Ministry of Higher Education and
Scientific Research
University of Diyala
College of Science
Department of Computer Science

Diseases Prediction and Detection in Iraqi Agriculture Using Internet of Things Based on Deep Learning Techniques

A Dissertation

Submitted to the Department of Computer Science\ College of Sciences\ University of Diyala in Partial Fulfillment of the Requirements for the Degree of PhD in Computer Science

By

Marwan Adnan Jasim

Supervised By

Prof. Dr. Jamal Mustafa AL-Tuwaijari

2025 A.D. **1447 A.H.**

Abstract

Plant health is essential for maintaining food availability and agricultural productivity. In Iraq, sustaining plant productivity remains a critical factor in ensuring national food security. The main causes of plant damage and reduced production are diseases. Therefore, this dissertation utilizes modern technologies, namely AI integrated with the IoT, to build an accurate architecture for managing plant illnesses with high precision and achieve precision agriculture in Iraq.

The comprehensive practical structure of this dissertation comprises two main subsystems.: The Early Disease Prediction System (EDPS), which predicts diseases in advance using key environmental factors collected. The relationship between these factors and disease occurrence is modeled using linear regression and random forests to estimate the probability of disease the next day. The second system, the Visual Disease Detection System (VDDS), processes real-world, field-acquired images captured using a Raspberry Pi camera. The images undergo a series of high-level preprocessing steps, including smart cropping, resizing, normalization, class-unbalanced handling, and data leakage prevention. The dataset is partitioned into "training, validation, and testing" subsets following a 70–10–20 split. Six models were trained, including two lightweight architectures developed in this dissertation (LightRes-SE and AttenDW-CBAM), the modified ResNet-18, a Baseline CNN, and the two pre-trained MobileNetV2 and EfficientNetB0. This diverse set provides a robust, comprehensive evaluation, demonstrating the overall strength and effectiveness of the proposed system across a broad comparative context, and the best-performing one was chosen. Random Forest and XGBoost algorithms were used for classification. Model optimization was performed using a teacher–student knowledge distillation framework, with one model serving as the teacher and another as the student, combined with PCA and FP16 quantization to prepare the models for deployment.

The results demonstrate that the EDPS system can effectively estimate disease probability using temperature and humidity data, as evaluated by RMSE and MAE.

Linear regression performed particularly well, giving a low error of **0.0073** for next-day prediction. Within the proposed VDDS, knowledge distillation demonstrated that transferring knowledge from the best-performing LightRes-SE model (teacher) to the AttenDW-CBAM model (student) not only preserved but slightly improved the classification accuracy to **0.9815**, while reducing both model size and inference time, the model size decreased from 10.399 MB to **3.927 MB**, and the inference time decreased from 216.11ms to **113.48ms**. Subsequent FP16 quantization further reduced size and computation time, with minimal accuracy loss of less than 3%. Additional metrics, including MCC, Kappa, and ECE, confirmed the consistency and reliability of the predictions. Grad-CAM was employed to interpret the model, while the combined system effectively reduced disease risk from prediction to visual detection.

Chapter One

General Introduction

Chapter One
General Introduction

1.1 Introduction

Plants are critical to all of humanity because they are the primary source of food and energy, and they form the basis for medical and industrial raw materials and other resources [1]. Agriculture has always been, and continues to be, the foundation for the continuity and sustainability of major food crops and other living organisms today [2]. Additionally, it plays a key role in providing oxygen and absorbing carbon dioxide, and is one of the most prominent elements in combating pollution on Earth. Plants are the primary source of food security in most countries worldwide, and they face challenges related to food security and global health [3].

Diseases are among the plants' most considerable issues threatening their production and growth. The plant's health and freedom from infections make its production more critical, reducing the consumption of various resources. Many illnesses affect plants, including fungi, bacteria, and other organisms that directly or indirectly affect crop health [4]. Worldwide, the greatest threat to plants is disease. They share a significant loss in agricultural yields, negatively affecting countries' economies. Recent research indicates that diseases cause many agricultural crop losses .Diseases that affect plants affect them and their production, and may lead to deterioration and death. They have an economic impact in decreased production, the high costs of combating and treating diseases, and the financial burdens on farmers [5].

Given the importance of the plants mentioned and the impact of diseases on them, attention must be paid to detecting and treating diseases early. The methods used in many agricultural countries are mostly outdated, relying on visual observation and the calling of farm experts. This requires time, effort, extensive experience, and knowledge of disease patterns. It often takes a long time to complete the detection and treatment process, indicating the disease has spread or the plant's condition has deteriorated. Therefore, it is indispensable to use modern technologies, such as artificial intelligence, to detect diseases early and treat them quickly before they spread [6].

Agriculture has been the major and central occupation of the people of Iraq and the old Mesopotamian cities from their very inception. Iraq's plentiful fresh water, favorable climate, and land suitable for cultivation made it significant in agriculture.

The interest of Iraqis in agriculture is based on several reasons, such as providing the market with plant products, opportunities for farmers, and strengthening the Iraqi economy as a whole [7, 8].

The import of plant products, particularly horticultural fruit and vegetables, from foreign countries and the inability to reach the goal of self-sufficiency is among the most significant influences on the Iraqi economy today. Low agricultural production resulting from symptoms plant diseases, water scarcity, and so on contributes to this dependency [9].

Plant diseases are one of the biggest challenges facing agriculture in Iraq. Diseases fundamentally affect plants, impairing their growth and production and leading to their death. The emergence and spread of viral, bacterial, and fungal diseases in Iraqi farms are due to several reasons, the most important of which is the use of traditional methods of detecting and treating diseases. Using outdated methods in Iraqi agriculture to detect and treat diseases is one of the most critical obstacles to the deterioration of the sector and the spread of disease. Instead, modern technologies such as AI and the IoT should be used to discover and treat diseases quickly [10].

1.2 Artificial Intelligence

"Artificial intelligence" (AI) seeks to imitate human perception in machines designed to behave and learn, such as humans. Its discovery dates back to the mid-twentieth century when John McCarthy first called it artificial intelligence at the Dartmouth Conference, which is considered the birth of AI. Since its inception, all research has focused on solving problems using AI . It began to develop over the decades. Expert systems started in the eighties, and machine learning and neural networks began to appear and grow in the twenty-first century to the point of development we are witnessing today [11].

Machine Learning (ML), Deep Learning (DL), computer vision, robotics, and pattern recognition are just a few of the subfields that make up artificial intelligence.

It creates diverse systems based on algorithms that adapt, analyze, learn, and make decisions independently. AI enters almost all areas of life, such as medicine, engineering, agriculture, industry, education, economy, and security [12].

AI in agriculture has revolutionized efficiency and productivity. As the world population grows, the need for food also increases. This requires adopting modern practices such as AI in agriculture, primarily by enhancing crop management and decision-making. Precision agriculture involves collecting diverse data on agricultural fields from various sources, including drones, field-mounted cameras, satellites, and sensors, to analyze and make correct decisions regarding diseases, cultivation, irrigation, fertilization, and harvesting. [13]. In developed countries, artificial intelligence is used in precision agriculture and integration with modern devices such as sensors and cameras in global farms and vegetable fields .It helps in the early and accurate detection of various plant illnesses, reducing crop loss [14].

1.3 Internet of Things (IoT) in Agriculture

The IoT is a network of interconnected devices that interact and share information via the Internet. This term appeared in the early 2000s and is continuously being developed to the present day. It is now causing a major revolution and trend in all fields [15]. The IoT is used in many fields, such as healthcare, hospitals, industry, smart cities, smart homes, agriculture, intelligent farms, traffic control, waste management, and many other fields [16].

IoT has entered the agricultural field firmly and has become a trend in developed countries. It is used in precision and modern agriculture by planting multiple sensors in the soil and field, cameras, and various sensors to monitor agricultural fields, collect data, and send it via the Internet to control centres. In addition to using irrigation systems, water levels are controlled, fertilization is applied, soil nutrients and moisture are tested, and diseases are monitored using cameras and multiple sensors [17]. The IoT is used to monitor water, soil, and plant conditions. Monitoring plants, soil, and their elements also reduces the use of fertilizers and pesticides, which supports the country's economy [18].

1.4 Related Works

This section explains and clarifies recent work related to disease prediction and diagnosis using modern technologies such as AI and the IoT:

In 2019, Hu et al. [19] proposed a new method for detecting tea leaf diseases. They used the support vector machine (SVM) algorithm to extract image features. They employed different optimization methods, such as deep conditional generative adversarial networks (C-DCGAN), to generate new data samples. A pre-trained model (VGG16) was used on these optimized samples to detect diseases. Using a traditional framework and modern pre-trained techniques such as VGG16, and generating new samples from the data, an accuracy of 90% was achieved in diagnosing tea diseases.

In 2019, Marceline Francis and C. Deasy [20] they offered a survey on the identification of infections in two plants apples and tomatoes using AI methods. This is obtained using a convolutional neural network trained in multiple stages, with a fully connected layer and a combination layer. It had 3663 images, labelled as healthy or unhealthy. Overfitting was minimized using various techniques (dropout, regularization), with the model achieving 87% diagnostic accuracy.

In 2020, Agarwal et al. [21] proposed a new strategy called ToLeD based on the CNN algorithm. They used a reliable global dataset called "Plant Village," which contains many images of healthy and unhealthy plant leaves. The number of classified categories was eight different categories of tomato diseases and one category for healthy ones. Several convolutional layers were used, followed by pooling and fully connected layers. They achieved high accuracy in diagnosing plant diseases (91%) and created a model that outperforms two pre-trained models, Inception V3 and VGG-16, which are highly efficient.

In 2020, Yan Guo et al. [22], they proposed a deep-learning model for plant disease diagnosis in smart agriculture. They combined several contemporary techniques, such as transfer learning, region proposal network (RPN), and another algorithm called the image segmentation algorithm Chan Vese (CV). Plant diseases in complex environments are detected from images using this combination. Features are extracted using the image segmentation algorithm, and transfer learning is used to detect multiple diseases, such as bacterial plaque and mildew. The model effectively handled complexity, enhancing mode1 generalization, and training efficiency. It achieved high accuracy in plant disease detection, reaching 83.57%, outperforming traditional methods, and improving smart agriculture. This study also pointed out some challenges, such as the regular iterative nature of the Chan-Vese algorithm, which affects the speed.

In 2020, Junde Chen et al. [23], they presented a model based on transfer learning and CNNs for disease recognition. This model was used and tested on big data to extract features and accurately classify diseases to improve agricultural environments. Pre-trained models such as VGG16 were used on the popular database ImageNet and modified to suit diverse agricultural data environments. They added “batch normalization,” swapped out the last layers of the VGG Net model for a second convolutional layer, and used the Swish activation function rather than the conventional ReLU function to recognize subtle plant disease symptoms and improve performance. The model achieved good results, represented by the high accuracy of 92 %, representing a clear improvement over other detection models on a general dataset that was used, which contained diverse images with different backgrounds. These results work to develop the agricultural sector and manage plant diseases in an advanced manner.

In 2021, Mohamed Samer I [24] presented an intelligent prediction system to predict late blight disease in potato plants before it occurs. Several techniques were used, such as machine learning and its integration with the IoT. The IoT was used to collect data from various sensors to measure environmental factors such as temperature, relative humidity, and relative pressure, which affect the appearance of the disease. Late blight is a globally widespread disease that causes significant losses in several crops, including potatoes and tomatoes. The authors integrated logistic regression techniques with neural networks to accurately predict diseases based on collected environmental data and to improve and develop agriculture through early disease detection, thereby reducing their spread later. The study results show that the suggested model's accuracy was significantly enhanced compared to Microsoft's cloud-based forecasting system, "Azure." The study attained up to 99.4% accuracy. The outcomes show how well the model manages diseases and how it might be used in more general precision agriculture contexts. According to the researcher, future research will concentrate on developing this proactive system to forecast various plant diseases and pests, increasing agricultural productivity.

In 2022, Ramana et al. [25], they proposed a study on disease detection in plant leaves using deep neural network architecture and IoT, as IoT opens a new and helpful field in smart agriculture by making users collect data from agricultural fields in real-time and transmit it to remote locations for processing and utilizing data derived from sensors and images captured from fields. A mixture of sensors was used, such as a moisture sensor, another sensor to monitor acidity (pH) level, a sensor to measure temperature, and a sensor to measure soil moisture. Several of these devices were installed in different locations of the farms, and RasPI 3 (RPI3) was used as a central controller to control the sensors. A camera was also connected to RasPi (RasPi) to monitor the plant leaves. A CNN architecture was utilized to detect diseases based on the captured images of the leaves. The results showed that the disease identification system using CNN achieved an accuracy of 96% after repeating the system for 50 training epochs. These results reinforce the importance of employing various AI and IoT technologies to enhance agricultural productivity by accurately predicting plant diseases.

In 2022, Hussain, Rafakat, et al. [26], they conducted research and demonstrated that rainfall, maximum and minimum relative humidity, and temperature extremes were all relevant in determining late blight disease development in potatoes. Linear regression and correlation coefficients were used to find the relationship between climate effects and disease occurrence. Correlation analyses showed a significant positive relationship between maximum temperature and late blight severity, while minimum relative humidity was significantly negatively associated with the severity of infection. Maximum temperature, maximum and lowest relative humidity, and rainfall were revealed to be responsible for 95.93% of the illness development throughout the research years using the multiple linear regression method. The results of this study indicate the possibility of using the predictive model to predict the emergence of the disease so that farmers can take appropriate early measures to eradicate diseases proactively. This reduces the spread of the disease by determining the optimal timing for applying control and treatment strategies.

In 2023, Challagundla, Yagnesh, et al. [27] they presented an advanced methodology for detecting citrus diseases using deep learning algorithms and machine learning models. Citrus is a crop of high economic importance, as it is grown in most countries. To address the problem of disease and its impact on countries' economies, researchers used three deep learning algorithms to generate plant image embedding: Inception V3, VGG16, and Squeeze Net. Also, five machine learning models were applied, including Random Forest, KNN, Gradient Boosting Neural Network, and Stochastic Gradient Descent (SGD), using a 5-fold cross-validation to accurately detect diseases. The results showed that the neural network integrated with Inception V3 as the image representation generator performed the best, achieving a classification accuracy of 96.6%, a recall of 96.6%, and an F1 score of 96.5 %."

In 2024, Shrestha, Tahmid Enam, et al. [28], they proposed DL based method to detect cucumber plant disease using a dataset hosted on Kaggle. Cucumber is an important vegetable for global production; thus, this study was conducted to increase productivity and maintain quality. MobileNetV2 is used as a pre-trained model, and classification was made on a modified image. This was achieved by applying a few layers, namely linear layers, and by using other methods and activation functions, such as the ReLU function, to extract important features and perform classification more accurately. Development of the Yolo v3 Tiny model for Cucumber Disease Detection with an Accuracy of 96.8%, helping with precision Farming and proper disease management.

In 2024, Nafil, Khalid, et al. [29], they implemented IoT and AI technologies by working on tomato disease detection technology. These included a variety of climate and soil data collection methods, as well as advanced use of the "LoRaWAN" technology. It integrates soil and climate sensors to collect data and then tracks the patterns of disease prediction accordingly. Correctly control plant diseases. Also, it finds and classifies diseases with the tomato plant photos. With that accuracy (86 %) in disease detection, a complete framework was developed to monitor agricultural diseases.

In 2024, Xu, Chang, and Lingxian Zhang [30] they develop a system to identify the diseases, downy mildew, powdery mildew, and gray mold that affect the cucumber. Fifteen diseases diagnosis models based on machine learning algorithms were constructed to accurately diagnose cucumber diseases. Of the remaining models developed, five multiclass SVM-based models yielded superior cucumber disease diagnostic performance, with disease prediction accuracies greater than 80%. Also, the predictive performance was even better after applying the method "downsampling". This method helps to handle unbalancedd data, which increases the accuracy of diagnosis and makes these algorithms more effective in analyzing agricultural data to detect diseases with high accuracy.

In 2025, Hosen, M. D., and Md Hasibul Islam [31] aim to improve tomato disease management using modern AI techniques. Tomato Village's dataset comprised 4,523 images of various tomato plants affected by diseases. The researchers used pre-processing, such as image resizing and interpolation, to train multiple models. Several models, such as VGG-16 and Inception V3, were used and modified with additional layers to improve them. Several metrics, including accuracy and F1 score, were used to evaluate the models' accuracy. Accuracy reached 93% for accurately detecting tomato diseases.

Table 1.1 shows the related works regarding the dataset, source, algorithm used, and accuracy achieved.

Table 1: Summary of related works

Authors (Year) & Reference	System Objective	Dataset (Size & Source)	Algorithm	Plant Type	Accuracy (%)	Main Contribution (in brief)
Mohamed Samer I. (2021) [24]	Disease Prediction	Environmental sensor data (IoT-based)	Logistic Regression + Neural Network	Potato	99%	Proposed an IoT-based early prediction system using environmental factors to forecast late blight before occurrence.
Hussain et al. (2022) [26]	Disease Prediction	Climatic data (rainfall, humidity, temperature)	Linear Regression, Correlation Analysis	Potato	95.93 (explained variance)	Developed a climate-based predictive model to forecast late blight occurrence and support proactive disease control.
Nafil et al. (2024) [29]	Disease Prediction	IoT soil & climate data	AI + IoT	Tomato	____	Introduced an integrated IoT–AI framework for disease prediction using climate and soil patterns.
Hu et al. (2019) [19]	Disease Detection	Augmented Data	SVM + VGG16	Tea	90.0%	Combine VGG16 features with SVM to improve accuracy and reduce training time.
Francis & Deisy (2019) [20]	Disease Detection	3,663 images	CNN	Apple, tomato	87.0%	They developed a simple CNN model to separate healthy from diseased leaves in apple and tomato, with 4 Conv +Pooling layers, as a preliminary study in the field.
Agarwal et al. (2020) [21]	Disease Detection	Plant Village	CNN (ToLeD)	Tomato	91.2%	They proposed a customized CNN architecture (3 Conv layers + 3 Max Pooling + 2 FC) on the Plant Village tomato disease dataset.

Guo et al. (2020) [22]	Disease Detection	_	CNN + RPN	Several Crops	83.57 %	They proposed a framework using a region proposal network (RPN) + CNN to identify and classify plant diseases in complex environments.
Chen et al. (2020) [23]	Disease Detection	A Public Dataset	VGG Net	Several Plants	92.0%	They used a VGG network model pre-trained on public leaf data to classify multiple diseases of diverse plants; they focused on generalizability from data.
Ramana et al. (2022) [25]	Disease Detection	_	CNN + IoT	Various crops	96.0%	They performed a preliminary integration of IoT with a CNN model to automate the detection of diseases of various crops in an innovative agricultural environment.
Challagundla et al. (2023) [27]	Disease Detection	_	Inception V3 + RF	Citrus	96.6%	They used the InceptionV3 network to detect diseases in citrus plants.
Shrestha et al. (2024) [28]	Disease Detection	Kaggle Dataset	MobileN etV2	Cucumb er	96.81 %	They adopted the lightweight MobileNetV2 on Kaggle data for optimization in a lightweight agricultural computing solution.
Xu et al. (2024) [30]	Disease Detection	_	Multi-Class SVM	Cucumb er	80.0%	They adopted a multi-class SVM classifier for cucumber diseases, indicating that traditional methods remain applicable in simpler cases.
Hosen, M. D., and Md Hasibul Islam [2025] [31]	Disease Detection	Tomato village(4525) images	VGG16	Tomato	93%	They used the VGG16 network on a dataset of 4,525 images from Tomato Village, achieving relatively good performance with local data.

1.5 Problem Definition

Managing crop diseases in the Iraqi agricultural environment is a multidimensional challenge, combining economic constraints and the methodological weaknesses of current AI-based systems. Despite advances in this field, these systems suffer from three fundamental problems and gaps:

First, reliance on traditional methods (such as manual visual inspection and laboratory testing) remains prevalent despite their high cost, slow turnaround times, and susceptibility to human error, resulting in delayed therapeutic intervention during disease outbreaks and significant economic losses.

Second, limited ability to make early and accurate diagnoses in changing field environments. Most previous studies rely on laboratory data or controlled environments, which limits their generalizability to real-world applications. Furthermore, most models focus on diagnosing disease after visual onset, without accounting for advanced prediction based on climatic factors such as temperature and humidity.

Third, pre-trained deep models are often large and slow to infer, making them unsuitable for practical applications on resource-limited devices such as the RasPi used in smart farms. This is especially true in Iraq, which lacks locally-adapted intelligent diagnostic systems. Lightweight models that can be implemented on the limited-capacity devices used in Iraqi agricultural fields are needed for intelligent disease diagnosis.

Hence, the fundamental problem of this dissertation arises:

How can a unified, intelligent, and advanced system be developed that integrates environmental risk prediction with visual symptom diagnosis, using advanced AI and IoT techniques, to address Iraq's agricultural challenges in terms of cost, efficiency, and reliability, while ensuring the highest methodological standards and enabling efficient deployment on edge devices?

1.6 Aim of the Dissertation

This dissertation aims to design and implement an advanced unified smart system based on Artificial Intelligence and the Internet of Things (AIoT) to enable early prediction and detection of plant diseases in Iraqi agriculture. This is done through proactive prediction of environmental risks, with real-time confirmatory visual diagnosis, within an advanced methodological environment. Modern methods, integrated steps, and capable of effective deployment on edge devices, to ensure high diagnostic accuracy, comprehensive assessment reliability, and operational efficiency in real fields.

1.7 Objective of the Dissertation

The main objective of the dissertation is to enhance crop productivity, reduce resource waste, and address the increasing demand for farm products in Iraq to support the Iraqi economy. This dissertation aims to achieve a set of methodological and operational objectives that form the backbone of the proposed system. These objectives are

1. Collect and document the real Iraqi dataset for diagnosing diseases of various Iraqi crops, while applying strict safeguards to prevent data leakage using pHash-based Group K-Fold and canonical stem deduplication techniques. In addition to ensuring fair class representation by employing a dedicated class-balancing mechanism based on inverse frequency weights with mathematical smoothing to enhance the accuracy of balanced classification and by using the

Matthews Correlation Coefficient (MCC) to address unbalanced data in field data.

2. Developing a proactive prediction system capable of predicting disease outbreaks (before they occur), using real climate data (temperature and humidity), by training two models and comparing and analyzing them: linear regression and random forest.

3. Developing a robust integrated visual diagnostic system based on optimized CNN architectures (LightRes-SE, AttenDW-CBAM, ResNet-18, Baseline CNN, MobileNetV2, EfficientNetB0) to extract multi-level spectral and structural features, followed by systematic comparisons and selecting the best model.

4. Optimizing the model through several techniques such as Knowledge Distillation (KD), selecting the best model as a teacher, and selecting the second best model as a student, with hybrid classification heads (Random Forest and XGBoost, to achieve a balance between deep representation and statistical accuracy. Another optimization technique is to quantize to 16 bits to reduce inference time and model size, making it suitable for use in peripheral devices.

5. Employing the Grad-CAM (Gradient-weighted Class Activation Mapping) interpreter in the final evaluation phase to ensure transparency of the decision-making process. This tool generates heat maps that accurately identify visual regions in the image critical for classification, interprets the model, clarifies its purpose, and explains what happens within it.

6. To address limited connectivity and enable IoT-based disease detection, a local in-field controller(RasPi) performs on-site analysis by integrating the

model deployed on the RasPi with a field camera and an IoT interface, and sends concise alerts directly to farmers without requiring high-speed internet.

1.8 Contribution

The fundamental contribution and methodological originality of this dissertation are crystallized in the unprecedented integration of the unique Iraqi agricultural context and the methodological rigor of the experimental design, achieving qualitative progress through the following axes:

1. Create the authentic Iraqi dataset for crop disease diagnosis, consisting of high-resolution images of eight classes, which consist of citrus, tomato, and cucumber cultivars collected from Iraqi fields under realistic field conditions. This fills a critical gap in the literature, which relies on datasets that are not representative of the local context (e.g., Plant Village or Kaggle).

2. Design a new dual system that combines an early disease prediction system (EDPS) using environmental and temporal data with an immediate visual disease detection system (VDDS). This hybrid system enables farmers to intervene proactively before disease onset while ensuring precise visual diagnosis. EDPS leverages machine learning algorithms and is evaluated using MAE and RMSE. VDDS employs a hybrid architecture, training six models—including two lightweight architectures developed in this dissertation (LightRes-SE and AttenDW-CBAM), a modified ResNet-18, a Baseline CNN, MobileNetV2, and EfficientNetB0—and applies optimization techniques to achieve high accuracy, efficiency, and smooth execution.

3. Implement rigorous methodological safeguards to prevent data leakage, including pHash-based clustering, canonical stem deduplication, and group K-fold partitioning, thereby ensuring the generalizability of the results. A comprehensive evaluation protocol combining RF/XGB classifiers with/without PCA, multiple metrics (MCC, Kappa, ECE), and Grad-CAM interpretability validated the model's performance and representations.

4. This dissertation enhanced the efficiency of deep plant disease diagnosis models by balancing performance and computational cost. Using knowledge distillation from a teacher to a lightweight student model, followed by FP16 quantization, the final model size was reduced by up to 80%, inference time and memory requirements were significantly lowered, and high diagnostic accuracy was maintained. This enabled deployment on resource-constrained edge devices such as RasPi, supporting practical AI adoption in field environments.

1.9 Outline of the Dissertation

This dissertation is divided into five chapters, the contents of which are briefly described below:

Chapter 2: This chapter provides an overview of the plants and diseases studied, the scientific backgrounds used, such as AI techniques and specialized algorithms in the proposed System, with examples and detailed equations. It also provides an overview of the IoT and its various components.

Chapter 3 presents the practical implementation of the proposed AI-based system, detailing the design, development, and deployment of the disease prediction and detection systems using IoT technologies within the context of Iraqi agriculture.

Chapter 4 shows the experimental results of implementing plant disease diagnosis system.

Chapter 5: This chapter presents the conclusions drawn from the accomplishments and results of the suggested System, along with several suggestions for future work.

الخــلاصـــة

تُعَدّ النباتات ركيزةً أساسيةً للأمن الغذائي العالمي، ولا سيما في الدول الزراعية مثل العراق، حيث تُشكّل الأمراض النباتية أحد الأسباب الرئيسة لتلف المحاصيل وانخفاض الإنتاج الزراعي. وانطلاقًا من هذه الإشكالية، توظّف هذه الأطروحة التقنيات الحديثة، وتحديدًا الذكاء الاصطناعي المتكامل مع إنترنت الأشياء، لبناء نظام عملي دقيق وموثوق لإدارة أمراض النباتات، بما يسهم في دعم مفاهيم الزراعة الدقيقة في البيئة الزراعية العراقية. يتكوّن النظام الشامل لهذه الأطروحة من نظامين رئيسيين. يتمثل النظام الأول في نظام التنبؤ المبكر بالأمراض(EDPS) ، الذي يهدف إلى التنبؤ باحتمالية الإصابة بالأمراض قبل حدوثها اعتمادًا على مجموعة من العوامل البيئية الرئيسة، شملت درجة الحرارة والرطوبة وقد نُمذجت العلاقة بين هذه العوامل وحدوث المرض باستخدام خوارزميتي الانحدار الخطي والغابات العشوائية، وذلك لتقدير احتمالية الإصابة بالمرض في اليوم التالي. وعند تجاوز هذه الاحتمالية نسبة 50%، يتم تفعيل النظام الثاني تلقائيًا. أما النظام الثاني، فهو نظام الكشف البصري عن الأمراض(VDDS) ، والذي يعالج صورًا حقيقية التُقطت ميدانيًا باستخدام كاميرا Raspberry Pi. تخضع هذه الصور لسلسلة متقدمة من خطوات المعالجة المسبقة، شملت القص الذكي، وتغيير الحجم، والتطبيع، ومعالجة عدم توازن الفئات، إضافةً إلى منع تسرب البيانات. تم تقسيم مجموعة البيانات إلى مجموعات فرعية للتدريب والتحقق والاختبار وفق نسبة 70%–10%–20% على التوالي. تم تدريب ستة نماذج مختلفة، تضمنت بنيتين خفيفتي الوزن طُوِّرتا ضمن هذه الأطروحة (LightRes-SE) و(AttenDW-CBAM)، إضافةً إلى شبكةResNet-18 ، وشبكة عصبية تلافيفيه أساسية، ونموذجين مدرّبين مسبقًا هما MobileNetV2 و .EfficientNetB0وقد وفّر هذا التنوع في النماذج تقييمًا شاملًا ومقارنًا أظهر القوة والفعالية العامة للنظام المقترح، حيث تم اختيار النموذج الأفضل أداءً. كما استُخدمت خوارزميتا Random Forest و XGBoostلأغراض التصنيف .أُجري تحسين النماذج باستخدام إطار تقطير المعرفة بنموذج المعلّم–الطالب، حيث عُيِّن أحد النماذج كمعلّم والآخر كطالب، مع دمج تحليل المكونات الرئيسية (PCA) وتقنية التكميم FP16 لتهيئة النماذج لمرحلة النشر العملي.

أظهرت النتائج أن نظام EDPS قادر على تقدير احتمالية الإصابة بالأمراض بكفاءة عالية اعتمادًا على بيانات درجة الحرارة والرطوبة، وذلك وفق مقاييس جذر متوسط مربع الخطأ (RMSE) ومتوسط الخطأ المطلق (MAE)، حيث حقق نموذج الانحدار الخطي أداءً متميزًا بخطأ منخفض بلغ 0.0073 للتنبؤ باليوم التالي. وضمن نظام VDDS ، بيّنت عملية تقطير المعرفة أن نقل المعرفة من نموذج (LightRes-SE) النموذج المعلّم) إلى نموذج (AttenDW-CBAM) النموذج الطالب لم يحافظ على دقة التصنيف فحسب، بل حسّنها بشكل طفيف لتصل إلى 0.9815، مع تقليل حجم النموذج وزمن الاستدلال بنسبة تجاوزت 50%. كما أسهم التكميم اللاحق باستخدام FP16 في تقليل الحجم وزمن الحساب بشكل أكبر، مع فقدان طفيف في الدقة لم يتجاوز 3%. علاوة على ذلك، أكدت مقاييس تقييم إضافية، شملت معامل ماثيو للارتباط (MCC)، ومعامل كابا (Kappa)، ومعامل معايرة الخطأ (ECE)، اتساق وموثوقية التنبوءات. كما استُخدمت تقنية Grad-CAM لتفسير قرارات النموذج بصريًا. وقد أثبت النظام المتكامل قدرته على تقليل مخاطر الإصابة بالأمراض من مرحلة التنبؤ المبكر إلى مرحلة الكشف البصري الفعلي.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة ديالى
كليـــة العلوم

التنبؤ بالأمراض والكشف عنها في الزراعة العراقية باستخدام إنترنت الأشياء القائم على تقنيات التعلم العميق

رسالة مقدمة الى قسم علوم الحاسوب / كلية العلوم / جامعة ديالى وهي جزء من متطلبات نيل درجة الدكتوراه في علوم الحاسوب

من قبل

مروان عدنان جاسم

بإشراف

أ.د جمال مصطفى التويجري

1447هـ **2025م**