Jimma University

Jimma Institute of Technology

Faculty of Computing and Informatics

M.Sc. in Computer Networking

Smart farming using IOT and machine learning in Jimma city Jiren kebele

MSc thesis

By

Zebider wuletaw

Advisor: Mr. Kebebew Ababu (Ass. Prof.)

Co-advisor:Worku Birhanie

December 2022 Jimma, Ethiopia

DECLARATION

I, the undersigned, declare that this thesis entitled Smart crop prediction using IOT and machine learning in Jimma city Jiren kebeleis my original work and has not been presented for a degree in this or any other universities, and all sources of references used for the thesis work have been appropriately acknowledged.

Name: Zebider wuletawSignature:________________________Date:______07/04/2022_____

This thesis has been submitted for examination with my approval as a University advisor and co-adviser

Advisor Name: Mr. kebebew ababu

Signature:____, Date:_______07/11/2022_________

Co-Advisor Name: Worku Birhanie

Signature: ____________________________________, Date :_____________

We the examiners approved that this thesis paper has passed through the review and defense process

	Name	Signature
1External examiner	Mekuwanint Agegnew (PhD),	______________________
2Internal Examiner	___________________________________,	______________________
3 Chairperson	___________________________________,	______________________

Dedication

——I dedicate this thesis for my beloved family——

Acknowledgments

I would like to thank my advisor Mr. Kebebew abebu (Ass. Prof.) and Mr. Worku Birhaniefor the assistance and comments that greatly improved the manuscript. I thank all my colleagues from Jimma University who provided insight and expertise that greatly assisted the research. I am also immensely grateful to all advisors and colleagues their sincere support, encouragement, and meaningful guidance on doing this work.

Abstract

Agriculture is one of the issues that no-way runs out to be examined. Since farming is one of the main sources of livelihood of the pastoral population. Thus, the construction of agrarian data-grounded internet of effects is veritably important to do. In Ethiopia collecting information like fertility of the soil, rainfall, growth of crops, temperature, downfall and information regarding colony of seeds, etc. can be collected with the help of IOT. It helps the growers to gain information regarding all farming conditions. With the help of internet technology, farming processes can be covered through detectors, smart cameras, mobile operations and bias like mini chips. Through IOT the automated internet technology, farmers can utilize farming resource efficiently.

In Jimma zone farmers rely on their experience to sow seeds on the soil. Without knowing what types of crops can be grown in the soil accurately, they cultivate the land and grow crops based on usual way of farming. Farmers do not try to crop new species in their land because if it fails to grow, they will lose revenue from farming. So, if there is a means that will provide the information to let them grow potential crops that will yield optimum production will be helpful to them. In addition the soil they cultivate may have enough minerals to grow surplus crops that are not grown in the region. Having the mineral content information of the soil in advance enhances optimum crop production.

Previously many researches have been done in crop prediction in another country. But up to my knowledge there is no existing crop prediction research that is done on Jimma zone in Jiren kebele. The data that is collected and the model that is built aimed for this kebele only. This research project is prediction of the types of crops that will grow in Jimma zone Jiren kebele. Based on the soil mineral and environmental condition data of the kebele a machine learning model was built to predict the types of crops that can be grown in the area. Different sensors such as moisture sensor, temperature sensor, pH sensor and NPK (nutrient sensor) were employed in the soil to get mineral and environmental condition data. And then decision tree algorithm was used to build a machine learning model that will predict the types of crops that will be grown in the soil. The average accuracy of the model in decision tree after 22 times run is 0.97.

Keywords: Wi-Fi, Arduino, IOT, PH sensor, Moisture sensor and NPK sensor.

Contents

List of figures

List of tables

List of Acronyms

IOT---Internet Of things.

LAN---Local area network.

WAN---wide area network.

NPK---Nitrogen Phosphorous Potassium

PH ---potential of hydrogen

ML---Machine learning

GUI ---Graphical user interface

GDP--Gross domestic product

RFID---Radio Frequency Identification

WSN---Wireless sensor network

MAC---Media access control

OSWA--Open Sensor Web Architecture

SOA--Service Oriented Architecture

OGC--Open Geospatial Consortium

PAN---Personal Area Network

M2M---machine to machine

GSM--Global system for mobile communication.

GPRS---General Packet Radio Services

LTE---Long termevolution.

LED--Light emitting diode.

IP---Internet protocol

PC---Personal computer

PCA---Principal component analysis

GPS--Global positioning system.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

The issue of agriculture is an important topic in particular in developing countries, this is because it will be influential in improving the quality of human resources and is also a means to maintain social and political stability[1][2]. The farm is the source of food security which became a fundamental right for every human being which must be fulfilled for the sake of maintaining survival. Ethiopia is a country which has a population of nearly the whole making of teff as a staple. Therefore, agriculture should be developed into a better and modern, especially in the use technology so that increased crop yields become larger and significant [3].

Agriculture is one of the economic populist bases in the world[4]. Agriculture also became the determining resilience, even food sovereignty. However, in the majority of fertile land from agricultural livelihoods depend is still not able to improve the standard of living that is more prosperous [5]. The crisis of the failed harvest always haunts in agricultural areas in several rural areas of agricultural producers[6]. So much is causing a growing rise in the price of agriculture in urban areas; it is swept up in the agricultural areas that are lacking in supplying his crop.

The utilization of technology is one of the conditions for making agriculture better because the data needed can be produced more so that the analysis will be better and more valid[7]. The utilization of technology in agriculture has so far been limited to the tools used in processing the land or processing crops, and is still very limited in producing data on the cropping process until there is harvest [8]. According to the data obtained by the researchers, there are still many farmers who use traditional methods in analyzing the planting process so that the analysis is still the same as previous farming methods[9].

Since the 1960s, the Internet, as a massive network of Local Area Networks (LAN) has been playing a vital role in connecting people, businesses and organizations together[10]. The Internethas removed geographical barriers between people and has given them a robust,efficient and cost effective means of communications. Now, it seems that things are about to change in the world of the Internet as a result of theappearance of smart objects that have the capability of generating and communicating dataover the Internet in a similar way to humans[11]. Internet of Things is the latest technologyand systems that holds the potential to change our way of life. One can think of IoT as atechnology made of two components; “Things” and “Internet”. The “Things” refers to anyobject or device that has the ability to perceive or collect data about it or the surroundingenvironment. Depending on what type, smartness and capabilities of this object, the objectmay be able to analyze and act smartly with other objects using “Internet” as a network forcommunications [5].

Traditional farming is the most widely used and known technique in Ethiopia[12] .Hence quality of product and productivity has not been achieved as compared to the number of farmers and farm land which account 80% of our population. So some form of technology must be introduced to improve the quality of agriculture and productivity. This thesis tries to implement IoT based smart agriculture to improve product quality and productivity.

1.2 Statement of the problem

In the broadest sense the problem of traditional agriculture is seen in the fact that the less developed countries are finding it more and more difficult to feed them. Two immediate facts stand out. First, the gap between current production and food needs in developing countries is widening. The deficit is now approaching 25 million metric tons of food per year. Second, food production per unit of land and labor in certain developing countries is declining relative to population growth.

The global population is predicted to touch 9.6 billion by 2050 – this poses a big problem for the agriculture industry[13]. In spite of severe challenges like rising weather conditions, rising climate change, and farming’s environmental impact, the high demand for more food has to be met. To meet these increasing needs, agriculture has to turn to new technology. New smart farming applications relied on IOT technologies will enable the agriculture industry to reduce wastage and improve productivity[14]. In Ethiopia farmers depend on traditional farming method to support their family as well as consumers. But there are some problems using traditional methods those are; to water their farm crops they totally depend on rain water that means if there is no rain there is no production. Now days they are using irrigation method using water pump but there is no mechanism to check whether this irrigation water actually reached crop root or not .There is no method to check the water level is excess or too little to the crop. In addition there is lack of technique to monitor the atmospheric temperature and humidity. Most crops require certain ranges of temperature and humidity to yield optimum product[15]. Moreover, crops require certain levels of nutrients such as Potassium, Phosphorous and Nitrogen to give maximum outcome. Hence, there is no a mechanism to monitor those nutrient content in soil in traditional farming leading to poor production.

Today farming in developed countries is supported by emerging technology. The Internet Of Things is restructuring agricultural business by enabling farmers and agriculturalists to use agricultural resource efficiently and yield optimum farm product[11]. IoT application helps to present summarized and combined information such as temperature, moisture, ph and humidity data to farmers in a way that can be used to make decision about farm crops. Integrating IoT with World Wide Web makes farmers to see their farm land by picture or video at real-time. More over entire nutrient capacity and other farm parameters can be seen from web pages using their smart phone from anywhere. In addition farmers can easily control their irrigation pump from home using IoT technology [8].

In Jimma zone farmers rely on their experience to sow seeds on the soil. Without knowing what types of crops can be grown in the soil accurately, they cultivate the land and grow crops based on usual way of farming. Farmers do not try to crop new species in their land because if it fails to grow, they will lose revenue from farming. So, if there is a means that will provide the information to let them grow potential crops that will yield optimum production will be helpful to them. In addition the soil they cultivate may have enough minerals to grow surplus crops that are not grown in the region. Having the mineral content information of the soil in advance enhances optimum crop production.

Previously many researches have been done in crop prediction in another country. But up to my knowledge there is no existing crop prediction research that is done on Jimma zone in Jiren kebele. The data that is collected and the model that is built aimed for this kebele only. This research project is prediction of the types of crops that will grow in Jimma zone Jiren kebele. Based on the soil mineral and environmental condition data of the kebele a machine learning model was built to predict the types of crops that can be grown in the area

This paper focuses on developing a smart agriculture system to monitor temperature, moisture content, PH and nutrient content using Arduino Uno board, Temperature sensor, Moisture sensor, PH sensor and NPK(Nitrogen, Phosphorous and Potassium) sensor and predict crop type recommended for that soil.

Therefore, this research will attempt to answer the following research question.

- How to improve the productivity of farming by predicting types of crop that will grow in the soil of Jimma zone Jiren kebele?
- How to provide nutrient content, environmental condition and soil condition information report to farmers?

1.3 Objectives

1.3.1 General objective

The objective of the study is to develop a machine learning model that predicts crop type and provide nutrient information reportfor Jimma zone Jiren kebele area.

1.3.2 Specific objective

To achieve the above general objective the following specific objectives will be followed:

- To review previously done research in the area of smart agriculture.
- To collect nutrient data from the soil of the kebele.
- To collect environmental data from the kebele.
- To preprocess collected data from sensors.
- To build a ML model that predicts crop type that a soil grows.
- Test and evaluate the system.

1.4 Methodology of the Study

Different methodologies will be used in this research in order to achieve the general and specific objectives of this study. The methodologies that will be used in conducting this research are described as follows:

1.4.1 Literature review

In order to achieve the objectives of the study relevant literature have been used different books, journal articles and conference papers on the areas related to smart agricultureusing IOT to have a detail understanding of the area. Revising the literature is a key component of the research method. The literature is reviewed throughout the duration of our research with consideration of new published works where necessary.

1.4.2 Data sources and Data collection

Live data is collected from farm land using moisture sensor, NPK sensor, PH sensor and moisture sensor. The collected data is sent to web pages.

1.4.3 Data Analysis

The data collected from sensors is interpreted to give meaning to farmers. The analysis contains parameters such as; normal moisture content, excess moisture, high ph value, low ph value, Nitrogen Phosphorous levels in relation to types of crops requirement. Which particular parameter is suitable for which crop type is identified on this step.

1.4.3 Presenting report

Report is generated on web pages using tables, graphs and GUI method. The report contains the entire summary of nutrient, moisture, temperature, and PH value of the farm land. Providing recommendation about which type of crop is suitable for the farm land is crucial.

1.4.4 Testing and evaluating

The system will be evaluated on a farm land by agricultural experts and technology experts.

1.4 Scope and limitation of the study

The main intent of this thesis is to develop smart agriculture system using Internet of Things by examining the existing researches in the area and come up with precision farming suitable and applicable to mycountry (Ethiopia). This work focuses on smart agriculture that provides information about farm land in terms of PH, nutrient, moisture and temperature. This work doesn't include automating pesticide spray and bug control.

1.5 Application of Results

Agriculture accounts for 46.3 percent of the Ethiopia's Gross domestic Product (GDP), 83.9 percent of trades, and 80% of the labor force [16].Many other economic activities depend on agriculture, including marketing, processing, and export of agricultural products. Production is overwhelming of a subsistence nature and a huge portion of product exports are given by the little agrarian cash crop division. Main crops include coffee, pulses (e.g., beans), oilseeds, cereals, potatoes, sugarcane, and vegetables. Exports are almost entirely agricultural commodities, and coffee is the largest foreign exchange earner. Ethiopia is also Africa's second biggest maize producer. So agriculture is back bone of Ethiopia's economy, using traditional farming we cannot compete with developed world farm production[17]. So implementing smart agriculture in Ethiopia is not an option rather it is mandatory. This system is applicable where automation of farming is required to yield more efficient production.

1.7 Organization of the thesis

The rest of the thesis is organized as follows:

Chapter Two: focuses on the related work in the field of smart agriculture. It also discusses how smart agriculture systems are classified, what are the key functions of an Smart Agriculture, what are key the components of this system , deployment techniques of the system.

Chapter Three: here it discusses about related works that have significant relation with this thesis. Even if there are a number of works done on this area, this Chapter selects the most related works to our thesis and presents them.

Chapter Four: concentrates on providing a design for the proposed work. How different components of the system are assembled will be shown.

Chapter Five: the proposed system is implemented in this Chapter by applying the proposed architecture. Evaluation of the work using experimental analysis and comparing to others works is presented here.

Chapter Six: summarizes the contributions made in the thesis, and concludes based on the results obtained from the thesis work. Furthermore, new issues that have been surfacing while working on the thesis are suggested as future work.

CHAPTER 2

LITERATURE REVIEW

This chapter provides the overview of smart agriculture from literature survey reviewed. And shed light on definition, components, and challenges of Internet Of things from smart agriculture view.

2.1 Overview

IoT is considered as an interconnected and distributed network of embedded systems communicating through wired or wireless communication technologies [5]. It is also defined as the network of physical objects or things empowered with limited computation, storage, and communication capabilities as well as embedded with electronics (such as sensors and actuators), software, and network connectivity that enables these objects to collect, sometime process, and exchange data[18].The goal of the Internet of Things is to enable things to be connected anytime, anyplace, with anything and anyone ideally using any path/network and any service[15].

Potential applications of the IoT are numerous and diverse, permeating into practically all areas of every-day life of individuals, enterprises, and society as a whole. The IoT application covers "smart" environments/spaces in domains such as: Transportation, Building, City, Lifestyle, Retail, Agriculture, Factory, Supply chain, Emergency, Healthcare, User interaction, Culture and tourism, Environment and Energy[19].

Internet of things in smart agriculture is being practiced worldwide nowadays[15]. In Green Houses it is used to Control micro-climate conditions to maximize the production of fruits and vegetables and its quality. In Compost is utilized in Control of moisture and temperature levels in in alfalfa, hay, straw, etc. to prevent fungus and other microbial contaminants[20]. In Animal Farming/Tracking it is used in Location and identification of animals grazing in open pastures or location in big stables, Study of ventilation and air quality in farms and detection of harmful gases from excrements, Offspring Care: Control of growing conditions of the offspring in animal farms to ensure its survival and health, field Monitoring[20]: Reducing spoilage and crop waste with better monitoring, accurate ongoing data obtaining, and management of the agriculture fields, including better control of fertilizing, humidity, temperature, nutrient, electricity and irrigation monitoring[21].

2.2 Characteristics of IoT

The distinctive features of the IoT are as follows [22]:

- **Interconnectivity**: With regard to the IoT, anything can be interconnected with the global information and communication infrastructure.
- **Things-related services:** The IoT is capable of providing thing-related services within the constraints of things, such as privacy protection and semantic consistency between physical things and their associated virtual things. In order to give thing-related services within the constraints of things, both the technologies in a physical world and an information world will change.
- **Heterogeneity**: IoT devices are heterogeneous as based on different hardware platforms and networks. They can interact with other devices or service platforms through different networks.
- **Dynamic changes**: The state of devices change dynamically, e.g., sleeping and waking up, connected and/or disconnected as well as the context of devices including location and speed. Moreover, the number of devices can change dynamically.
- **Enormous scale**: The number of devices that need to be managed and that communicate with each other will be at least an order of magnitude larger than the devices connected to the current Internet. Even more critical will be the management of the data generated and their interpretation for application purposes. This relates to semantics of data, as well as efficient data handling.
- **Safety**: As we gain benefits from the IoT, we must not forget about safety. As both the creators and recipients of the IoT, we must design for safety. This includes the safety of our personal data and the safety of our physical well-being. Securing the endpoints, the networks, and the data moving across all of it means creating a security paradigm that will scale.

- **Connectivity**: Connectivity enables network accessibility and compatibility. Accessibility is getting on a network while compatibility provides the common ability to consume and produce data.

2.3 IoT Elements

We present a taxonomy that will aid in defining the components required for Internet of Things from a high level perspective. Specific taxonomies of each component can be found elsewhere [23]. There are three IoT components which enables seamless ubicomp: a) Hardware - made up of sensors, actuators and embedded communication hardware b) Middleware - on demand storage and computing tools for data analytics and c) Presentation - novel easy to understand visualization and interpretation tools which can be widely accessed on different platforms and which can be designed for different applications. In this section, we discuss a few enabling technologies in these categories which will make up the three components stated above.

2.3.1 Radio Frequency Identification (RFID)

RFID technology is a major breakthrough in the embedded communication paradigm which enables design of microchips for wireless data communication[24]. They help in automatic identification of anything they are attached to acting as an electronic barcode[23]. The passive RFID tags are not battery powered and they use the power of the reader‘s interrogation signal to communicate the ID to the RFID reader. This has resulted in many applications particularly in retail and supply chain management[25]. The applications can be found in transportation (replacement of tickets, registration stickers) and access control applications as well. The passive tags are currently being used in many bank cards and road toll tags which is among the first global deployments. Active RFID readers have their own battery supply and can instantiate the communication. Of the several applications, the main application of active RFID tags is in port containers[26] for monitoring cargo.

2.3.2 Wireless Sensor Networks (WSN)

Recent technological advances in low power integrated circuits and wireless communications have made available efficient, low cost, low power miniature devices for use in remote sensing applications. The combination of these factors has improved the viability of utilizing a sensor network consisting of a large number of intelligent sensors, enabling the collection, processing, analysis and dissemination of valuable information, gathered in a variety of environments [27]. Active RFID is nearly the same as the lower end WSN nodes with limited processing capability and storage. The scientific challenges that must be overcome in order to realize the enormous potential of WSNs are substantial and multidisciplinary in nature [27]. Sensor data are shared among sensor nodes and sent to a distributed or centralized system for analytics. The components that make up the WSN monitoring network include:

a) WSN hardware - Typically a node (WSN core hardware) contains sensor interfaces, processing units, transceiver units and power supply. Almost always, they comprise of multiple A/D converters for sensor interfacing and more modern sensor nodes have the ability to communicate using one frequency band making them more versatile [23].

b) WSN communication stack - The nodes are expected to be deployed in an adhoc manner for most applications. Designing an appropriate topology, routing and MAC layer is critical for scalability and longevity of the deployed network. Nodes in a WSN need to communicate among themselves to transmit data in single or multi-hop to a base station. Node drop outs, and consequent degraded network lifetimes, are frequent. The communication stack at the sink node should be able to interact with the outside world through the Internet to act as a gateway to the WSN subnet and the Internet [21].

c) Middleware - A mechanism to combine cyber infrastructure with a Service Oriented Architecture (SOA) and sensor networks to provide access to heterogeneous sensor resources in a deployment independent manner[28]. This is based on the idea to isolate resources that can be used by several applications. A platform independent middleware for developing sensor applications is required, such as an Open Sensor Web Architecture (OSWA) [4]. OSWA is built upon a uniform set of operations and standard data representations as defined in the Sensor Web Enablement Method (SWE) by the Open Geospatial Consortium (OGC).

d) Secure Data aggregation - An efficient and secure data aggregation method is required for extending the lifetime of the network as well as ensuring reliable data collected from sensors [29]. Node failure being a common characteristic of WSNs, the network topology should have the capability to heal itself. Ensuring security is critical as the system is automatically linked to actuators and protecting the systems from intruders becomes very important.

2.3.3 Data storage and analytics

One of the most important outcomes of this emerging field is the creation of an unprecedented amount of data. Storage, ownership and expiry of the data become critical issues. The internet consumes up to 5% of the total energy generated today and with these types of demands, it is sure to go up even further[30]. Hence data centers which run on harvested energy and which are centralized will ensure energy efficiency as well as reliability. The data have to be stored and used intelligently for smart monitoring and actuation. It is important to develop artificial intelligence algorithms which could be centralized or distributed based on the need. Novel fusion algorithms need to be developed to make sense of the data collected. State-of-the-art nonlinear, temporal machine learning methods based on evolutionary algorithms, genetic algorithms, neural networks, and other artificial intelligence techniques are necessary to achieve automated decision making. These systems show characteristics such as interoperability, integration and adaptive communications[31]. They also have a modular architecture both in terms of hardware system design as well as software development and are usually very well-suited for IoT applications

2.3.4 Visualization

Visualization is critical for an IoT application as this allows interaction of the user with the environment. With recent advances in touch screen technologies, use of smart tablets and phones has become very intuitive[20]. For a lay person to fully benefit from the IoT revolution, attractive and easy to understand visualization have to be created. As we move from 2D to 3D screens, more information can be provided to the user in meaningful ways for consumers. This will also enable policy makers to convert data into knowledge which is critical in fast decision making[32]. Extraction of meaningful information from raw data is nontrivial. This encompasses both event detection and visualization of the associated raw and modeled data, with information represented according to the needs of the end user.

2.4Architecture of IoT

IOT architecture consists of different layers of technologies supporting IOT. It serves to illustrate how various technologies relate to each other and to communicate the scalability, modularity and configuration of IOT deployments in different scenarios. Figure 1 shows detailed architecture of IOT. The functionality of each layer is described below:

Figure 1IoT architecture

2.4.1 Smart device or sensor layer

The lowest layer is made up of smart objects integrated with sensors. The sensors enable the interconnection of the physical and digital worlds allowing real-time information to be collected and processed. There are various types of sensors for different purposes. The sensors have the capacity to take measurements such as temperature, air quality, speed, humidity, pressure, flow, movement and electricity etc. In some cases, they may also have a degree of memory, enabling them to record a certain number of measurements. A sensor can measure the physical property and convert it into signal that can be understood by an instrument. Sensors are grouped according to their unique purpose such as environmental sensors, body sensors, home appliance sensors and vehicle telematics sensors, etc. Most sensors require connectivity to the sensor gateways. They can use Local area network; such as Ethernet and Wi-Fi connections or Personal Area Network (PAN) such as ZigBee, Bluetooth and Ultra-Wideband(UWB). Sensors that don't need network to sensor cluster head, their network to backend servers can be given utilizing Wide Area Network such as LTE,GSM, and GPRS. Sensors that use low power and low data rate connectivity, they typically form networks commonly known as wireless sensor networks (WSNs).. WSNs are becoming popular due to allowing far more sensor nodes while providing adequate battery life and covering large geographical areas.

2.4.2 Gateways and Networks

To transport large amount of data generated by these small sensors requires a high performance wired or wireless network infrastructure. Very different protocols have been used to allow machine-to-machine (M2M) networks and their applications.

With demand needed to serve a wider range of IOT services and applications such as high speed transactional services, context-aware applications, etc, multiple networks with various technologies and access protocols are needed to work with each other in a heterogeneous configuration. These networks can be in the form of a private, public or hybrid models and are built to support the communication requirements for latency, bandwidth or security.

2.4.3 Management Service Layer

The management service renders the processing of information possible through analytics, security controls, process modeling and management of devices. One of the important features of the management service layer is the business and process rule engines. IOT brings connection and interaction of objects and systems together providing information in the form of events or contextual data such as temperature of goods, current location and traffic data. Some of these events require filtering or routing to post processing systems such as capturing of periodic sensory data, while others require response to the immediate situations such as reacting to emergencies on patient's health conditions. The rule engines support the formulation of decision logics and trigger interactive and automated processes to enable a more responsive IOT system. In the area of analytics, various analytics tools are used to extract relevant information from massive amount of raw data and to be processed at a much faster rate. Analytics such as in memory analytics allows large volumes of data to be cached in random access memory (RAM) rather than stored in physical disks. In-memory analytics reduces data query time and augments the speed of decision making. Streaming analytics is another form of analytics where analysis of data, considered as data-in-motion, is required to be carried out in real time so that decisions can be made in a matter of seconds.

Data management is the capability to control data information flow. With data management in the management service layer, information can be accessed, integrated and controlled. Higher layer applications can be shielded from the need to process unnecessary data and reduce the risk of privacy disclosure of the data source. Data filtering methods such as data integration and data synchronization are utilized to abstract the details of the information while giving only useful information that is usable for the relevant applications. With the use of data abstraction, information can be extracted to provide a common business view of data to gain greater agility and reuse across domains. Security must be enforced across the whole dimension of the IOT architecture right from the smart object layer all the way to the application layer. Security of the system prevents system hacking and compromises by unauthorized personnel, thus reducing the possibility of risks.

2.4.4 Application Layer

The IoT application covers "smart" environments/spaces in domains such as: Transportation, Building, City, Lifestyle, Retail, Agriculture, Factory, Supply chain, Emergency, Healthcare, User interaction, Culture and tourism, Environment and Energy.

2.5Challenges of IoT

There are key challenges and implications today that need to be addressed before mass adoption of IOT can occur [33].

2.5.1 Privacy and Security

As the IoT become a key element of the Future Internet and the usage of the Internet of Things for large-scale, partially mission-critical systems creates the need to address trust and security functions adequately. New challenges identified for privacy, trust and reliability are:

- Providing trust and quality of information in shared information models to enable re-use across many applications.
- Providing secure exchange of data between IoT devices and consumers of their information.
- Providing protection mechanisms for vulnerable devices

2.5.2 Cost versus Usability

IOT uses technology to connect physical objects to the Internet. For IOT adoption to grow, the cost of components that are needed to support capabilities such as sensing, tracking and control mechanisms need to be relatively inexpensive in the coming years.

2.5.3 Interoperability

In the traditional Internet, interoperability is the most basic core value; the first requirement of Internet connectivity is that “connected” systems be able to “talk the same language” of protocols and encodings[34]. Different industries today use different standards to support their applications. With numerous sources of data and heterogeneous devices, the use of standard interfaces between these diverse entities becomes important. This is especially so for applications that supports cross organizational and various system boundaries. Thus the IOT systems need to handle high degree of interoperability.

2.5.4 Data Management

Data management is a crucial aspect in the Internet of Things. When considering a world of objects interconnected and constantly exchanging all types of information, the volume of the generated data and the processes involved in the handling of those data become critical.

2.5.5 Device Level Energy Issues

One of the essential challenges in IoT is how to interconnect “things” in an interoperable way while taking into account the energy constraints, knowing that the communication is the most energy consuming task on devices.

2.6 IoT Sensors

There are different types of sensors based on their there pupose. Here are the sensors that are most widely used int this research project.

2.6.1 Node Mcu

According to [32] Node Mcu is an open source programmable circuit board that can be integrated into a wide variety of makerspace projects both simple and complex. This board contains a microcontroller which is able to be programmed to sense and control objects in the physical world. By responding to sensors and inputs, the Node Mcu is able to interact with a large array of outputs such as LEDs, sensors, motors and displays. Node Mcu acts as a coordinator that receives sensed data from sensors and forwards the data to processing server. The following figure shows a Node Mcu.

Figure 2 Node Mcu

2.6.2 Moisture sensor

The soil moisture sensor is used to collect moisture content of a soil and transfer the collected data to Arduino Uno micro controller. The following figure shows moisture sensor;

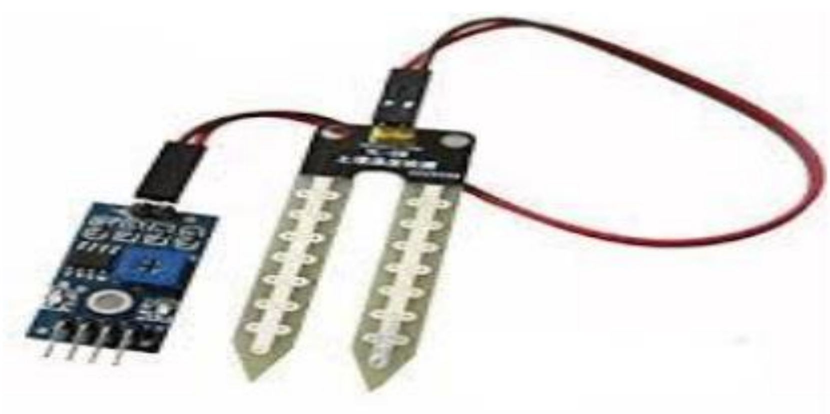

Figure 3Moisture sensor

2.6.3 Temperature sensor

Temperature sensor retrieve the current temperature level and it approximates the result in degree centigrade (C^o).The following figure shows temperature sensor;

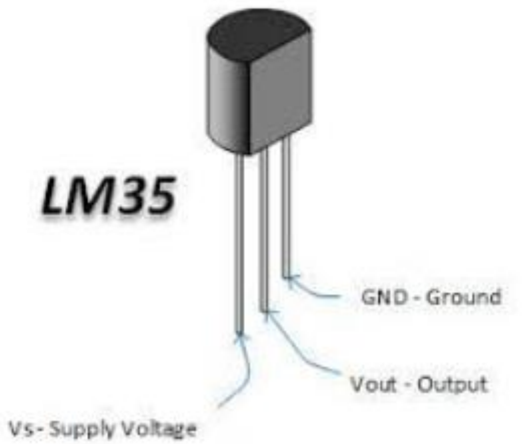

Figure 4Temperature sensor

2.6.4 Nutrient sensor

In this research project we use NPK (Nitrogen, Phosphorous and Potassium sensor). This sensor collects soil nutrient content and fertilizer concentration in the soil.

2.6.5 PH sensor

PH sensor is used to measure soil acidic or basic property and sends the measurement to Arduino Uno board.

2.6.6 Wi-Fi module

Wi-Fi module is used to send the collected data to remote server such as web server or database server.

2.7Application of IoT inSmart Agriculture

2.7.1 Smart Farming

Smart Farming is a farming management concept using modern technology to increase the quantity and quality of agricultural products[35]. Farmers in the 21st century have access to GPS, soil scanning, data management, and Internet of Things technologies. In IOT-based smart agriculture, a system is built for monitoring the crop field with the help of sensors (light, humidity, temperature, soil moisture, etc.) and automating the irrigation system. IoT (Internet of things) in an agricultural context refers to the use of sensors, cameras, and other devices to turn every element and action involved in farming into data [16].

The goal of smart agriculture research is to ground a decision making support system for farm management[28]. Smart farming deems it necessary to address the issues of population growth, climate change and labor that has gained a lot of technological attention, from planting and watering of crops to health and harvesting.

IoT devices are the basic enabling modules that are used to implement smart agriculture. The following devices are most commonly used IoT sensors used in smart agriculture:

- Temperature Sensor
  This sensor is mainly used inside the warehouse, and connected to the microcontroller. It detects temperature inside the warehouse, which can then be modified by switching the heater ON/OFF.
- Humidity sensor
  This sensor is used in the warehouse to detect the humidity inside. Accordingly, the cooling fan is switched ON/OFF.
- Moisture sensor
  The soil moisture sensor is used to sense the moisture in the soil of the field and transfer it to microcontroller of block 2 in order to take controlling action of switching water pump ON/OFF[36]. It generates a digital signal when the moisture in the soil increases above some threshold value. The automatic self-watering plant is designed using this sensor with Arduino. Microscope sensor is used to detect the sound. It generates a signal when the intensity of detected sound increases beyond some threshold value[36].

2.7.2 IoT in Crop and Water Management

Generally, a farmer fed either less or more water to cultivate his field which results in the problem of wastage of water or inadequate to the crops[37]. It sends a notification message to farmers when the level of moisture increases or decreases. Agriculture in IOT is connecting with both Web Map Service and Sensor Observation Service to ensure proper management of water for farming and removes the problem of wastage of water[38].

Advantages

- Generate Extra income through cost savings

- Increasing efficiency
- Perfect time, information about water usage data and analysis for smart and fast decisions.

Disadvantages

- Lack of indicator lights
- Neither Warranty nor Guarantee
- More difficulty in installation

2.7.3 IoT in Pest Controlling and Its Working

Sometimes farmer's hard work is damaged by insects and pests therefore results in big loss to farmers. To prevent this problem, IOT in agriculture has a system that detects their movement of predators using PIR sensors [39]. This information can be made for the farmers to remove the damaging done by predators.

Advantages

- Automatic decision making
- Beneficial for nature
- Reduces the hazards of pesticides

Disadvantages

- It consumes energy and time
- More involvement in the methods of technicalities

2.7.4 IoT in Precision Agriculture

High reliability is required about whether the information which decreases the problems of crop damage. Agriculture in IOT makes exact delivery of real-time information on the weather changing, soil quality, labor cost and many more to farmers[17].

Advantages

- It provides the farming process more accurate.
- It controls for increasing warehouse and growing ofcrops

Disadvantages

- Precision farming cannot be utilized completely in everycrop.
- Needs to install different base stations

CHAPTER 3

RELATED WORK

Internet of things is rapidly being applied in smart farming to improve the quality of farm and farm product. The following section presents review of papers that are related to smart farming in IoT.

In [2] tried to implement smart agriculture monitoring using IoT. They used four sensors namely water pH sensor; temperature and humidity sensors; soil pH sensor; and groundwater level sensors. They also used solar panel for energy source to provide current for sensors. The project also has front end system to help users to easily interact with system. For famers they designed an android app to make report available on smart phone. It has also web based interface that can be viewed on web browsers. Generally they did a good job in integrating hardware and software together. But the system only provides current report of farm land it doesn't provide recommendation to farmers what to do based on the presented data. The other thing is that they didn't other sensors such as NPK (nitrogen phosphorus potassium) sensors that could easily detect the concentration of fertilizers in farm land.

In [4] this paper smart farming using IoT is presented. The goal of this paper is to implement a smart vehicle that is equipped with GPS, Moisture sensor, cutter, sprayer, and obstacle sensor to perform multipurpose task. The first task is to detect moisture content of soil if low it spray water by itself .The other task is it can weed the farm and spray fertilizers. In addition it used to scare birds, other crop eating animals and prevents theft. And finally the vehicle can be controlled remotely using computer at home. The researchers did good job in implementing weeding, watering, fertilizer spraying and bird scaring in single smart vehicle. But the sensed data about farm land that can be used for future recommendations is not recorded in database. And it doesn't provide general report about the farm land.

In [5] the authors implemented smart agriculture using IoT technology. They used soil moisture sensor, humidity sensor and Temperature sensors to get reading from farm land. The collected sensor data from sensors is sent to a website called www.thingspeak.com using Wi-Fi module. This website is an official website that can easily connect to our sensor and get report from it. But the report it generates cannot be customized as needed and new features cannot be added as farmers need and requirement. So, it is better to have its own web page than using public website to get a sensor data report.

In [7] the authors tried to implement smart agriculture using IoT. They used to sensors namely soil moisture sensor, water level sensors and Arduino Uno board. But one cannot get entire farm land soil condition using only two sensors it is better to add temperature sensor, ph sensor and nutrient sensor (NPK). This paper doesn’t show how they tried to implement it in real farm land. The project also lacks front end user interface for farmers so they can easily interact with system.

In [1] the authors discussed the relevance of IoT in agriculture. They pointed out the applications of IoT in agriculture. The main intent of this paper is to explain the advantages, scope and limitations of using IoT technology in farming. They also elaborated the architecture of IoT and how its different components are interconnected together. But no implementation detail was discussed on the research work.

In [28] the role of Internet of Things and robotics in agriculture were discussed. Big data processing applications in smart farming is also elaborated in this research work. The authors tried to point out using satellite data in smart farming plays a major role in improving efficiency of smart farming. They suggested using Artificial Intelligence to generate report and recommendation from sensor collected farm data brings better performance for the system.

In [30] the aim of the paper is to design IoT based system for operation of agricultural activities which is scalable and cost effective. But it only describes the advantage of wireless network in agriculture. It doesn’t discuss what type of sensors they used to collect farm land data. They do not elaborated how they integrated the different components of wireless sensor network (WSN).

CHAPTER 4

Proposed System Architecture

4.1 Introduction

In this chapter the proposed architecture for smart agriculture using IoT will be discussed. Different components of the proposed architecture are described with their relevance and techniques to use while building those components. The Chapter presents the architecture with implemented algorithms. It uses Node Mcu board to receive sensor data and Wi-Fi module to send those data into MySQL database. The collected data is used to build machine learning model that can predict the types of crops it will grow.

4.2 System Architecture

In this Section, I have proposed architecture for smart agriculture system that can increase the performance of the existing traditional farming. It has different components as shown in Figure below;

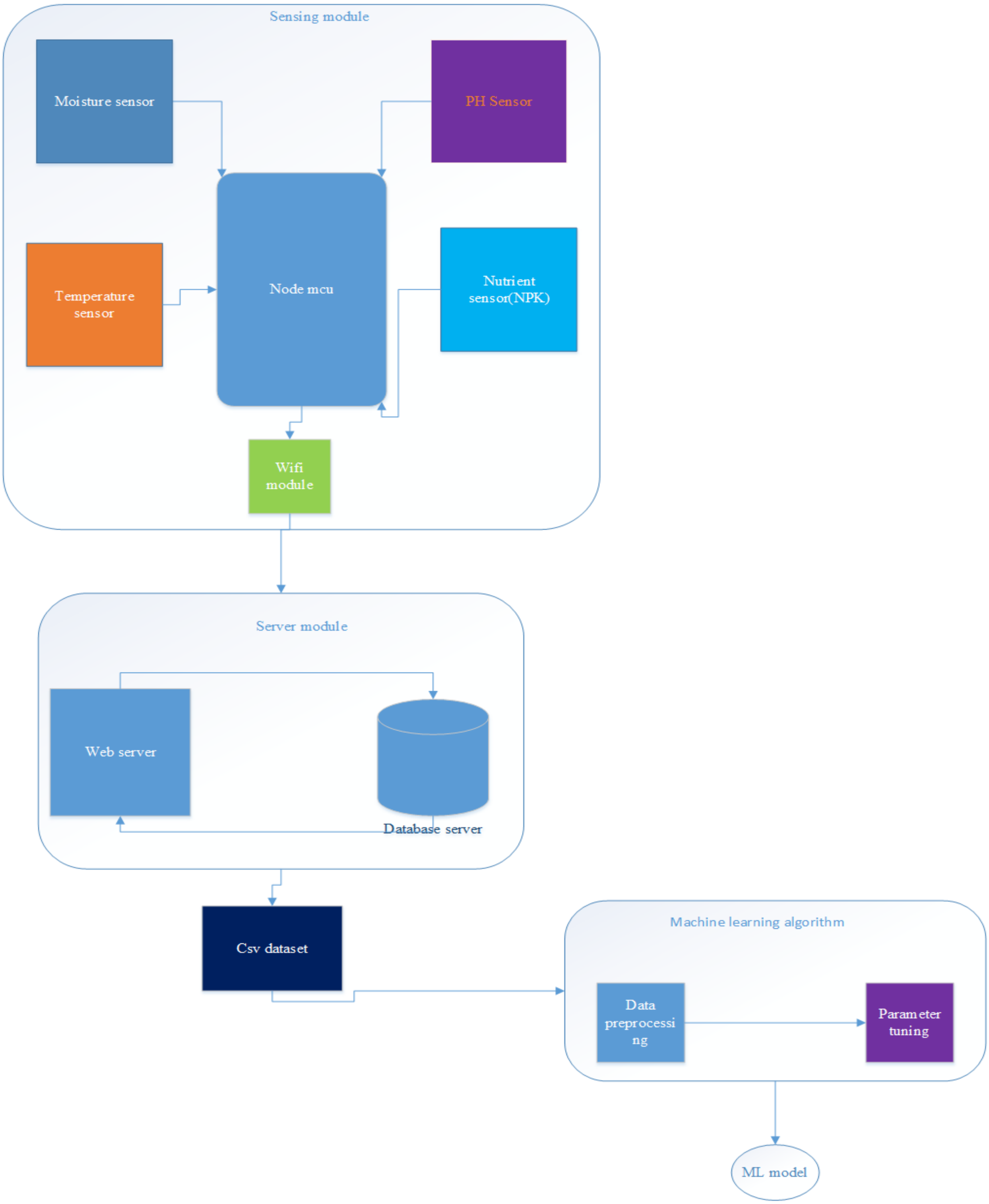

Figure 5 Proposed system architecture.

4.2.1 Components of the Proposed System

4.2.1.1Machine learning model

Crop prediction machine learning model is used to predict what types of crops can a soil grows using environmental and chemical components. The chemical components include Nitrogen, Phosphorous and Potassium. The environmental characteristics includes; temperature, humidity, ph and rainfall levels. Crop prediction machine learning model was built using decision tree mathematical algorithm in Matlab 2018 software. To build this model many steps were followed. The first step is data collection; it is process that I used many sensors to collect soil characteristics data. The datacollected contains 2,202 raw data with 22 classes of crops.Eighty percent of the dataset is used for training and twenty percent for testing purpose.The second step is cleaning or data preprocessing step, in this step the data is cleaned to make it ready to be processed. Next step is featureextraction; this step involves selection of important variables from the data and removing redundant features from it. Finally, parameter tuning, this step involves tuning of machine learning model parameter to get the best result out of it. The detail discussion of the above components is discussed in the following section.

4.2.1.1.1 Data preprocessing

First data is collected from sensing module using different sensors from a soil. The sensing module contains moisture sensor, temperature sensor and nutrient sensors to collect chemical and environmental data from the soil. The sensed data is sent to MySQL database to be stored as a raw data. The raw data is exported from MySQL database as .csv file dataset. This dataset is raw data so; removal of empty values in the data is performed. Or population of empty values by median values is done. Removing of redundant or duplicate entries was carried out. Visualization of dataset distribution as a graphical plot and summary was done to decide what kind of machine learning model to use.

4.2.1.1.2 Parameter tuning

Parameter tuning or hyper parameter optimization is the process of fine tuning the machine learning parameters. In my case I tried out to tune decision tree criterion, random state, depth,and information gain. Parameter optimization plays great role for improving the performance of the machine learning model.

4.2.1.2 Sensing module

This module contains components that are responsible to collect environmental and soil data. These components are sensors and central board which are used to collect data from a soil. I used the sensors in sensing model to get environmental and nutrient data from the soil. The collected data is then passed to the server module via Wi-Fi module.

4.2.1.2 Server module

The server module contains two servers namely web server and database server. They are used to store or process the received sensor data.

4.2.1.2.1 Web server

Web server is used to provide access to data stored in the database. It usually contains code to provide user interface to end users. And it also performs some processing to provide report users. We can retrieve data from web server by using PC or Smart phone by providing the web address of web server (IP/Domain name).

4.2.1.2.2 Database server

Database is used to store collected data from sensors on permanent basis. Additional report could be generated from the database as needed. Dataset that will be used to build a machine learning model is extracted from this database server.

CHAPTER 5

Implementation and Performance Evaluation

5.1 Overview

I used NodeMcu, moisture sensor, temperature sensor and nutrient sensor to collect dataset from soil in Jimma. The collected data is stored in MySQL database, so it can be retrieved easily using apache web server. The implementation took place using NodeMcu, sensors, apache web server, and MySQL database in xampp package. Dataset cleaning and preprocessing took place using excel file and Matlab script. The dataset training took place using Matlab 2018a software through decision tree algorithm.

5.2 Tools used

Different tools have been used for implementing the proposed architecture. The tools are listed below:

- Moisture sensor, Temperature sensor and Humidity sensor to collect weather data
- Matlab for data preprocessing and training
- Excel to process csv dataset exported from MySQL database
- MySQL package to store collected data

5.3Data collection

I used NodeMcu, moisture sensor, temperature sensor and nutrient sensor to collect dataset from soil in Jimma. The collected data is stored in MySQL database, so it can be retrieved easily using apache webserver. The dataset contains 22 classes of crops and their respective nutrient requirement. The nutrients are Nitrogen, Potassium, and Phosphorous. In addition it contains rainfall level, Temperature and humidity of the environment.

The following figure shows data collection scenario:

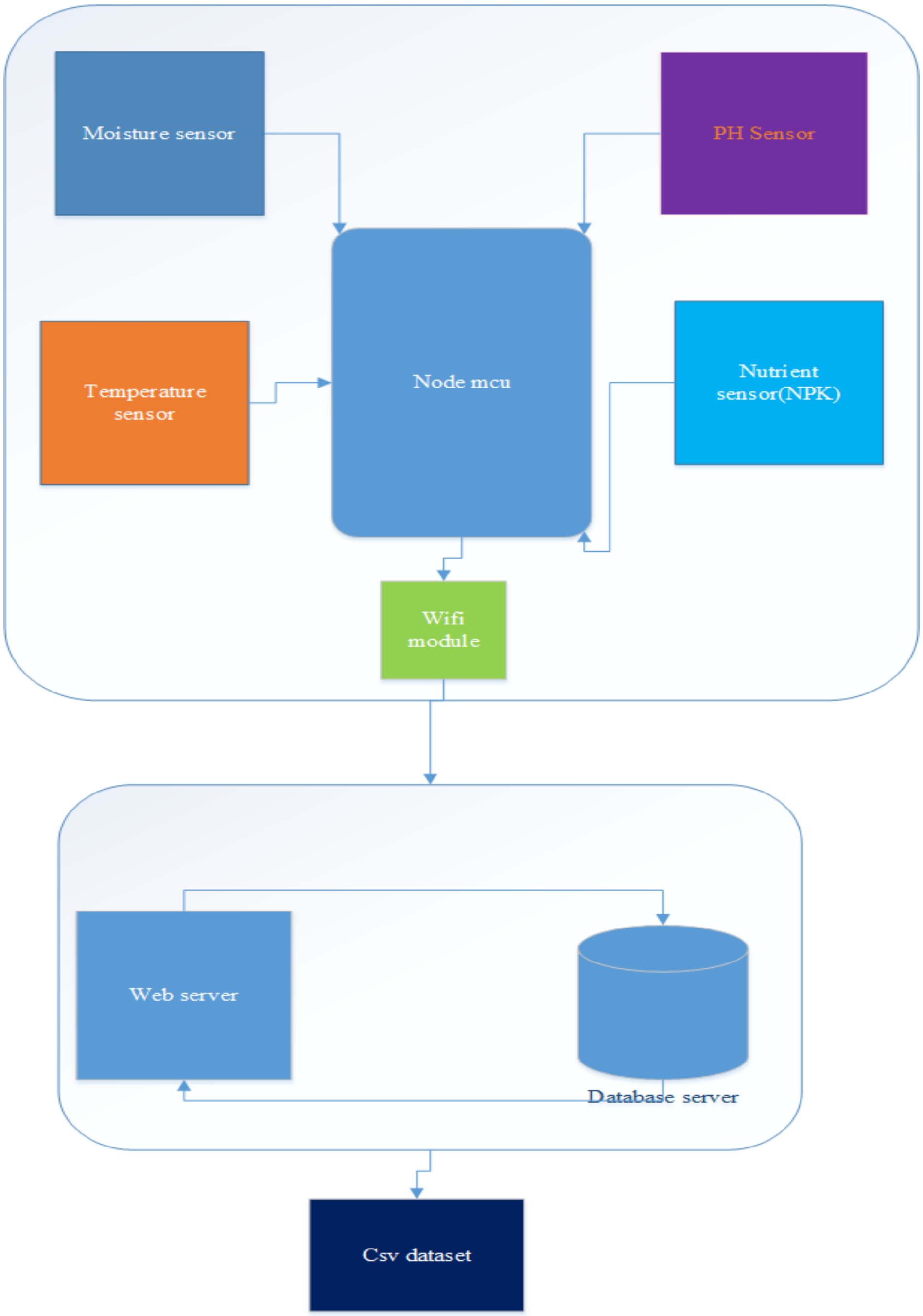

Figure 6data collection scenario

5.3.1 Dataset description

The dataset consists of 2,201entries from which contains equally distributed percentages of apple, banana, black gram, chickpea, coconut, coffee, cotton, grapes, jute, kidney-beans, lentil, maize, mango, moth-beans, mung-bean, muskmelon, orange, papaya, pigeon-peas, pomegranate, rice, and watermelon.

Table 1Dataset distribution.

Crop type	Count	Percentage
apple	101	4.5
banana	101	4.5
Blackgram	101	4.5
chickpea	101	4.5
coconut	101	4.5
coffee	101	4.5
cotton	101	4.5
grapes	101	4.5
jute	101	4.5
kidney-beans	101	4.5
lentil	101	4.5
watermelon	101	4.5
maize	101	4.5
mango	101	4.5
moth-beans	101	4.5
mung-bean	101	4.5

muskmelon,	101	4.5
orange	101	4.5
papaya	101	4.5
pigeon-peas	101	4.5
pomegranate	101	4.5
rice	101	4.5
Total		2201

Table 2 training dataset distribution

Crop type	Count	Percentage
apple	75	3.4
banana	75	3.4
Black-gram	75	3.4
chickpea	75	3.4
coconut	75	3.4
coffee	75	3.4
cotton	75	3.4
grapes	75	3.4

jute	75	3.4
kidney-beans	75	3.4
lentil	75	3.4
watermelon	75	3.4
maize	75	3.4
mango	75	3.4
moth-beans	75	3.4
mung-bean	75	3.4
muskmelon,	75	3.4
orange	75	3.4
papaya	75	3.4
pigeon-peas	75	3.4
pomegranate	75	3.4
rice	75	3.4
Total		1,650

Table 3 testing dataset distribution

Crop type	Count	Percentage
apple	101	4.5
banana	101	4.5
Blackgram	101	4.5
chickpea	101	4.5
coconut	101	4.5
coffee	101	4.5
cotton	101	4.5
grapes	101	4.5
jute	101	4.5
kidney-beans	101	4.5
lentil	101	4.5
watermelon	101	4.5
maize	101	4.5
mango	101	4.5
moth-beans	101	4.5
mung-bean	101	4.5
muskmelon,	101	4.5

orange	101	4.5
papaya	101	4.5
pigeon-peas	101	4.5
pomegranate	101	4.5
rice	101	4.5
Total		2201

Out of 2201 seventy five percent of the dataset is used for training purpose where as 25 % of the dataset is used for testing purpose.

5.3.2 Dataset Features

The collected data has 7 features which are listed in the following table below.

Table 4 packet capture features

Feature name	**Feature description**
N	Nitrogen content of a soil
P	Phosphorous content of a soil
K	Potassium content of a soil
Humidity	Humidity of the environment
Temperature	Temperature of the environment
Rainfall	Rainfall of the environment

PH	PH value of a soil.

5.4 Data preprocessing

Data preprocessing or data cleaning is an important part of building machine learning model. It is preparing the data in way that can be fed into data processing algorithms.

The collected data is in a csv format, the processing task is filling empty values with zero, converting text values into numbers, dropping duplicate values and removing rows which have only single entry and removing entry that has low variance.

5.5 Training

The preprocessed dataset is fed into machine learning algorithm to build crop type classification model. The decision tree algorithm is preferred machine learning algorithm in our case because it was able to give better accuracy than others available in Matlab. The tree has 1000 nodes and uses Gini Index as tree split criterion. It took 1.0561 seconds to train the entire dataset with prediction speed of 43,000 observations per second.

Table 5 parameter setting

Parameter	**Value**
Splits	100
Split criterion	Gini diversity index
Feature selection	Principal component analysis

5.6 Implementation of the components

We used Matlab 2018a for training our dataset using decision tree algorithm. The dataset collected using Node Mcu is stored in MySQL database in xampp data management system. The dataset exported from database is preprocessed using Excel and Matlab. The preprocessed dataset is fed into machine learning algorithm to produce classification model.

5.6.1 Training Matlab code

Training code used to train the model in Matlab is listed below. Detail configurations are defined in appendix1.

```
function [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)
inputTable = trainingData;
predictorNames = {'N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall'};
predictors = inputTable(:, predictorNames);
response = inputTable.label;
isCategoricalPredictor = [false, false, false, false, false, false, false];

% Train a classifier
% This code specifies all the classifier options and trains the classifier.
classificationTree = fitctree(...
    predictors, ...
    response, ...
    'SplitCriterion', 'gdi', ...
    'MaxNumSplits', 100, ...
    'Surrogate', 'off', ...
    'ClassNames', [1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 21; 22]);

% Create the result struct with predict function
predictorExtractionFcn = @(t) t(:, predictorNames);
treePredictFcn = @(x) predict(classificationTree, x);
trainedClassifier.predictFcn = @(x) treePredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct
```

```
trainedClassifier.RequiredVariables = {'N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall'};
trainedClassifier.ClassificationTree = classificationTree;
trainedClassifier.About = 'This struct is a trained model exported from Classification Learner R2018a.';
trainedClassifier.HowToPredict = sprintf('To make predictions on a new table, T, use: \n  yfit = c.predictFcn(T) \nreplacing ''c'' with the name of the variable that is this struct, e.g. ''trainedModel''. \n \nThe table, T, must contain the variables returned by: \n c.RequiredVariables \nVariable formats (e.g. matrix/vector, datatype) must match the original training data. \nAdditional variables are ignored. \n \nFor more information, see <a href="matlab:helpview(fullfile(docroot, ''stats'', ''stats.map''), ''appclassification_exportmodeltoworkspace'')">How to predict using an exported model</a>.');
% Extract predictors and response
% This code processes the data into the right shape for training the
% model.
inputTable = trainingData;
predictorNames = {'N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall'};
predictors = inputTable(:, predictorNames);
response = inputTable.label;
isCategoricalPredictor = [false, false, false, false, false, false, false];

% Set up holdout validation
cvp = cvpartition(response, 'Holdout', 0.25);
trainingPredictors = predictors(cvp.training, :);
trainingResponse = response(cvp.training, :);
trainingIsCategoricalPredictor = isCategoricalPredictor;

% Train a classifier
% This code specifies all the classifier options and trains the classifier.
classificationTree = fitctree(...
```

```
    trainingPredictors, ...
    trainingResponse, ...
    'SplitCriterion', 'gdi', ...
    'MaxNumSplits', 100, ...
    'Surrogate', 'off', ...
    'ClassNames', [1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 21; 22]);
% Create the result struct with predict function
treePredictFcn = @(x) predict(classificationTree, x);
validationPredictFcn = @(x) treePredictFcn(x);
% Add additional fields to the result struct
% Compute validation predictions
validationPredictors = predictors(cvp.test, :);
validationResponse = response(cvp.test, :);
[validationPredictions, validationScores] = validationPredictFcn(validationPredictors);
% Compute validation accuracy
correctPredictions = (validationPredictions == validationResponse);
isMissing = isnan(validationResponse);
correctPredictions = correctPredictions(~isMissing);
validationAccuracy = sum(correctPredictions)/length(correctPredictions);
```

5.7 Experiments and Results

This section discuses experiments performed while doing the thesis. The experiment involves getting better accuracy by altering the training parameters, algorithm and dataset.

Experiment1

Using decision tree algorithm with 100 splitswithout using PCA (principal component analysis) as feature selection we have the following result:

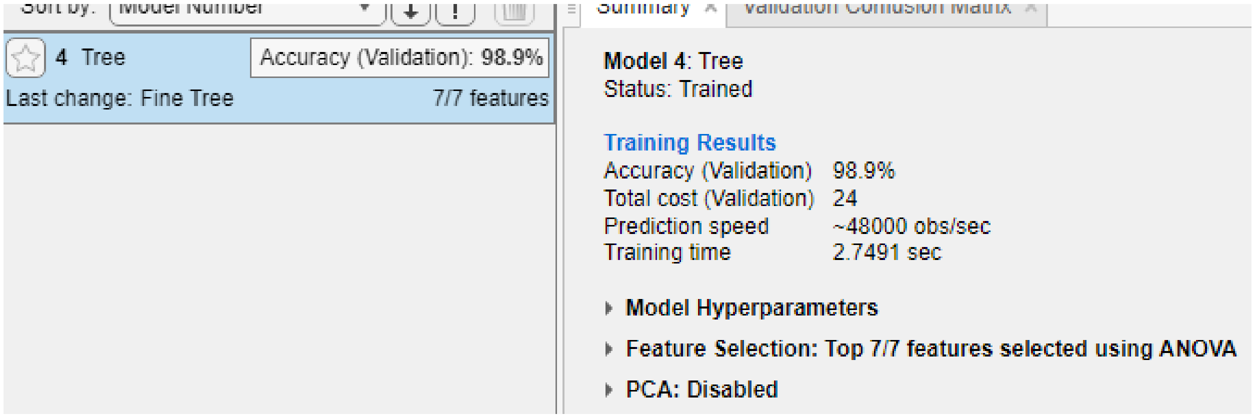

Figure 7 accuracy experiment 1

That means 98.9 %accuracy. The following Table shows the confusion matrix

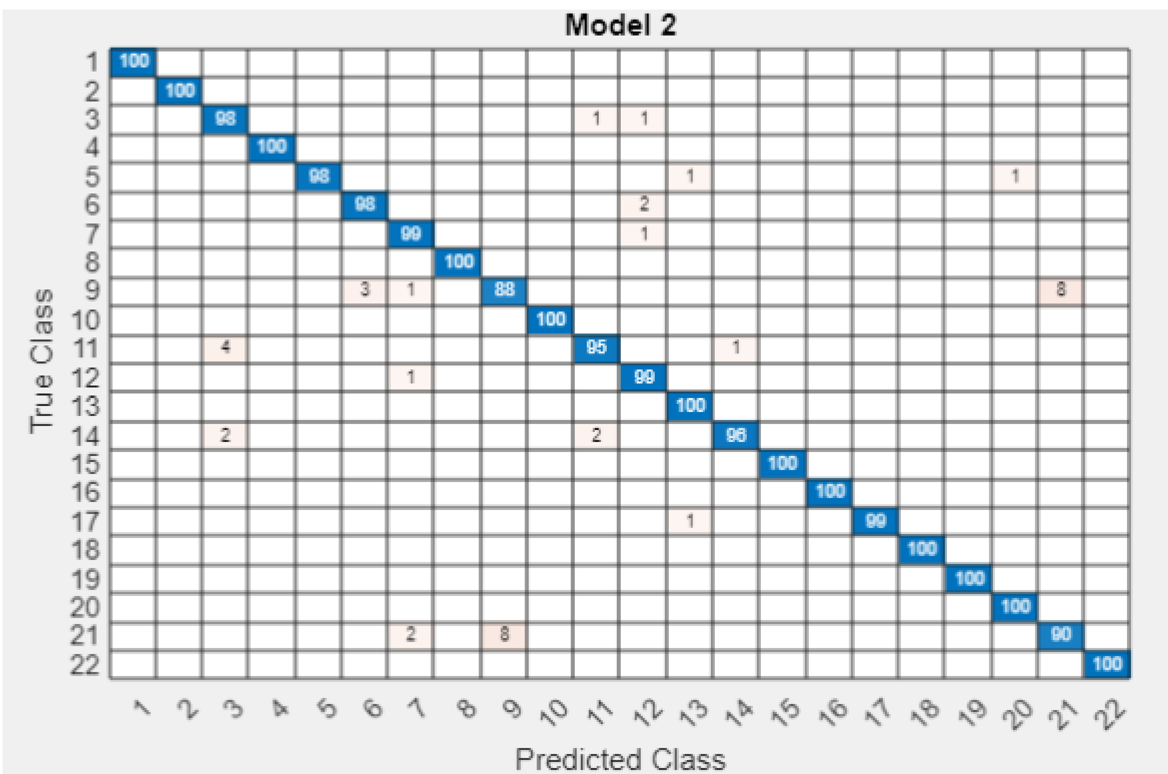

Table 6 confusion matrix experiment1

Table 7confusion matrix for class 6 (coffee)

TN 450	**FP 0**
FN 2	TP 98

Precision

Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives. Equation 5.7.1 shows the formula how to calculate precision.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5.7.1)$$

= <u>1</u>

Recall

Recall is the ability of a model to find all the relevant cases within a dataset. The precise definition of recall is the number of true positives divided by the number of true positives plus the number of false negatives. True positives are data point classified as positive by the model that actually are positive (meaning they are correct), and false negatives are data points the model identifies as negative that actually are positive (incorrect). Equation 5.7.2 shows the formula how to calculate recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5.7.2)$$

=<u>0.98</u>

F1 score

The F1 score is the harmonic mean of precision and recall taking both metrics into account in the following equation:

$$F1 = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (5.7.3)$$

$= \underline{0.98}$

Experiment2

Using PCA as feature and trees with number of split 100 following is the result

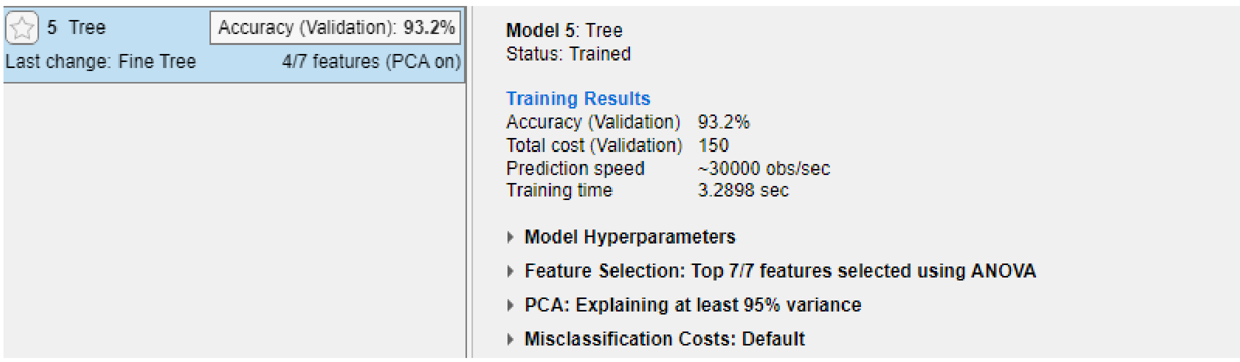

Figure 8 accuracy experiment2

The above figure shows the accuracy obtained using decision tree algorithm and PCA as a feature selector. And we obtained an accuracy of 93.2%.

Table 8 confusion matrix experiment2

TN 498	**FP 5**
FN 3	TP 22

Precision

Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5.7.4)$$

$= \underline{0.81}$

Recall

Recall is the ability of a model to find all the relevant cases within a dataset. The precise definition of recall is the number of true positives divided by the number of true positives plus the number of false negatives. True positives are data point classified as positive by the model that actually are positive (meaning they are correct), and false negatives are data points the model identifies as negative that actually are positive (incorrect). The equation 5.7.5 shows how recall is calculated;

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5.7.5)$$

$= 0\underline{.88}$

F1 score

The F1 score is the harmonic mean of precision and recall taking both metrics into account in the following equation5.7.6:

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5.7.6)$$

$= \underline{0.84}$

Experiment3

I modified experimental variables such as number of splitsand without using PCA I got an accuracy of 99.1%. It is shown in the figure20 below.

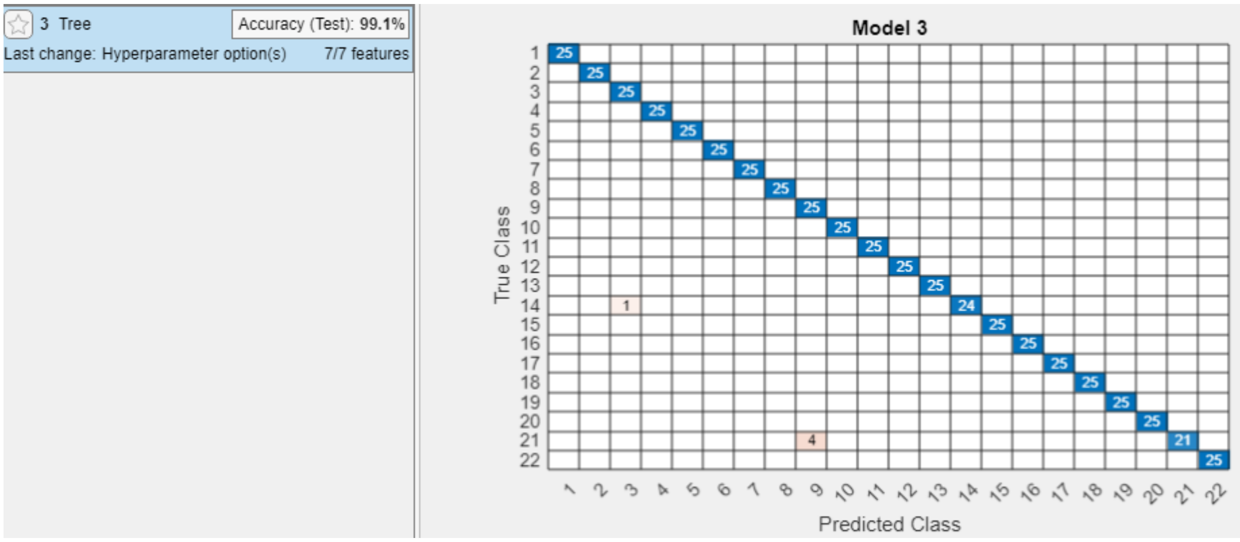

Figure 9 accuracy graph

5.8 Performance Evaluation

In this section, I evaluated the proposed crop prediction model using conventional performance evaluation methods. The evaluation includes those measures like accuracy and performance metrics. Most widely used standard metrics for measuring the performance of machine learning model is shown below.

Table 9 Confusion matrix for class14 (mothbeans)

		Actual Class	
		Negative class	Positive class
Predicted Class	Negative class	525	0
	Positive class	1	25

5.8.1 Performance Evaluation: Accuracy

As mentioned earlier the effectiveness of machine learning model is measured in terms of accuracy in which it identifies how much does it can predict the types of crop the soil grows. The accuracy of the proposed system is calculated using Equation 5.8.1.

$$\text{Accuracy} = \frac{TP+TN}{TN+TP+FP+FN} \quad (5.8.1)$$

In measuring the accuracy of machine learning module system I used 550 total test dataset from which I got TN=525, TP=24, FP=0 and FN=1 from this we can calculate the accuracy as 99.8%.

5.8.2 Performance Evaluation: Performance

The main objective of this performance evaluation is to identify whether it add noticeable overload or not to the ML model. In particular, the following measures will be used to assess the performance including accuracy

$$\text{False Positive Rate (FPR)} = \frac{FP}{TN+FP} \quad (5.8.2)$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad (5.8.3)$$

$$\text{Recall or True Positive Rate or Detection Rate (DR)= Recall}=\frac{TP}{TP+FN} \quad (5.8.4)$$

$$\text{F1score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5.8.5)$$

The other metrics which are useful to measure the performance of machine learning model are training time and testing time in which training time is the time needed by the algorithm to build the model and testing time is the time needed by the classifier to classify new example using a given model. Even if these issues are dependent on data set size doing it fast enough is important to avoid the delay. But the most commonly used are detection rate and false alarm rate. From this we can calculate each for detection approaches. For proposed approach I get FPR = 0.0, Precision = 1, Recall = 0.96 and F1score = 0.979.

5.8.3 Performance Evaluation: Completeness

The aim of this completeness is to minimize false alarm rate. When the false alarm rate decreases we can assume that the ML model will starts predicting types of crops that the soil will grow based on given environmental factors.

5.8.4 Performance Evaluation: Scalability

Scalability is the other issue which responsibly identify whether the ML model works in large dataset. And I have test it using dataset having 2,200 instances and we got accuracy of 99.8%. The proposed system easily accepts inputs in csv formats and predicts types of crops the soil supports to grow.

5.9 Discussion

The previous chapters explained the techniques I used how to make crop prediction machine learning model based on environmental factors. This paper goal is to achieve the objective listed in the chapter1 of this thesis. The objective of the study was to develop smart Machine Learning based crop prediction system using environmental and soil nutrient composition data.

I can say that this work have answered the research questions rose in this paper and achieved the objective of this research. Besides prediction of crop type further analysis on the grouping of crops based different factors has been done. For example we analyzed which groups of crops grow in various N and P content. The following figure shows the phenomena;

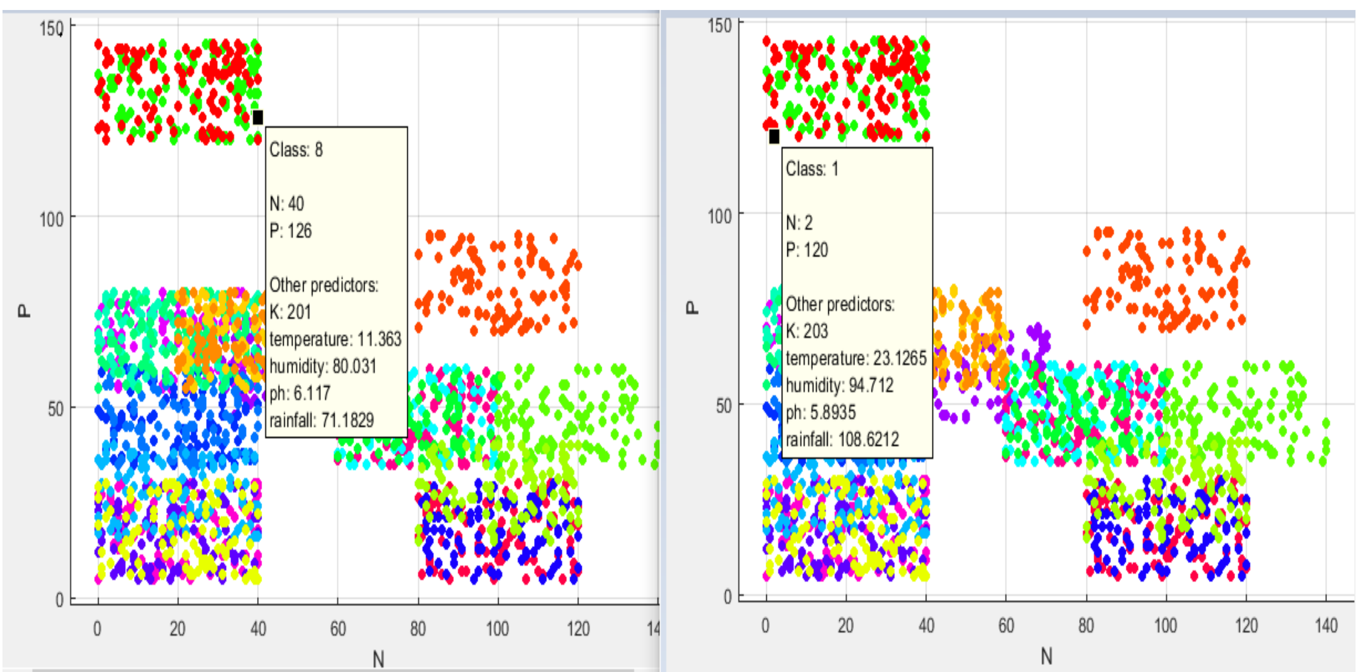

Figure 10 Crop distribution based on N and P.

From the figure above we can see that class 8(grapes) and class1(apple) has similar N and P requirement, which is high number of Phosphorus and lower level of Nitrogen. Similarly, crops in class 5 indicated with yellow color (coconut) and crop with purple color(orange fruit) needs low levels of Nitrogen and low Phosphorous.

The machine learning model was built using IoT dataset, which is collected from sensor nodes and web server and database server. The collected data is preprocessed and fed to decision tree algorithm to build the model. The built model was tested using 75% of dataset for training and 25% for testing. Based on testing result I have got an average accuracy value of 98.8%.

CHAPTER 6

Conclusion and Future work

6.1 Conclusion

Nowadays food shortage is eminent globally, even worse in African countries due to lack of mechanized farming. So using technology assisted farming promotes crop production in agriculture.

In this thesis I tried to develop Machine Learning model that predicts crop type for a given a soil based on environmental and nutrient condition of the soil. In addition of crop type prediction detail analysis has been done how certain group of crops grows with in similar environmental condition.

The data preprocessing and analysis has been done in Matlab software 2018a package. The built model has been tested with conventional performance measurement techniques and proven to be effective by providing high result. It is achieved an average accuracy of 98.8% with low false positive rate.

The main contribution of this thesis is it has achieved building a Machine Learning model that predicts crops type based on environmental and nutritional factors.

6.2 Future work

Even though this thesis result is promising and achieved its objective, it needs to be improved for the future. Future work that is continuation of this paper is listed below.

- This work is limited to predict only 22 types of crops, so further increasing number of crops is required.
- Increasing the number of dataset and collecting data outside Jimma area to make the model generalize the result to work for Ethiopia.

References

[1] F. Hussain, R. Hussain, S. A. Hassan, E. Hossain, and C. R. Mar, "Machine Learning in IoT Security : Current Solutions and Future Challenges," pp. 1–23.

[2] D. Kumar, K. Shen, B. Case, D. Garg, G. Alperovich, and D. Kuznetsov, "All Things Considered : An Analysis of IoT Devices on Home Networks," 2019.

[3] M. R. Shahid, G. Blanc, and Z. Zhang, "Machine Learning for IoT Network Monitoring".

[4] P. Caroline, E. Fiorenza, S. Sharma, and S. Ranjan, "SMART e-AGRICULTURE MONITORING BASED ON," *Int. J. Sci. Dev. Res.*, vol. 3, no. 10, pp. 223–227, 2018.

[5] V. N. Malavade and P. K. Akulwar, "Role of IoT in Agriculture," *Natl. Conf. "Changing Technol. Rural Dev.*, vol. 1, no. 13, pp. 422–425, 2019, [Online]. Available: www.iosrjournals.org

[6] R. Gómez-Chabla, K. Real-Avilés, C. Morán, P. Grijalva, and T. Recalde, "IoT Applications in Agriculture: A Systematic Literature Review," *Adv. Intell. Syst. Comput.*, vol. 901, pp. 68–76, 2019, doi: 10.1007/978-3-030-10728-4_8.

[7] F. Production *et al.*, *The future of food and agriculture: trends and challenges*, vol. 4, no. 4. 2014. [Online]. Available: www.fao.org/publications%0Ahttp://www.fao.org/3/a-i6583e.pdf%0Ahttp://siteresources.worldbank.org/INTARD/825826-1111044795683/20424536/Ag_ed_Africa.pdf%0Awww.fao.org/cfs%0Ahttp://www.jstor.org/stable/4356839%0Ahttps://ediss.uni-goettingen.de/bitstream/han

[8] N. C. Babu, "ENHANCING AGRICULTURAL PRODUCTIVITY WITH IOT," pp. 88–90.

[9] Novi yulia Budiarti, "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title," *Sustain.*, vol. 4, no. 1, pp. 1–9, 2020, [Online]. Available: https://pesquisa.bvsalud.org/portal/resource/en/mdl-20203177951%0Ahttp://dx.doi.org/10.1038/s41562-020-0887-

9%0Ahttp://dx.doi.org/10.1038/s41562-020-0884-z%0Ahttps://doi.org/10.1080/13669877.2020.1758193%0Ahttp://sersc.org/journals/index.php/IJAST/article

[10] "IoT use cases for agriculture Precisely monitor weather".

[11] D. J. Langley, J. van Doorn, I. C. L. Ng, S. Stieglitz, A. Lazovik, and A. Boonstra, "The Internet of Everything: Smart things and their impact on business models," *J. Bus. Res.*, vol. 122, no. December 2019, pp. 853–863, 2021, doi: 10.1016/j.jbusres.2019.12.035.

[12] G. Zerssa, D. Feyssa, D. G. Kim, and B. Eichler-Löbermann, "Challenges of smallholder farming in Ethiopia and opportunities by adopting climate-smart agriculture," *Agric.*, vol. 11, no. 3, pp. 1–26, 2021, doi: 10.3390/agriculture11030192.

[13] D. Summary, *Victoria ' s On-Farm Internet of Things Trial*, no. June. 2020.

[14] C. Industry and G. Partners, "Introduction to IoT in Agriculture".

[15] R. R. Shamshiri, J. W. Jones, K. R. Thorp, D. Ahmad, H. C. Man, and S. Taheri, "Review of optimum temperature, humidity, and vapour pressure deficit for microclimate evaluation and control in greenhouse cultivation of tomato: A review," *Int. Agrophysics*, vol. 32, no. 2, pp. 287–302, 2018, doi: 10.1515/intag-2017-0005.

[16] Y. Abbassi and H. Benlahmer, "The Internet of Things at the service of tomorrow's agriculture," *Procedia Comput. Sci.*, vol. 191, pp. 475–480, 2021, doi: 10.1016/j.procs.2021.07.060.

[17] S. S. Patil, "Internet-of-Things (IoT) Based Smart Agriculture," vol. 10, no. 5, 2020.

[18] A. H S*, D. Bhavani, A. Z. Fayaz, and A. Benny, "Smart Farming: IoT Based Water Managing System," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 4, pp. 2383–2385, 2020, doi: 10.35940/ijitee.d1796.029420.

[19] M. S. G. Kadam and M. P. B. Rajegore, "Internet of Things (IOT)," pp. 69–74.

[20] V. K. Quy *et al.*, “IoT-Enabled Smart Agriculture: Architecture, Applications, and Challenges,” *Appl. Sci.*, vol. 12, no. 7, 2022, doi: 10.3390/app12073396.

[21] V. Kumar and R. K. Chawda, “a Research Paper on Smart Home,” *Int. J. Eng. Appl. Sci. Technol.*, vol. 5, no. 3, pp. 530–532, 2020, doi: 10.33564/ijeast.2020.v05i03.088.

[22] C. Ave, “FT - NIR Analysis of Coffee,” pp. 1–3.

[23] S. Nagpure, S. Ingale, S. Pahurkar, A. M Bobade, M. Ghosal, and T. Dhope, “Smart Agriculture Using IOT,” *Helix*, vol. 9, no. 3, pp. 5081–5083, 2019, doi: 10.29042/2019-5081-5083.

[24] P. Tennevall and S. J. Hultgren, “Saving resources through smart farming - An IoT experiment study,” p. 55, 2019, [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:bth-17968

[25] D. Vasisht *et al.*, “Farmbeats: An IoT platform for data-driven agriculture,” *Proc. 14th USENIX Symp. Networked Syst. Des. Implementation, NSDI 2017*, pp. 515–529, 2017.

[26] E. J. A. Z. A. Hmed, I. B. Y. Aqoob, A. B. G. Ani, M. U. I. Mran, and M. O. G. Uizani, “N Etworking T Echnologies for the I Nternet of T Hings I Nternet - of -T Hings -B Ased S Mart E Nvironments : S Tate of the a Rt ,” vol. 12, no. October, pp. 10–16, 2016.

[27] V. Shankar, S. Kandagatla, K. Dorthi, V. C. S. Rao, P. Ankam, and N. Swathi, “Smart farming using IOT,” *AIP Conf. Proc.*, vol. 2418, no. 1, pp. 791–796, 2022, doi: 10.1063/5.0082563.

[28] A. Swami, A. Patil, and K. Kharade, “Applications of IoT for Smart Agriculture or Farming,” *Int. J. Res. Anal. Rev.*, vol. 6, no. 2, pp. 537–540, 2019.

[29] C. M. Swaraj and K. M. Sowmyashree, “IOT based Smart Agriculture Monitoring and Irrigation System,” *Int. J. Eng. Res. Technol.*, vol. 8, no. 14, pp. 245–249, 2020, [Online]. Available: www.ijert.org

[30] A. Srilakshmi, J. Rakkini, K. R. Sekar, and R. Manikandan, “A comparative study on

Internet of Things (IoT) and its applications in smart agriculture," *Pharmacogn. J.*, vol. 10, no. 2, pp. 260–264, 2018, doi: 10.5530/pj.2018.2.46.

[31] L. Battezzati, "Internet of Things Archives | Internet of Things," *Cisco*, no. July 2016, pp. 1–45, 2016, [Online]. Available: https://www.gsma.com/iot/search/internet-of-things/

[32] D. V. Gowda, S. M. Prabhu, M. Ramesha, J. M. Kudari, and A. Samal, "Smart Agriculture and Smart Farming using IoT Technology," *J. Phys. Conf. Ser.*, vol. 2089, no. 1, 2021, doi: 10.1088/1742-6596/2089/1/012038.

[33] V. V. Reddy, B. Sneha, M. Vinothini, R. V. Kritthika, and Y. Deepika, "Iot Based Smart Precision," *Eur. J. Mol. Clin. Med.*, vol. 7, no. 4, pp. 1443–1451, 2020.

[34] P. P. Ray, "Internet of things for smart agriculture: Technologies, practices and future direction," *J. Ambient Intell. Smart Environ.*, vol. 9, no. 4, pp. 395–420, 2017, doi: 10.3233/AIS-170440.

[35] A. Risdawati AP, C. Pertiwi, and A. Oktarina, "Integrated smart farming system in developing potential products of the village," *E3S Web Conf.*, vol. 306, p. 05014, 2021, doi: 10.1051/e3sconf/202130605014.

[36] H. Zia, A. Rehman, N. R. Harris, S. Fatima, and M. Khurram, "An experimental comparison of iot-based and traditional irrigation scheduling on a flood-irrigated subtropical lemon farm," *Sensors*, vol. 21, no. 12, 2021, doi: 10.3390/s21124175.

[37] S. Santiteerakul, A. Sopadang, K. Y. Tippayawong, and K. Tamvimol, "The role of smart technology in sustainable agriculture: A case study of wangree plant factory," *Sustain.*, vol. 12, no. 11, pp. 1–13, 2020, doi: 10.3390/su12114640.

[38] W. L. Hsu, W. K. Wang, W. H. Fan, Y. C. Shiau, M. L. Yang, and D. J. D. Lopez, "Application of internet of things in smart farm watering system," *Sensors Mater.*, vol. 33, no. 1, pp. 269–283, 2021, doi: 10.18494/SAM.2021.3164.

[39] M. Suchithra, T. Asuwini, M. C. Charumathi, and R. N. Lal, "Monitoring of Agricultural

Crops using Cloud and IOT with Sensor Data Validation," *Int. J. Pure Appl. Math.*, vol. 119, no. 12, pp. 14327–14335, 2018.

[40] J. Martis, "An Architectural design proposal for IoT in Agriculture," *Int. J. Sci. Res. Eng. Dev.*, vol. X, no. X, pp. 1–6, 2020, [Online]. Available: www.ijsred.com

Appendix 1

Machine learning training code in Matlab

```
function [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)
inputTable = trainingData;
predictorNames = {'N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall'};
predictors = inputTable(:, predictorNames);
response = inputTable.label;
isCategoricalPredictor = [false, false, false, false, false, false, false];

% Train a classifier
% This code specifies all the classifier options and trains the classifier.
classificationTree = fitctree(...
    predictors, ...
    response, ...
    'SplitCriterion', 'gdi', ...
    'MaxNumSplits', 100, ...
    'Surrogate', 'off', ...
    'ClassNames', [1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 21; 22]);

% Create the result struct with predict function
predictorExtractionFcn = @(t) t(:, predictorNames);
treePredictFcn = @(x) predict(classificationTree, x);
trainedClassifier.predictFcn = @(x) treePredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct
trainedClassifier.RequiredVariables = {'N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall'};
trainedClassifier.ClassificationTree = classificationTree;
trainedClassifier.About = 'This struct is a trained model exported from Classification Learner R2018a.';
```

```
trainedClassifier.HowToPredict = sprintf('To make predictions on a new table, T, use: \n  yfit = c.predictFcn(T) \nreplacing ''c'' with the name of the variable that is this struct, e.g. ''trainedModel''. \n \nThe table, T, must contain the variables returned by: \n c.RequiredVariables \nVariable formats (e.g. matrix/vector, datatype) must match the original training data. \nAdditional variables are ignored. \n \nFor more information, see <a href="matlab:helpview(fullfile(docroot, ''stats'', ''stats.map''), ''appclassification_exportmodeltoworkspace'')">How to predict using an exported model</a>.');
% Extract predictors and response
% This code processes the data into the right shape for training the
% model.
inputTable = trainingData;
predictorNames = {'N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall'};
predictors = inputTable(:, predictorNames);
response = inputTable.label;
isCategoricalPredictor = [false, false, false, false, false, false, false];

% Set up holdout validation
cvp = cvpartition(response, 'Holdout', 0.25);
trainingPredictors = predictors(cvp.training, :);
trainingResponse = response(cvp.training, :);
trainingIsCategoricalPredictor = isCategoricalPredictor;

% Train a classifier
% This code specifies all the classifier options and trains the classifier.
classificationTree = fitctree(...
    trainingPredictors, ...
    trainingResponse, ...
    'SplitCriterion', 'gdi', ...
    'MaxNumSplits', 100, ...
```

```
    'Surrogate', 'off', ...
    'ClassNames', [1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 21; 22]);
% Create the result struct with predict function
treePredictFcn = @(x) predict(classificationTree, x);
validationPredictFcn = @(x) treePredictFcn(x);
% Add additional fields to the result struct
% Compute validation predictions
validationPredictors = predictors(cvp.test, :);
validationResponse = response(cvp.test, :);
[validationPredictions, validationScores] = validationPredictFcn(validationPredictors);
% Compute validation accuracy
correctPredictions = (validationPredictions == validationResponse);
isMissing = isnan(validationResponse);
correctPredictions = correctPredictions(~isMissing);
validationAccuracy = sum(correctPredictions)/length(correctPredictions);
```

Appendix 2

Machine learning code for testing in Matlab

```
%%
tic;

s = RandStream('mt19937ar','Seed',20120214);

nKeepGene = 2201;
nKeepSnps = 57333;
nKeepCols = 840;

snps = SlicedData;
snps.dataSlices{1} = floor(s.rand( nKeepSnps, nKeepCols)*3);
snps.rowNameSlices{1} = cellstr(num2str((1:nKeepSnps)','Snp_%05d'));
snps.columnNames = cellstr(num2str((1:nKeepCols)','Sample_%03d'));

gene = SlicedData;
gene.dataSlices{1} = s.randn( nKeepGene, nKeepCols);
gene.rowNameSlices{1} = cellstr(num2str((1:nKeepGene)','Gene_%04d'));
gene.columnNames = cellstr(num2str((1:nKeepCols)','Sample_%03d'));

snps_aaa = snps.Clone();
snps_aaa.RowCenter();
[covariates, ~] = eigs((snps_aaa.dataSlices{1})'*snps_aaa.dataSlices{1},10);
clear snps_aaa;

% correct for randomness of eigs
covariates = bsxfun(@times, covariates, sign(covariates(1,:)));

% add random gene mean
gene.dataSlices{1} = bsxfun(@plus, gene.dataSlices{1}, s.randn(nKeepGene,1).^2);

% add signal to the first two genes

gene.dataSlices{1}(1,:) = snps.dataSlices{1}(1,:) + s.randn(1,nKeepCols);
gene.dataSlices{1}(2,:) = (snps.dataSlices{1}(2,:)==1) + s.randn(1,nKeepCols);

% add covariates to gene expression

gene.dataSlices{1} = gene.dataSlices{1} + s.randn(nKeepGene,10) * covariates';
```

```
clear s nKeepGene nKeepSnps nKeepCols;
toc;
```

```
%% Merlin
tic;
folder = 'Merlin';
if(~exist(folder,'dir'))
   mkdir(folder)
end;

[fid, msg] = fopen('Merlin\Testing.dat','w');
   if(fid==-1); error(msg); end;
   fprintf(fid,'C\tCvrt_%d\n',1:size(covariates,2));
   fprintf(fid,'T\t%s\n',gene.rowNameSlices{1}{:});
   fprintf(fid,'M\t%s\n',snps.rowNameSlices{1}{:});
fclose(fid);

[fid, msg] = fopen('Merlin\Testing.map','w');
   if(fid==-1); error(msg); end;
   z = [snps.rowNameSlices{1} num2cell((1:snps.nRows())') ]';
   fprintf(fid,'1\t%s\t%d\n',z{:});
fclose(fid);

toAC = [9 'A' 9 'A'; 9 'A' 9 'C';9 'C' 9 'C'];
[fid, msg] = fopen('Merlin\Testing.ped','w');
   if(fid==-1); error(msg); end;
   for s=1:gene.nCols
      fprintf(fid,'%i\t1\t0\t0\t1',s);
      fprintf(fid,'\t%f',covariates(s,:));
      fprintf(fid,'\t%f',gene.dataSlices{1}(:,s));
      z = toAC(snps.dataSlices{1}(:,s)+1,:);
      z = z';
      fprintf(fid,z(:));
      fprintf(fid,'\n');
   end;
fclose(fid);

clear ans fid msg folder toAC s z;
toc;
```

```
%% Plink
tic;
folder = 'Plink';
if(~exist(folder,'dir'))
```

```
    mkdir(folder)
end;

[fid, msg] = fopen('Plink\Testing.map','w');
    if(fid==-1); error(msg); end;
    z = [snps.rowNameSlices{1} num2cell((1:snps.nRows())') ]';
    fprintf(fid,'1\t%s\t0\t%d\n',z{:});
fclose(fid);

toAC = [9 'A' 9 'A'; 9 'A' 9 'C';9 'C' 9 'C'];
[fid, msg] = fopen('Plink\Testing.ped','w');
    if(fid==-1); error(msg); end;
    for s=1:gene.nCols
        fprintf(fid,'%i\t1\t0\t0\t1\t0',s);
        z = toAC(snps.dataSlices{1}(:,s)+1,:);
        z = z';
        fprintf(fid,z(:));
        fprintf(fid,'\n');
    end;
fclose(fid);

[fid, msg] = fopen('Plink\TestingPheno.txt','w');
    if(fid==-1); error(msg); end;
    for s=1:gene.nCols
        fprintf(fid,'%i\t1',s);
        fprintf(fid,'\t%f',gene.dataSlices{1}(:,s));
        fprintf(fid,'\n');
    end;
fclose(fid);

[fid, msg] = fopen('Plink\TestingCvrt.txt','w');
    if(fid==-1); error(msg); end;
    for s=1:gene.nCols
        fprintf(fid,'%i\t1',s);
        fprintf(fid,'\t%f',covariates(s,:));
        fprintf(fid,'\n');
    end;
fclose(fid);

clear fid ans folder msg toAC z s
toc;
```

```
%% Rqtl
tic;
```

```
folder = 'Rqtl';
if(~exist(folder,'dir'))
   mkdir(folder)
end;

[fid, msg] = fopen('Rqtl\Testing_genes.csv','w');
   if(fid==-1); error(msg); end;
   fprintf(fid,'id');
   fprintf(fid,',%d',1:gene.nCols);
   fprintf(fid,'\n');
   for i=1:gene.nRows
      fprintf(fid,'%s',gene.rowNameSlices{1}{i});
      fprintf(fid,',%f',gene.dataSlices{1}(i,:));
      fprintf(fid,'\n');
   end;
fclose(fid);

toAHB = [',A';',H';',B'];
zz = cell(3,snps.nRows);
zz(1,:) = snps.rowNameSlices{1};
zz(2,:) = num2cell(1:snps.nRows);
for i=1:snps.nRows
   z = toAHB(snps.dataSlices{1}(i,:)+1,:);
   z = z';
   zz{3,i} = z(:);
end;
[fid, msg] = fopen('Rqtl\Testing_snps.csv','w');
   if(fid==-1); error(msg); end;
   fprintf(fid,'id,,');
   fprintf(fid,',%d',1:gene.nCols);
   fprintf(fid,'\n');
   fprintf(fid,'%s,1,%d%s\n',zz{:});
fclose(fid);
clear z zz;

[fid, msg] = fopen('Rqtl\Testing_new_genes.csv','w');
   if(fid==-1); error(msg); end;
   fprintf(fid,'%s,',gene.rowNameSlices{1}{:});
   fprintf(fid,'id\n');
   if(fid==-1); error(msg); end;
   for s=1:gene.nCols
      fprintf(fid,'%f,',gene.dataSlices{1}(:,s));
      fprintf(fid,'%d',s);
      fprintf(fid,'\n');
```

```
    end;
fclose(fid);

[fid, msg] = fopen('Rqtl\Testing_new_snps.csv','w');
    if(fid==-1); error(msg); end;
    fprintf(fid,'id');
    fprintf(fid,',%s',snps.rowNameSlices{1}{:});
    fprintf(fid,'\n');
    fprintf(fid,',%d',1:snps.nRows());
    fprintf(fid,'\n');
    for s=1:gene.nCols
        fprintf(fid,'%d',s);
    %     fprintf(fid,'\t%s',toAC{snps.dataSlices{1}(:,s)+1});
        z = toAHB(snps.dataSlices{1}(:,s)+1,:);
        z = z';
        fprintf(fid,z(:));
        fprintf(fid,'\n');
    end;
fclose(fid);

save('Rqtl\Testing_cvrt.txt','covariates','-ascii','-tabs','-double');

clear ans folder fid msg i s toAC toAHB z
toc;
```

```
%% Matrix eQTL
tic;
folder = 'Matrix_eQTL';
if(~exist(folder,'dir'))
    mkdir(folder)
end;

[fid, msg] = fopen('Matrix_eQTL\Testing_genes.txt','w');
    if(fid==-1); error(msg); end;
    fprintf(fid,'id');
    fprintf(fid,'\t%s',gene.columnNames{:});
    fprintf(fid,'\n');
    for i=1:gene.nRows
        fprintf(fid,'%s',gene.rowNameSlices{1}{i});
        fprintf(fid,'\t%f',gene.dataSlices{1}(i,:));
        fprintf(fid,'\n');
    end;
fclose(fid);
```

```
toAC = [9 '0';9 '1';9 '2'];
   zz = cell(2,snps.nRows);
   zz(1,:) = snps.rowNameSlices{1};
   for i=1:snps.nRows
      z = toAC(snps.dataSlices{1}(i,:)+1,:);
      z = z';
      zz{2,i} = z(:);
   end;
   [fid, msg] = fopen('Matrix_eQTL\Testing_snps.txt','w');
   if(fid==-1); error(msg); end;
   fprintf(fid,'id');
   fprintf(fid,'\t%s',snps.columnNames{:});
   fprintf(fid,'\n');
   fprintf(fid,'%s%s\n',zz{:});
fclose(fid);

[fid, msg] = fopen('Matrix_eQTL\Testing_cvrt.txt','w');
   if(fid==-1); error(msg); end;
   fprintf(fid,'id');
   fprintf(fid,'\t%s',snps.columnNames{:});
   fprintf(fid,'\n');
   for i=1:size(covariates,2)
      fprintf(fid,'Cvrt_%d',i);
      fprintf(fid,'\t%d',covariates(:,i));
      fprintf(fid,'\n');
   end;
fclose(fid);

clear folder z zz ans fid msg i;
toc;
```

```
%%                                                          FastMap

tic;

folder = 'FastMap';

if(~exist(folder,'dir'))

   mkdir(folder)

end;


[fid, msg] = fopen('FastMap\Testing_genes.txt','w');

   if(fid==-1); error(msg); end;
```

```
    fprintf(fid,'id\tid');
    fprintf(fid,'\t%s',gene.columnNames{:});
    fprintf(fid,'\n');
    for i=1:gene.nRows
        fprintf(fid,'%s\t%s',gene.rowNameSlices{1}{i},gene.rowNameSlices{1}{i});
        fprintf(fid,'\t%f',gene.dataSlices{1}(i,:));
        fprintf(fid,'\n');
    end;
fclose(fid);

toAC = [9 '0';9 '1';9 '2'];
zz = cell(2,snps.nRows);
zz(1,:) = num2cell(1:snps.nRows);
for i=1:snps.nRows
    z = toAC(snps.dataSlices{1}(i,:)+1,:);
    z = z';
    zz{2,i} = z(:);
end;
[fid, msg] = fopen('FastMap\Testing_snps.txt','w');
    if(fid==-1); error(msg); end;
    fprintf(fid,'id\tid');
    fprintf(fid,'\t%s',snps.columnNames{:});
    fprintf(fid,'\n');
    fprintf(fid,'1\t%d\t%s\n',zz{:});
fclose(fid);

clear folder z zz ans fid msg i toAC;
toc
```