K. FruitsGB: top indian fruits with quality. IEEE Dataport 2020 https://dx.doi.org/10.21227/gzkn-f379.
[73] Meshram VA, Patil K. FruitNet: Indian Fruits Dataset with quality (Good, Bad & Mixed quality). Mendeley Data 2021;V1. doi:10.17632/b6fftwbr2v.1.
[74] Tripathi M, Maktedar D, Dhanajay D. Fruits and Vegetables. Mendeley Data 2020;V2 10.17632/73kpfrbcck.2.
[75] Math M, Kumar R, Dharwadkar V, Nagaraj Dr. Real-world tomato image dataset for deep learning and computer vision applications involving precision agriculture. Mendeley Data 2020;V1 10.17632/9zyvdgp83m.1.