FOM Hochschule für Oekonomie & Management

Study Centre Essen

Master-Thesis

in degree program Big Data & Business Analytics

to obtain the degree of

Master of Science (M.Sc.)

on the subject

Effect of Artificial Intelligence on Agriculture Industry Transformation

from

Urvi Patel

Supervisor	Prof. Dr. Dirk Stein MBA
Matriculation number	586462
Date of submission	16th August 2022

Table of Contents

Index of Figures

List of Abbreviations

AI	Artificial Intelligence
AT&T	American Telephone and Telegraph Company
AWS	Amazon Web Services
CAGR	Compound Annual Growth Rate
CDO	Chief Data Officer
CIO	Chief Information Officer
CNN	Convolutional Neural Network
DARPA	Defense Advanced Research Projects Agency
DL	Deep Learning
DSRPAI	Dartmouth Summer Research Project on Artificial Intelligence
DSS	Decision Support Systems
DX	Digital Transformation
ETL	Extract, Transform, and Load
FGCP	Fifth Generation Computing Project
FNR	False Negative Rates
FPR	False Positive Rates
GPS	Global Positioning System
GPU	Graphics Processing Unit
GUI	Graphical User Interface
HTML	Hypertext Markup Language
ICT	Information and Communication Technologies
IoT	Internet of things
IT	Information Technology

JSON	JavaScript Object Notation
KPI	Key Performance Indicator
ML	Machine Learning
NAS	Neural Architecture Search
NLP	Natural Language Processing
NoSQL	Not Only SQL
R&D	Research and Development
ReLU	Rectified Linear Unit
RFID	Radio Frequency Identification
ROI	Return on Investment
SQL	Structured Query Language
UAV	Unmanned Aerial Vehicle
XML	Extensible Markup Language
YAML	Ain't Markup Language

1. Introduction

1.1. Importance of Topic

The world's population is increasing, and urbanization is accelerating. Consumption habits are shifting as disposable income rises. Farmers are under a lot of pressure to meet rising demand, and they're looking for a means to boost output. There will be more people to feed in thirty years. Because there is a finite amount of rich soil, there will be a need to expand beyond traditional farming.

There is a need to find ways to assist farmers in reducing their risks or at the very least, making them more manageable. Plant diseases pose a significant threat to the world economy, environment, consumers, and farmers. Pests and infections kill 35% of crops in India alone, causing farmers to lose a significant amount of money.[1] Yield of plant also decreases due to these diseases.

Plant disease symptoms can be seen in various sections of the plant; however, leaves are the most usually seen part for diagnosing an infection.[2] Crop leaf disease identification must be swift and precise to increase agricultural productivity in a sustainable manner.[3] One of the most intriguing potentials is to use artificial intelligence in agriculture to detect disease symptoms.

Artificial Intelligence (AI) in agriculture is projected to offer benefits such as reduced soil compaction and savings in production resources, labor expenses, and working hours.[4] AI can enhance currently installed technology as the next stage on the path from traditional to innovative farming.

---

[1] Cp. *Robin Sharma*., Artificial Intelligence, 2021, pp. 938.
[2] Cp. *Sukhvir Kaur, Shreelekha Pandey, and Shivani Goel*., Plants Disease, 2019, pp. 508.
[3] Cp. *Pallepati Vasavi, Arumugam Punitha, and T. Venkat Narayana Rao*., Disease Detection, 2022, pp. 2079.
[4] Cp. *Svenja Mohr and Rainer Kühl*., Artificial Intelligence in German Agriculture, 2021, pp. 1818.

1.2. Problem Statement

The problem statement derived from this introduction is that "Plant diseases pose a significant risk to the environment, the economy, and food security. For effective disease control, early detection of crop disease is critical."

1.3. Research Question

Based on the problem statement research questions could be formulated as follows:

RQ1. How can we detect the crop disease using AI?

RQ2. How can we predict the yield of crop based on the disease?

1.4. Planned Output

The expected outcome of this thesis is to do a feasibility study to know the prediction of crop yield based on crop damage.

1.5. Thesis Structure

This thesis follows the standard guidelines provided by the university. This paper starts with an introduction to the topic which provides a stage for the explanation of the topic followed by the problem statement and research questions which is then followed by the methodology of research used in summary. The second chapter of the theoretical foundation entails the background and literature of Deep Learning, Digital Transformation, Smart Farming and A.I. starting from the roots of big data and various types of deep learning algorithms as well. Later it follows to explain the different analytical methods for the different types of research methodology. The fourth chapter research process explains the concepts of A.I. in the Agriculture industry and its various application incurred from the literature review done from various articles and books based on certain criterion. This chapter also explains a case point of leaf disease detection through a Deep Learning algorithm using EfficientNetB3 model. The fifth chapter discusses the output and the insights of the thesis providing the summary of it followed by its limitation and the conclusion discussed in Chapter 6.

2. Theoretical Foundation

This chapter comprises the theoretical aspects used for the reference of this paper. The chapter explains the origin of A.I., Digital Transformation, Agriculture, Smart Farming, and Machine Learning from its roots in Big Data. Also, the chapter discusses the different analytical methods for Big Data and other types of Machine Learning algorithms and different kinds of AI and their evolution. This chapter also explains the various research methodologies used for this paper.

2.1. Definition of Big Data

Human society has entered an age of digital information due to tremendous advances in scientific technology. The amount of data collected in daily work and living has considerably expanded, ushering in the era of big data.[5] Big Data is a concept that has permeated our daily lives. Big data promises to solve some of the world's most complex challenges, from commercial applications to research in various sectors.[6] Big data is also common in most academic areas, including science, psychology, geography, humanities (now sometimes known as digital humanities), and healthcare.[7]

In recent years, computing technology has progressed, and the availability and popularity of social media and cell phones have created new sources for gathering research data.[8] The world has evolved into an "age of data" due to these technological breakthroughs.[9] Massive amounts of data can now be generated at any time and from any planet.[10] This is unquestionably true for scientists, particularly nursing scholars, whose study is no longer constrained by traditional data collection methods.[11]

What is data?

Quantities, letters, or symbols used by a computer can be stored and sent as electrical

---

[5] Cp. *Jane Hyatt Thorpe and Elizabeth Alexandra Gray*., Big Data, 2015, pp. 172.
[6] Cp. *Maddalena Favaretto et a*l., Definition of Big Data, 2020, no page number.
[7]Cp. *No Author.*, https://research.polyu.edu.hk/en/publications/the-opportunity-of-big-data-research-for-health-geographers-a-cas, accessed May 19, 2022.
[8] Cp. *Rob Kitchin.*, Big Data Challenges and Risks, 2013, pp. 264.
[9]Cp. *No Author.*, Big Data: How Data Analytics Is Transforming the... (PDF)," accessed May 19, 2022.
[10] Cp. *Viktor Mayer-Schönberger and Kenneth Cukier*., Big Data: A Revolution, 2014, pp. 1143.
[11] Cp. *Ho Ting Wong et al*., Disaster Preparedness, 2016, pp. 1015.

impulses and recorded on magnetic, optical, or mechanical media.[12]

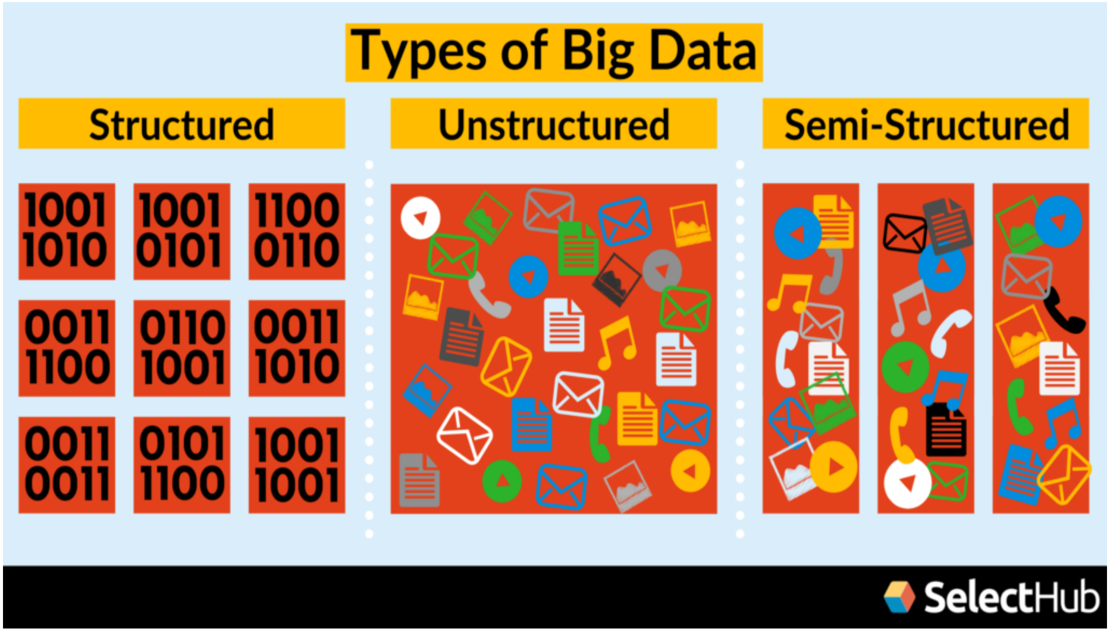
What is Big Data?

Big Data is a massive collection of data growing exponentially over time. Because data is so complex, no standard data management systems can effectively store or process it.[13] Big data is data, but it is data to a large extent.[14]

2.1.1 Types of Data

Every day, people generate 2.5 quintillion bytes of data. According to Statista, the internet will generate 74 Zettabytes (74 trillion GBs) of data by 2021.[15] Managing such tedious and ongoing data discoursing is becoming increasingly challenging; big data was created to collect such enormous and complex data; it is concerned with converting large and complex data into meaningful data that cannot be collected or analysed using traditional methods.[16]

Figure 1: Types of Big Data



Source: https://www.selecthub.com/big-data-analytics/types-of-big-data-analytics/

---

[12] Cp. *No Author*.," https://www.mathsisfun.com/data/data.html" Access on – 27-05-2022.
[13] Cp. *Andrea De Mauro, Marco Greco, and Michele Grimaldi*., What Is Big Data, 2015, pp. 97.
[14] Cp. *Andrea De Mauro, Marco Greco, and Michele Grimaldi*., Big Data, 2016, pp. 125.
[15] Cp. *Liam J. Caffery et al.*, Transforming Dermatologic, 2018, pp. 570.
[16] Cp. *Mohammad Fikry Abdullah and Kamsuriah Ahmad*., Unstructured Data to Structured Data, 2013, pp. 153.

It is impossible to save all data in the same way. After the type of data has been recognized, the methods for data storage may be accurately examined.[17] A Cloud Service, such as Microsoft Azure, provides a one-stop-shop for all data types, including blobs, queues, files, tables, desks, and application data. However, there are unique services for dealing with various data sub-categories even within the Cloud.[18]

Big data is classified in three ways:

- Structured Data
- Semi-Structured Data
- Unstructured Data

These three terms are popular in big data, even though they are technically relevant at all levels of analytics.[19] When working with large amounts of data, it's even more important to understand where the raw data originates and how it needs to be processed before being analysed. Because there is so much of it, data extraction must be efficient for the project to be worthwhile.[20]

The data structure is crucial to understanding how to work with it and what insights it can yield. Before being processed, all data goes through an extract, transform, and load (ETL) procedure.[21] It's a reasonably literal term: data is collected, structured so that an application can read it, and then stored for later use. Each data structure requires a different ETL technique.[22]

I. Structured Data

Data in a specific field within a record can be informally characterised as structured data. It is bound by a particular schema, which ensures that all data has the same features. Relational data is another name for structured data.[23] It is divided into many

---

[17] Cp. *Ibid*.
[18] Cp. *Ibid*.
[19] Cp. *Arvind Arasu and Hector Garcia-Molina.*, Structured Data, 2003, no page number.
[20] Cp. *Tomasz Wiktorski, Yuri Demchenko, and Oleg Chertov*., Big Data Infrastructure, 2019, pp. 541–47.
[21] Cp. *Alkis Simitsis, Panos Vassiliadis, and Timos Sellis*., ETL Processes, 2005, pp. 567.
[22] Cp. *Ibid*.
[23] Cp. *No Author.*, https://eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=3&sid=aaf1d93e-7777-44f8-b739-f40b767a2011%40redis, accessed May 28, 2022.

tables to improve the data's integrity by producing a single record representing each entity. Relationships are enforced by using table constraints.[24]

The data must be brought together using Structured Query Language (SQL). Data that is structured is simple to enter, query, and analyse.[25] The data is all formatted in the same way. However, mandating a consistent structure means that any data change will be too harsh, as each record will need to be modified to conform to the new system.

Figure 2: Structured Data Table

CUSTOMER

CUSTOMER_ID	LAST_NAME	FIRST_NAME	STREET	CITY	ZIP_CODE	COUNTRY
10302	Boucher	Leo	54, rue Royale	Nantes	44000	France
11244	Smith	Laurent	8489 Strong St	Las Vegas	83030	USA
11405	Han	James	636 St Kilda Road	Sydney	3004	Australia
11993	Mueller	Tomas	Berliner Weg 15	Tamm	71732	Germany
12111	Carter	Nataly	5 Tomahawk	Los Angeles	90006	USA
14121	Cortez	Nola	Av. Grande, 86	Madrid	28034	Spain
14400	Brown	Frank	165 S 7th St	Chester	33134	USA
14578	Wilson	Sarah	Seestreet #6101	Emory	1734	USA

Source: https://www.michael-gramlich.com/what-is-structured-semi-structured-

Numbers, dates, strings, and other types of organised data are examples. An e-commerce website's business data can be structured data.[26]

Cons of Structured Data

1. Structured data can only be used when established functionalities are required. Structured data, as a result, has limited flexibility and is only appropriate for specific use cases.[27]

2. Structured data is kept in a data warehouse with strict constraints and a well-defined structure.[28] Any change in requirements would necessitate changing all of that structured data to accommodate the new needs. In terms of resource and time management, this is a significant disadvantage.[29]

---

[24] Cp. *Syed Iftikhar Hussain Shah, Vassilios Peristeras, and Ioannis Magnisalis.*, Ecosystem, 2021, no page number.
[25] Cp. *Xindong Wu et al*., Data Mining, 2014, pp. 100.
[26] Cp. *Wiktorski, Demchenko, and Chertov*,, Big Data Infrastructure, no page number.
[27] Cp. *Arvind Arasu and Hector Garcia-Molina.*, Structured Data, 2003, no page number.
[28] Cp. *Ibid.*
[29] Cp. *Ibid.*

II. Semi-Structured Data

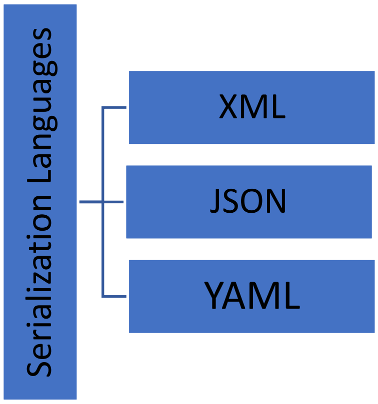
Any schema does not bind semi-structured data for data storage and handling. The information neither in a relational structure nor orderly sorted into rows and columns as it would be in a spare however, other properties, such as key-value pairs, aid in distinguishing between the various entities.[30] Semi-structured data is referred to as NoSQL data since it does not require a structured query language.[31] A data serialization language is used to transport semi-structured data across systems with the different underlying technology.[32]

Data is written in plain text so that various text-editing tools can be used to extract useful information.[33] Data serialization readers can be implemented on devices with minimal processing resources and bandwidth due to their simple format.

Data Serialization Languages

Software developers use serialization languages to write memory-based data to files, transport it, store it, and parse it.[34] The sender and recipient are not required to be aware of the other system. Both systems can quickly grasp the data using the same serialization language. Three serialization languages are widely used.[35]

Figure 3: Types of Data Serialization languages



Source: Author Own Representation

---

[30] Cp. *G. Suseendran et al*., Semi Structured Data, 2021, no page number.
[31] Cp. *Lin Jiang and Zhijia Zhao*., Streaming, 2022, pp. 12.
[32] Cp. *Maria A. Poltavtseva, Pavel V. Semyanov, and Elizaveta A. Zaitzeva*., Information Security, 2020, no page number.
[33] Cp. *Rosa Maria Canton-Croda and Damian Emilio Gibaja-Romero*., Transformation, 2020, pp. 119.
[34] Cp. *Dirk Eddelbuettel, Murray Stokely, and Jeroen Ooms*., Serialization, 2016, pp. 7.
[35] Cp. *Jan Newmarch*., *Network Programming*, 2017, pp. 60.

1. XML

eXtensible Markup Language is the abbreviation for eXtensible Markup Language. It's a text-based markup language for storing and transmitting data.[36] XML parsers are available in nearly every major development platform. Both humans and machines can read it. For schema, transformation, and presentation, XML has specific standards. It's self-explanatory.[37] An XML example of a programmer's information is shown below.

Example:

```
<ProgrammerDetails>
      <FirstName>Jane</FirstName>
      <LastName>Doe</LastName>
      <CodingPlatforms>
            <CodingPlatform Type="Fav">W3School</CodingPlatform>
            <CodingPlatform Type="2ndFav">Code4Eva!</CodingPlatform>
            <CodingPlatform Type="3rdFav">CodeisLife</CodingPlatform>
</CodingPlatforms>
</ProgrammerDetails>
```

XML uses tags (text enclosed in angular brackets) to structure the data (for example, FirstName) and attributes (for example, Type) to highlight the data.[38] Other forms have acquired more popularity due to the language's verbose and voluminous nature.[39]

2. JSON

JSON (JavaScript Object Notation) is a simple open-standard data interchange format.[40] JSON is simple and stores and transmits data items as human/machine-readable language.[41]

---

[36] Cp., *No Author.*, http://www.w3.org/TR/1998/REC-xml-19980210, accessed May 28, 2022.
[37] Cp. *Ibid.*
[38] Cp. *Ibid.*
[39] Cp. *Mary Fernandez, Dan Suciu, and Chun-Nan Hsu.*, http://www.w3.org/TR/1998/NOTE-xml-ql-19980819http://www.w3.org/TR/1998/NOTE-xml-ql-19980819.htmlLatestversion:http://www.w3.org/TR/NOTE-xml-ql,accessed May 28, 2022,.
[40] Cp. *Felipe Pezoa et al.*, JSON, 2016, pp. 267.
[41] Cp. *Ibid.*

Example:

```
{
      "firstName": "Jane",
      "lastName": "Doe",
      "codingPlatforms": [
            { "type": "Fav", "value": "Geeksforgeeks" },
            { "type": "2ndFav", "value": "Code4Eva!" },
            { "type": "3rdFav", "value": "CodeisLife" }
      ]
}
```

This is a less formal format than XML. It resembles a key/value pair approach rather than a typical data representation. [42]JSON is supported by default in JavaScript. Despite its popularity among web developers, non-technical employees find working with JSON unpleasant due to its firm reliance on JavaScript and structural elements (braces, commas, etc.)[43]

3. YAML

YAML is a data serialization language that is simple to use.[44] It stands for YAML Isn't Markup Language in a figurative sense. Because of its simplicity, it is used by both technical and non-technical handlers worldwide.[45] Line separation and indentation establish the data structure, which lowers the reliance on structural characters. YAML is a very comprehensive language, and its appeal stems from the fact that it is human-machine readable.[46]

Example:

Personal data stored in an XML file-

```
<rec><name>Prashant Rao</name><sex>Male</sex><age>35</age></rec>
```

---

[42] Cp. *Guanhua Wang.*, Web Applications, 2011, pp. 182.
[43] Cp. *Pierre Bourhis et al.*, Data Model, 2017, pp. 124.
[44] Cp. *Oren Ben-Kiki et al*., YAML, 2001, pp. 19.
[45] Cp. *Ibid*.
[46] Cp. *V. Sinha et al*., Visualization, 2000, pp. 11.

```
<rec><name>Seema R.</name><sex>Female</sex><age>41</age></rec>

<rec><name>Satish Mane</name><sex>Male</sex><age>29</age></rec>

<rec><name>Subrato Roy</name><sex>Male</sex><age>26</age></rec>

<rec><name>Jeremiah J.</name><sex>Male</sex><age>35</age></rec>
```

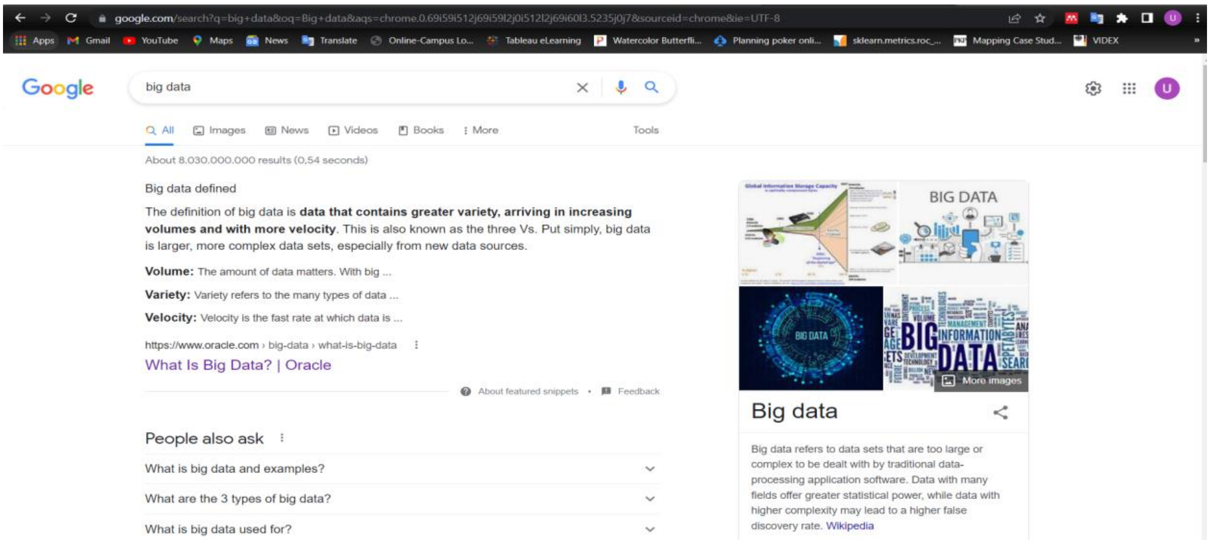
III. Unstructured Data

Unstructured data is any data that has an undetermined shape or organization.[47] Unstructured data, in addition to its enormous bulk, poses several processing issues to extract value from it.[48] A good example is a heterogeneous data source including a mix of simple text files, photos, videos, and other types of unstructured data. Organizations nowadays have a plethora of data at their disposal, but they don't know how to extract value because the data is in its raw or unstructured format.[49]

Example:

The output returned by 'Google Search.'

Figure 4: Google Search Results



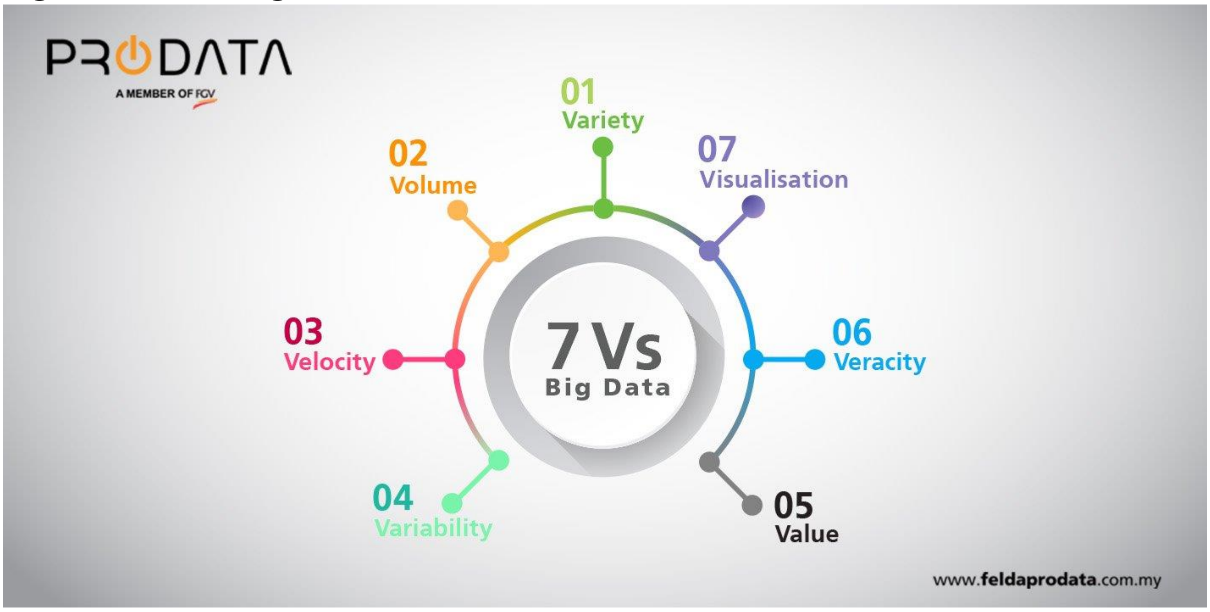
Source: Author Own Representation

---

[47] Cp. *Peter Buneman et al*., Unstructured Data, 1996, pp.336.
[48]Cp.*NoAuthor*.,https://books.google.de/books?hl=en&lr=&id=BEDTrvUH8NcC&oi=fnd&pg=PA243&dq=unstructured+data&ots=WySzfTLIK8&sig=17kARegGpQeTZAgdjXg4E56vj88&redir_esc=y#v=onepage&q=unstructured data&f=false, accessed May 28, 2022.
[49] Cp. *Adanma Cecilia Eberendu*., Unstructured Data, 2016, no page number.

2.1.2 The 7V’s of Big Data

Figure 5: 7V's of Big Data



Source: https://twitter.com/fgvprodata/status/1090909343850082304

Big data is a difficulty not only because of its size but also because of the numerous additional elements that must be considered.[50] There are five generally discussed Vs. of Big Data, but we may add two additional Vs. that can be viewed as a factor in the Big Data aspect.[51]

The 7 Vs. of big data are the following:

Volume: The Volume, as the name implies, deals with massive amounts of data created by a variety of sources such as text, audio, photos, videos, and so on. For example, this massive amount of data could be generated by social media chatter, health sensors, web server logs, financial markets, and so on.[52] We need specific and strong tools to query and modify such a large amount of data that cannot be queried and edited in typical ways. Kafka, Hadoop, and others are examples of these tools.[53]

Velocity: Underlying the volume figures is a wider trend: 90% of all existing data has been created in the last two years. The rate at which data is generated, gathered, and

---

[50] Cp. *Seon-Cheol Yu, Dong-Bin Shin, and Jong-Wook Ah*n., Big Data, 2016, pp. 2893.
[51] Cp. *Ibid*.
[52] Cp. *Yogesh Kumar Gupta and Surbhi Kumari*., Big Data Analytics, 2020, pp. 474.
[53] Cp. *Nawsher Khan et al*., Big Data: Survey, 2014, no page number.

processed is increasing rapidly. Globally, there will be 19 billion network connections feeding this velocity by next year.[54]

Although most data are stored before analysis, there is a growing demand for real-time processing of massive volumes like the 200 million emails, 300,000 tweets, and 100 hours of YouTube videos that are sent every minute of the day.[55] Real-time processing lowers storage requirements while increasing responsiveness, accuracy, and profitability.

Variety: Another issue of Big Data processing is handling the large variety of data, which is in addition to the massive volumes and growing velocities of data. When seen as a whole, these data appear to be an incomprehensible blob with no discernible pattern. The extraction of meaning from such diversity, which includes natural language, hashtags, geospatial data, multimedia, sensor events, and much more, necessitates ever-increasing algorithmic and computational capability.[56]

Variability: Furthermore, the inherent meanings and interpretations of these jumbles of raw data are contextually dependent. This is especially true when it comes to processing natural language. A single word might have several different meanings.[57] Over time, new meanings emerge, and old meanings fade away. When it comes to evaluating and responding to social media buzz, for example, understanding implications is critical. Because of its infinite diversity, Big Data presents a unique decoding problem if its full worth is to be realized.

Veracity: The scientific community used to believe that the data coming in was clean and exact. As a result of the increased generation of data, veracity was introduced. This calls into question the data's veracity, as well as the significance of the conclusions drawn from the analysis. The issue we're supposed to be asking is whether we can trust all of the data that comes in. To address this issue, we normalize data in relational and conventional databases, ensuring that data integrity is maintained.[58]

Visualization: One of the most important tasks for any Big Data processing system is to

---

[54] Cp. *Thibaud Chardonnens et al*., High Velocity, 2013, pp. 785.
[55] Cp. *Ibid*.
[56] Cp. *Sören Auer et al*., Dimension of Big Data, 2017, pp. 50.
[57] Cp. *Abou Zakaria Faroukhi et al.*, Big Data, 2022, no page number.
[58] Cp. *Ibid*.

reduce the enormous scale of the data to something that can be simply understood and used. Converting data into graphical formats is one of the finest strategies for human consumption. Due to the properties of pace and variety, spreadsheets and even three-dimensional visualizations are frequently inadequate.[59] To compress into visual shapes, there may be a plethora of spatial and temporal dimensions, as well as interactions between them. The major motivation behind AT&T's Nanotubes visual representation package is to solve these challenges.

Value: The data's worth is always unique for a reason. Unlike the other Vs of big data, value corresponds to the outcome of data processing, i.e., how useful are the insights derived from the data. In this case, the value of the data should always outweigh the cost of data management and storage. Data frequently moves across value tiers. The higher the tier, the more valuable the data is, and vice versa. The value of data can be determined by a variety of criteria that must be considered by organizations processing the data.[60]

2.1.3 Problems of Big Data

The main implementation difficulties in Big Data are the challenges. These require quick attention and must be dealt with because if they are not dealt with, the technology may fail, resulting in some unpleasant outcomes. The problems of big data include storing and processing extraordinarily vast and rapidly rising data sets.[61]

Some of the Big Data challenges are:

1. **Sharing and Accessing Data:**

• The inaccessibility of data sets from external sources is perhaps the most common obstacle un big data endeavors.

• Data sharing can be quite difficult.

• It involves the requirement for legal documents that are both inter- and intra-institutional.

• Getting data from public repositories is problematic for a variety of reasons.[62]

---

[59] Cp. *Seon-Cheol Yu, Dong-Bin Shin, and Jong-Wook Ah*n., Big Data, 2016, pp. 2893.
[60] Cp. *Yogesh Kumar Gupta and Surbhi Kumari*., Big Data Analytics, 2020, pp. 474.
[61] Cp. *Evgeniy N. Pavlovskiy*., Prospectives of Big Data, 2019, pp. 995.
[62] Cp. *Zongben Xu and Yong Shi*., Big Data Analysis, 2015, pp. 367.

• It is vital for data to be provided in an accurate, full, and timely manner because data in the company's information system will be used to make informed decisions in a timely manner.

2. Privacy and Security:

• Big Data is another significant challenge. This problem has sensitive, philosophical, technical, and legal implications.[63]

• Due to the vast volume of data generated, most firms are unable to maintain regular checks. However, doing security checks and observations in real time should be required because it is the most useful.

• There is some information about a person that, when combined with external vast data, may lead to certain secretive facts about that person, and he may not want the owner to know about it.

• To bring value to their business, several organizations collect information about people. This is accomplished by providing them with hitherto unseen insights into their lives.[64]

3. Analytical Challenges:

• Big data presents a number of significant analytical issues, including how to deal with a problem when data volume grows too enormous.

• Or how do you figure out what the most important data points are?

• Or how to get the most out of data?

• The huge amount of data to be analyzed can be structured (organized data), semi-structured (semi-organized data), or unstructured (unstructured data) (unorganized data). There are two methods for making decisions: either incorporate enormous data volumes into the research or establish whether big data is relevant ahead of time.[65]

---

[63] Cp. *Alfredo Cuzzocrea and Ernesto Damiani.*, Big Data Exchange, 2021, pp. 5082.
[64] Cp. *Ibid.*
[65] Cp. *Evgeniy Yur Evich Gorodov and Vasiliy Vasil Evich Gubarev.*, Data Visualization, 2013, no page number.

4. **Technical challenges:**

- **Quality of data:**

  - There is a cost associated with collecting a big volume of data and storing that data. Large data storage is constantly desired by big enterprises, business leaders, and IT professionals.[66]
  - Big data, rather than having irrelevant data, focus on quality data storage for better findings and conclusions.
  - This raises the dilemma of how to verify that data is useful, how much data is sufficient for decision-making, and whether or not the data saved is correct.

- **Fault tolerance:**
  - Another technical challenge is fault tolerance, and fault tolerance computing is exceedingly difficult, requiring complex algorithms.
  - Nowadays, some of the latest technologies, such as cloud computing and big data, are designed to ensure that when a failure occurs, the damage caused is within an acceptable range, and the task does not have to be restarted from the beginning.[67]

- **Scalability:**

  • Big data projects can quickly develop and evolve. Cloud computing has emerged as a solution to Big Data's scalability problem.[68]

  • It raises a number of issues, such as how to operate and execute numerous jobs so that the goals of each workload can be met in a cost-effective manner.

  • It also necessitates an efficient response to system breakdowns. This raises the issue of what kind of storage devices should be used.

2.2 Analytical methods for Big Data Analytics

Although there are numerous ways to analyze big data the most common ones that are frequently used are Descriptive, Predictive, and Prescriptive.[69] The Descriptive method

---

[66] Cp. *H. V. Jagadish et al*., Technical Challenges, 2014, pp. 89.
[67] Cp. *Ibid*.
[68] Cp. *Ibid*.
[69] Cp. *Adeleh Asemi et al*., Recommender Systems, 2022, pp. 15.

describes the data and Predictive method predicts future events based on the data and the Prescriptive method prescribes possible actions that can be taken on the data.[70] The desired type of analysis is often chosen by the problem at hand.

2.2.1. Descriptive Analytics

This is the most fundamental form of analytics for big data that describes the data at a tedious level. This form of analytics discusses the key performance indicators (KPIs) of the data. Moreover, the analysis is straightforward and does not require complex machine learning algorithms and advanced statistical techniques. This analysis can be done using normal programming languages like Python, R, etc. MS Excel and Tableau are the tools that can be used to generate reports on this analysis.[71] There are various techniques described in descriptive statistics such as Measures of central tendency (Mean, Median and Mode), Measure of Position (Percentiles), and Measure of Variation (Standard Deviation, Variance, and Skewness).

Figure 6: Representation of Descriptive Statistics

	Hour	Class	Min	Max	Transactions	Sum	Mean	Median	Var
0	0.0	0	0.0	7712.43	3961	256572.87	64.774772	12.990	45615.821201
1	0.0	1	0.0	529.00	2	529.00	264.500000	264.500	139920.500000
2	1.0	0	0.0	1769.69	2215	145806.76	65.826980	22.820	20053.615770
3	1.0	1	59.0	239.93	2	298.93	149.465000	149.465	16367.832450
4	2.0	0	0.0	4002.88	1555	106989.39	68.803466	17.900	45355.430437

Source: Author Own Representation

2.2.2 Predictive Analytics

Although the descriptive methods describe the data the predictive methods draw from the information to predict future events with a certain probability. This empowers companies to optimize their profit and loss which means to say that they can increase their risk to reward ratio. Different metrics are used to perform predictive analytics such as False Positive Rates (FPR), False Negative Rates (FNR), R-Square, Odds Ratio, Recall, Precision, etc.[72]

---

[70] Cp. *Ibid*.
[71] Cp. *Harkiran Kaur and Aanchal Phutela.*, Descriptive Data Analytics, 2018, pp. 679.
[72] Cp. *Ibid*.

Figure 7: Representation of Metrics for Predictive Analysis

	precision	recall	f1-score	support
No Fraud	1.00	0.99	0.99	56863
Fraud	0.10	0.86	0.19	98
accuracy			0.99	56961
macro avg	0.55	0.92	0.59	56961
weighted avg	1.00	0.99	0.99	56961

Source: Author Own Representation

Complex algorithms like Time Series analysis, Deep Learning, etc. are used to perform the predictive analysis. To carry out these algorithms one needs to have good knowledge of statistical methods and requires by and large good computational power and business knowledge.[73]

There are several other applications for predictive analytics such as time series analysis can predict the price of the stocks listed on the stock exchange.[74] Predictive analytics can be performed using programming languages such as Python, R, etc. Several other tools such as SPSS can be used to build predictive models. There are several other algorithms such as Linear and Logistic Regression and several packages of Python such as Sci-Kit Learn to aid in predictive analysis.

2.2.3 Prescriptive Analytics

According to Frazzetto, D. et.al. Prescriptive analytics is commonly used for the decision-making of the data.[75] This method can be used to take steps in the right direction and achieve the desired result ensuing in the generation of a better return on the investment (ROI). The prescriptive analysis derives from predictive analytics and concludes to make the right decision to optimize the working or the process efficiency of the company.[76]

---

[73] Cp. *Galit Shmueli and Otto R. Koppius*., Predictive Analytics, 2011, pp. 556.
[74] Cp. Charles Nyce., www.aicpcu.org, accessed May 29, 2022.
[75] Cp. *Sa Kwang Song et al*., Prescriptive Analytics, 2013, pp. 1144.
[76] Cp. *Reza Soltanpoor and Timos Selli*s., Prescriptive Analytics for Big Data, 2016, pp. 250.

2.3 Definition of Transformation

2.3.1. Definition of Digital Transformation

The adoption of digital technologies by an organization is known as digital transformation (DX).[77]

To meet changing business and market requirements, digital transformation is the process of employing digital technology to build new — or adapt current — business processes, culture, and consumer experiences.[78] Digital transformation is the redesigning of business in the digital age.

According to Deloitte, “digital transformation is all about becoming a digital enterprise—an organization that uses technology to continuously evolve all aspects of its business models (what it offers, how it interacts with customers, and how it operates).”

Digital transformation entails experimenting with new technology and rethinking your present approach to common problems in order to evolve your firm.[79] A transition does not always have a clear endpoint because it is an evolution.[80] "Digital transformation is better understood of as perpetual adaptation to a constantly changing environment," according to the MIT Sloan Management Review, a publication that focuses on how management transforms in the digital era.

Digital transformation is the intentional and prioritized transformation of company and organizational activities, processes, competences, and models to fully harness the changes and opportunities of a mix of digital technologies and their growing influence across society.[81]

Digital transformation can take many different forms, and each company's journey will be different. For example, a corporation might employ AI or cloud computing to improve its client experience, or it might revamp its supply chain to include machine learning more effectively. A corporation can even estimate what products customers will want in a few

---

[77] Cp. *Christof Ebert, Carlos Henrique, and Cabral Duarte.*, Digital Transformation, 2018, no page number.
[78] Cp. *Ibid*.
[79] Cp. *Haluk Demirkan, James C. Spohrer, and Jeffrey J. Welser.*, Digital Innovation, 2016, pp. 16.
[80] Cp. *Ibid*.
[81] Cp. *Gregory Vial.*, Digital Transformation, 2019, no page number.

months and change manufacturing to match demand.[82]

Starting a digital transformation journey, in any case, necessitates a shift in mentality. It's an opportunity to rethink how businesses operate, frequently from the ground up.[83]

Two concepts related to digital transformation are digitization and digitalization.[84]

Digitization:

Digitization is the act of converting analog data and information into digital form, such as scanning a photograph or document and putting it on a computer.[85]

Digitalization:

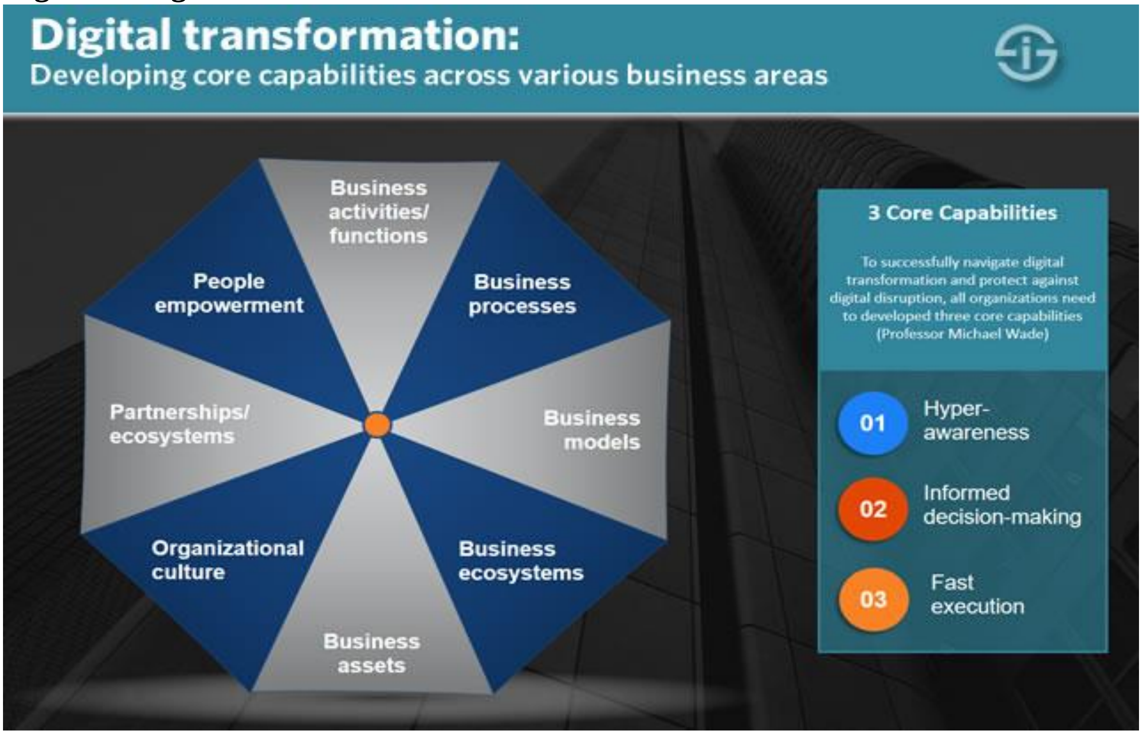
Digitalization is the application of digital technologies to alter company processes and projects, such as training people to use new software platforms that enable companies launch goods more quickly.[86] While digital transformation may entail digitalization activities, it affects the entire business and extends beyond the project level.

Digital Transformation Areas: Digital transformation in the comprehensive and interconnected meaning that it necessitates might include, among other things, the transformation of:[87]

---

[82] Cp., *No Author.*, https://www.techtarget.com/searchcio/definition/digital-transformation, accessed May 29, 2022.
[83] Cp. *K Schwertner.*, DIGITAL TRANSFORMATION, 2017, pp. 391.
[84] Cp., *No Author.*, https://www.salesforce.com/eu/products/platform/what-is-digital-transformation/ accessed May 29, 2022.
[85] Cp. *Jason Bloomberg Contributor.*, https://www.forbes.com/sites/jasonbloomberg/2018/04/29/digitization-digitalization-and-digital-transformation-confuse-them-at-your-peril/#78e677fd2f2c, accessed May 29, 2022.
[86] Cp. *Johan Hagberg, Malin Sundstrom, and Niklas Egels-Zandén.*, Digitalization, 2016, pp. 700.
[87] Cp., *No Author.*, https://www.i-scoop.eu/digital-transformation/, accessed May 29, 2022.

Figure 8: Digital Transformation Areas



Source: https://www.i-scoop.eu/digital-transformation/

- **Business activities/functions**: marketing, operations, human resources, administration, customer service, etc.[88]

- **Business processes**: Business process management, business process optimization, and company process automation are all used to achieve a specific business goal by connecting one or more associated operations, activities, and sets (with new technologies such as robotic process automation).[89] In today's digital transformation initiatives, business process optimization is critical, and most sectors and situations are a combination of customer-facing and internal aims.[90]

- **Business models**: how firms operate, from their go-to-market strategy and value proposition to how they make money and effectively shift their core company, tapping into fresh revenue sources and methodologies, and sometimes even abandoning the traditional core business.[91]

---

[88] Cp. *Ibid*.
[89] Cp. *Resego Morakanyane and Audrey A Grace*., Digital Transformation, 2017, no page number.
[90] Cp. *Janet L. Hartley and William J. Sawaya*., Business Processes, 2019, pp. 709.
[91] Cp. *Nico Roedder et al*., Data Analytics, 2016, pp. 2795.

- **Business ecosystems**: the business's network of partners and stakeholders, as well as contextual elements such as legislative or economic agendas and evolutions. On the fabric of digital transformation and information, new ecosystems are being established between organizations with varied backgrounds, where data and actionable intelligence become innovation assets.[92]

- **Business asset management**: where the emphasis is shifting away from traditional assets and toward less 'tangible' assets like information and customers (improving customer experience is a top goal of many digital transformation "projects," and information is the lifeblood of business, technological evolutions, and any human relationship).[93] Customers and information must be considered as valuable assets from every angle.

- **Organizational culture**: whereby a clear customer-centric, agile, and hyper-aware goal must be established, which is accomplished by developing core competencies across the board in areas such as digital maturity, leadership, knowledge worker silos, and so on, allowing the organization to become more future-proof.[94] Processes, commercial activities, cooperation, and the IT side of digital transformation all intersect with culture. Changes are required to get applications to market faster. Development and operations are at the heart of DevOps. Change is required to integrate IT and OT in businesses/processes/activities (not just the information and operational technologies, but also the procedures, culture, and teamwork). Etc.[95]

- **Ecosystem and partnership models**: With the advent of co-optative, collaborative, co-creating, and lastly, wholly new business ecosystem techniques, new business models and revenue sources are being created.[96] Ecosystems will be critical to the success of the as-a-service economy and digital transformation.

- **Customer, worker, and partner approach:** People and strategy come first in digital change. Any stakeholder's shifting behavior, expectations, and demands are

---

[92] Cp., *No Author.*, https://www.sydle.com/blog/digital-transformation-60d63745755d340eceb04c7e/, accessed May 29, 2022.
[93] Cp. *Ibid*.
[94] Cp. *Ibid*.
[95] Cp. *Ibid*.
[96] Cp. *Ibid*.

critical. Customer-centricity, user experience, worker empowerment, new workplace models, shifting channel partner dynamics, and other factors all play a role in many change subprojects.[97] It's vital to remember that digital solutions are never the only way to address any of these human issues, from employee satisfaction to improving customer experience. People first and foremost involve, respect, and empower others; technology is an extra enabler and part of the equation of choice and basic requirements.[98]

2.3.2 Types of Digital Transformation

Digital transformation isn't a single big thing. Often, businesses are just focused on organizational change, ignoring the fact that there are four different types of digital transformations. As a result, they are unable to fully benefit from all that digital transformation has to offer.[99]

Figure 9: Types of Digital Transformation

Source: https://www.fasterpr.com/2021/10/digital-transformation-4-types.html

1. Process Transformation

Process transformation requires altering components of a company's processes in order to reach new objectives. When a company's business processes need to be overhauled, it usually embarks on a business process transformation. An organization's procedures will

---

[97]Cp., *No Author.*, https://www.sydle.com/blog/digital-transformation-60d63745755d340eceb04c7e/, accessed May 29, 2022.
[98] Cp. *Ibid.*
[99] Cp., *No Author.*, https://www.servicesbytechdata.com/resources/the-4-types-of-digital-transformation-explained, accessed May 29, 2022.

be modernized, new technology will be integrated, money will be saved, and essential systems will be better integrated as a result of such a transformation.[100]

On the shop floor, we witness process transformation, with corporations such as Airbus using heads-up display glasses to increase the quality of human inspection of airplanes. We've also seen process revolutions in the consumer experience, with companies like Domino's Pizza entirely reimagining the food ordering process; Domino's Any Ware allows customers to purchase from any device. Because of this invention, the company was able to surpass Pizza Hut in sales. We're also seeing organizations use robotic process automation to improve back-office procedures like accounting and legal. Process transformation may provide a lot of value, and technology adoption in these areas is quickly becoming a must. These transitions are characterized by concentrated efforts centered on specific business areas.[101] Because these transformations tend to be focused efforts around specific areas of the business, they are often successfully led by a CIO or CDO.

Important Steps in Process Transformation:

Business process transformation takes the same stages as business process management but aims for a more significant change. When it comes to process transformation, there are a few things to consider:[102]

- **Identify the goals of the transformation:** Is the company's main goal to improve its systems? How about incorporating current technology? Processes to be adapted to a new organizational structure? What prompted the requirement?
- **Establish baseline metrics:** Organizations should gather the necessary data to demonstrate that the business process transformation will be successful. Cost, time, amount of errors, and other measurable indicators should all be included.
- **Bring in all stakeholders:** This includes soliciting feedback from all individuals involved in the process on what worked well in the previous one and what they expect from the current one.

---

[100] Cp. *Ibid*.
[101] Cp. *Terry McNulty and Ewan Ferlie.*, Process Transformation, 2016, pp. 1391.
[102] Cp., *No Author.*, https://kissflow.com/workflow/bpm/business-process-transformation/, accessed May 29, 2022.

- **Map out the best scenario:** Businesses should create the ideal workflow path using a diagramming tool that includes both human and system tasks that must be completed.
- **Set live and monitor:** Companies should start by introducing small teams to the new process and regularly monitor progress and any necessary changes to ensure success.

2. **Business Model Transformation**

Process transformation focuses on certain business sectors. The underlying building blocks of how value is delivered in a certain industry are the focus of business model modifications.[103] Traditional business models are being challenged by digital revolution. Netflix's redesign of video distribution and Apple's reinvention of music delivery: iTunes are two examples of this type of business model innovation.[104]

3. **Domain Transformation**

- Domain transformation entails gaining access to new markets and regions. They can reach a larger number of clients and even overtake established leaders in the domain by expanding into new sectors they have not before attempted to enter.[105]
- This is an area of digital transformation that is often disregarded and demands more attention. As the landscape shifts, organizations should look for new ways to contact and serve new customers, thereby expanding their capabilities. This type of transition is becoming more common as different types of firms try to provide new value and widen their horizons.
- This is made possible by digital technologies. Organizations can utilize their vast resources to make their mark and expand their horizons with a wide range of tools at their disposal, such as the Internet of Things and AI. Opportunities to tap into a variety of markets are sometimes already present; firms only need to dig beneath the surface to see them.
- The case of online retailer Amazon, which moved into a new market domain with the launch of Amazon Web Services, is a great example of how domain transformation works (AWS). In a field historically dominated by high-profile behemoths Microsoft

---

[103] Cp. *Andreas Zolnowski and Research Zolnowski.*, Business Model, 2016, no page number.
[104] Cp. *Ibid*.
[105] Cp., *No Author.*, https://www.bairesdev.com/blog/digital-transformation-types-to-understand/, accessed May 29, 2022.

and IBM, AWS is now the world's largest cloud computing/infrastructure service provider.[106]

- While it may not have appeared to be a clear move on Amazon's part, the online retailer's entrance into cloud services saw the company leverage capabilities and services it already had. For example, it has previously established large storage capabilities as part of its effort to underpin its global business while also providing crucial computing services to Amazon's start-ups and other emerging firms. So, everything was in place; all that was missing was the vision and confidence to make the leap into the cloud.
- Any company undergoing a digital transformation should be aware of the new prospects for domain change that come with modern technology integration.[107]

4. Cultural/Organizational Transformation

- A successful digital transformation involves more than simply technological upgrades or product redesigns. If a company's digital transformation efforts aren't aligned with its internal values and habits, it can have a negative impact on the company's culture.[108]
- The negative consequences range from a slow adoption of digital technology to a loss of market competitiveness, as well as the initiative's ultimate failure and loss of productivity and money. A thorough and collaborative effort, on the other hand, can assist in shifting the culture to comprehend, accept, and promote digital transformation.
- Leaders should develop a clear vision for the transition and convey it effectively throughout the organization on a regular basis. They must understand what sensible risks are and why they are worthwhile to take.
- Experian, a consumer credit agency, is one example of this cultural/organizational shift. By incorporating collaboration and agile development into its workflows, it was able to transform its company. It also sparked a fundamental change in employee concentration from equipment to data across the organization.

---

[106] Cp., *No Author.*, https://www.poppulo.com/blog/what-are-the-4-main-areas-of-digital-transformation, accessed May 29, 2022.
[107] Cp. *Ibid*.
[108] Cp. *Ibid*.

A core element of digital culture:[109]

- Promotes an external, rather than an internal, orientation.
- Prizes delegation over control
- Encourages boldness over caution
- Emphasizes more action and less planning
- Values collaboration more than individual effort

2.3.3 Approaches of Digital Transformation

1. Wait for proof of digital success

This first method will assist businesses in concentrating on empirical results. The problem with this method is that you run the danger of being stuck in limbo until your digital transformation takes off.[110] It gives competitors a significant edge if their digital success occurs sooner than yours. Because most businesses have a combination of digital and non-digital solutions, the first strategy is frequently the preferred option.

2. Develop an all-inclusive digital strategy

This method focuses on creating a strategy that is as complete as feasible right away. It necessitates a well-thought-out long-term strategy. The all-encompassing digital transformation strategy emphasizes culture change and quick innovation implementation. It is usually a costly strategy with numerous risks. Only companies with a high level of patience and a willingness to take risks should apply.[111]

3. Incremental delivery of digital skill

After selecting an initial aim and route, this method focuses on delivering benefits as a corporation travel toward a potential changing digital destination. However, as the company grows, lessons learned and fresh inputs are considered, affecting the digital goal and, in some cases, the path to the destination. As a result, this strategy focuses on producing a robust yet pliable plan that can adapt to industry developments over time.[112]

---

[109] Cp. *Ibid*.
[110] Cp. *No Author.,* https://www.chaione.com/blog/4-digital-transformation-types, accessed May 29, 2022.
[111] Cp. *Ibid*.
[112] Cp. *Ibid*.

2.4 Definition of Artificial Intelligence

While several definitions of artificial intelligence (AI) have emerged over the past few decades, John McCarthy offers the following definition in this 2004 paper, It's the science and engineering of artificial intelligence.[113] intelligence, especially intelligent computer programs. with a similar task of using computers to understand human intelligence, but AI need not be limited to methods of biological observation. works, "Computers and intelligence" published in 1950. In this paper, Turing, often called the "father of computers," asks the question, "Can machines think?" From there, he proposed a test, now known as the "Turing test", in which a human questioner would attempt to distinguish between a computer and human text responses. While this test has been the subject of numerous reviews since its publication, it remains an important part of the history of AI as well as a concept continuum in philosophy as it employs ideas around linguistics.

In its simplest form, artificial intelligence is a field, which combines computer science and robust datasets, to enable problem-solving. It also encompasses subfields of machine learning and deep learning, which are frequently mentioned in conjunction with artificial intelligence. These disciplines are comprised of AI algorithms that seek to create expert systems that make predictions or classifications based on input data.[114] Today, a lot of hype still surrounds AI development, which is expected of any new emerging technology in the market. As noted in Gartner’s hype cycle product innovations like self-driving cars and personal assistants, follow “a typical progression of innovation, from overenthusiasm through a period of disillusionment to an eventual understanding of the innovation’s relevance and role in a market or domain.” As Lex Fridman notes here in his MIT lecture in 2019, we are at the peak of inflated expectations, approaching the trough of disillusionment. As conversations emerge around the ethics of AI, we can begin to see the initial glimpses of the trough of disillusionment.

2.4.1. Types of Artificial Intelligence

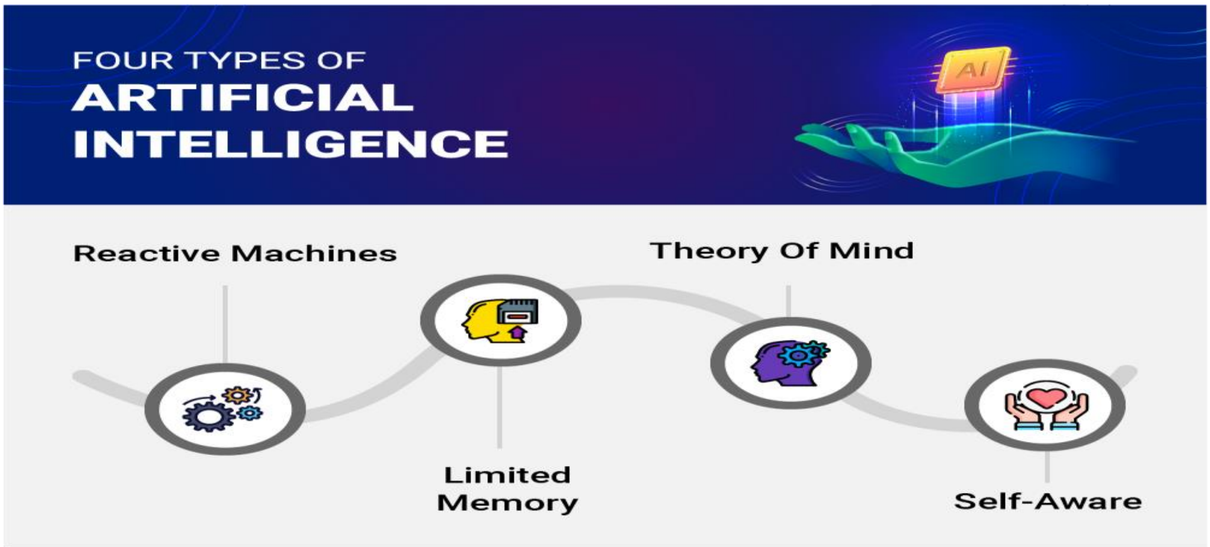
There are four types of artificial intelligence: reactive machines, limited memory, theory

---

[113] Cp. *Jonas Schuett.*, Definition of AI, 2019, no page number.
[114] Cp. *Dimiter Dobrev.*, Artificial Intelligence, 2012, no page number.

of mind, and self-perception.[115]

Figure 10: Types of Artificial Intelligence



Source: https://www.hdatasystems.com/blog/importance-and-benefits-of-artificial-intelligence

1. **Reactive Machines**

To inform current decisions. Deep Blue, the IBM chess supercomputer that beat international grandmaster Garry Kasparov in the late 1990s, is the perfect example of this. Deep Blue can identify pieces on the board. chess and know how to make each move. He can make predictions about possible next moves for him and his opponents. And he can choose the most optimal moves among the possibilities. But he has no concept of the past, nor any memory of what happened before. Unlike a rarely used chess-specific rule that repeats the same move three times, Deep Blue ignores everything before the current time. All it does is look at the pieces as it stands and choose from among possible next moves. This type of intelligence involves the computer perceiving the world directly. and acting on what he sees. an intrinsic concept of the world. In a paper, artificial intelligence researcher Rodney Brooks argues that we should just build machines like this. His main reason is that people are not good at programming accurate simulation worlds for computers, the so-called Artificial Intelligence is the "representative" of the world.

[115] Cp. *No Author.*, https://www.hdatasystems.com/blog/importance-and-benefits-of-artificial-intelligence, accessed June 1, 2022.

Current intelligent machines that surprise us There is no such concept of the world or a limited and specialized one for its functions. The design innovation of Deep Blue is not to expand the range of possible movies that the computer has considered. Instead, the developers have found a way to narrow their horizons and stop pursuing some potential future moves, based on how they evaluate their results. Without this ability, Deep Blue would need an even more powerful computer to defeat Kasparov.[116]

Similarly, Google's AlphaGo, which has beaten the top human Go experts, is also unable to assess all potential future moves. Its analytical methods are more sophisticated than Deep Blue, which uses neural networks to assess game development. These methods improve the AI system's ability to better play specific games., but they cannot be easily modified or applied to other situations easily fooled.

They are unable to participate in the interactive world, as we imagine AI systems might one day happen. Instead, these machines will behave the same every time they encounter the same situation. Trusted AI system: you want your self-driving car to be a reliable driver. But that's bad if we want machines to interact and respond to the world. These simplest AI systems will never get bored, overrated, or sad.

2. Limited Memory

This type contains machines that can see into the past. Self-driving cars are already part of this. For example, they observe the speed and direction of other cars. This cannot be done immediately; it is necessary to define specific objects. and track them over time. These observations are added to pre-programmed representations of the world of self-driving cars, which also include lane markings, traffic lights, and other key elements., like a bend in the road.[117] They are included when the vehicle decides when to change lanes, to avoid crossing other drivers or being hit by a nearby vehicle. But this simple information about the past is fleeting. how drivers synthesize their experiences over the years behind the wheel.

So how can we build AI systems that generate comprehensive representations, remember their experiences, and learn to handle new situations? Brooks was right that it was very

[116] Cp. *Ibid*.
[117] Cp. *Ibid.*

difficult to do. My research into methods inspired by Darwin's evolution may begin to fill in the gaps left by humans by letting machines construct their representations.

3. Theory of Mind

We can stop here and call this the big gap between the machines we have and the machines we will build in the future. However, it is best to be more specific in discussing what types of representational machines should form and which of them Machines are the next, more advanced type representing not only the world but also the agents or other entities in the world. In psychology, this is known as the "theory of mind" - the understanding that people, creatures, and objects in the world can have thoughts and feelings that influence their behavior. We have social interactions. Without understanding each other's motives and intentions, and without considering what others know about me or the environment, working together is at best difficult, at worst impossible. and that each of us has thoughts, feelings, and expectations about how we will be treated.[118] And they will have to adjust their behavior accordingly.

4. Self-Perception

The final step in the development of AI is to build systems capable of forming representations of them. Ultimately, our AI researchers will not only understand consciousness but also build machines with it. “Theory of mind” is possessed by type III artificial intelligence. Awareness is also called "self-awareness" for a reason. (“I want this item” is a very different statement from “I know I want this item”). Sensual beings are aware of themselves, know their inner state, and can predict the emotions of others. make those kinds of inferences.[119]

While we may be a long way from creating self-aware machines, we should focus our efforts on understanding memory, learning, and the ability to make decisions based on them. on past experiences. This is an important step in understanding human intelligence itself.[120] And it is paramount if we want to design or develop more special machines to categorize what they see in front of them.

---

[118] Cp. *Ibid*.
[119] Cp. *Ibid*.
[120] Cp. *Ibid*.

2.4.2 History of Artificial Intelligence

Five years later, the proof concept was initiated by Allen Newell, Cliff Shaw, and Herbert Simon, Logic Theorists. The Logic Theorist is a program designed to mimic human problem-solving skills and is funded by Research and Development (RAND). It is widely considered to be the first artificial intelligence program and was presented at the Dartmouth Summer Research Project on Artificial Intelligence (DSRPAI) organized by John McCarthy and Marvin Minsky in 1956.[121] created a term he coined at the very event. Unfortunately, the conference did not live up to McCarthy's expectations; People come and go as they please, and there is no agreement on standard methods for the field. Even so, everyone agrees with the idea that AI is possible.[122] of AI Research a Successful Glider and Back to Between 1957 and 1974, AI flourished. Computers can store more information and become faster, cheaper, and more accessible. Machine learning algorithms have also improved, and people have a better idea of which algorithm to apply to their problems. Early demonstrations such as Newell and Simon's Joint Problem Solver and Joseph Weizenbaum's ELIZA show promise for speech-language interpretation and problem-solving purposes.[123] These successes, coupled with advocacy from leading researchers (specifically DSRPAI participants) have convinced government agencies such as the Defense Advanced Research Projects Agency (DARPA) to finance support AI research at multiple institutions. The government is particularly interested in a machine capable of transcribing and translating spoken language as well as high-speed data processing. even higher. In 1970, Marvin Minsky told Life Magazine, "three to eight years from now we'll have a machine with the general intelligence of an average human."[124] However, while the basic principle is proven, there is still a long way to go to achieve the ultimate goals of natural language processing, abstract thinking, and self-recognition.

The original AI that broke through the fog revealed a mountain of obstacles. Most important was the lack of computing power to do anything important: computers simply couldn't store enough information or process it fast enough. For example, to

[121] Cp. *Vivek Kaul, Sarah Enslin, and Seth A. Gross*., Artificial Intelligence, 2020, pp. 810.
[122] Cp. *Michael Haenlein and Andreas Kaplan*., History of Artificial Intelligence, 2019, pp. 10.
[123]Cp. *Kaul, Enslin, and Gross*., History of Artificial Intelligence in Medicine, no page number.
[124] Cp. *Marvin Minsky*., Steps Toward Artificial Intelligence, 1961, pp. 12.

communicate, one must know the meaning of many words and understand them in many combinations.[125] Hans Moravec, McCarthy's doctoral student at the time, said that "computers are still too weak to be millions of times smart".

As patience dwindled, funding ran out, and research was delayed for ten years.

In the 1980s, AI was revived from two sources: expanding the algorithmic toolkit and raising funds. The expert system mimics the decision-making process of a human expert.[126] The program asked an expert in a field how to react in a certain situation, and when this has been learned for most situations, non-experts can get advice from the program. Expert systems have been widely used in industries. The Japanese government has heavily funded expert systems and other AI-related efforts under the Fifth Generation Computing Project (FGCP). and improve artificial intelligence. Unfortunately, most of the ambitious goals were not achieved. The indirect effects of FGCP have inspired a generation of talented young engineers and scientists. Either way, funding for the FGCP is gone and AI is on the radar. Ironically, in the absence of government funding and public hype, AI flourished.[127] During the 1990s and 2000s, many of the historic goals of artificial intelligence were achieved. In 1997, reigning world chess champion and grandmaster Gary Kasparov was defeated by Deep Blue, an IBM chess-playing computer program. artificial intelligence decision-making program.[128] That same year, speech recognition software developed by Dragon Systems was deployed on Windows. That's another big step but in the direction of trying to explain spoken language. A mechanical breakdown cannot be handled. Even human emotions are fair game, as evidenced by Kismet, a robot developed by Cynthia Breazeal that can recognize and express emotions. Time Heals All Wounds We haven't gotten smarter about how we code AI, so what's changed? It turns out that the fundamental limitation of computer storage that stopped us 30 years ago is no longer an issue. Law, who estimates that computer memory and speed double every year, has finally caught up and, in many cases, exceeded our needs. This is exactly how Deep Blue was able to beat Gary Kasparov in 1997 and how Google's Alpha Go was able to

[125] Cp. *Ibid*.
[126] Cp. *Baochuan Lu and Pam Smallwood*., Community, 2020, no page number.
[127] Cp. *J Paul Myers*., Computing Project, 2015, no page number.
[128] Cp. *Feng Hsiung Hsu*., Deep Blue, 1999, pp. 73.

beat Chinese Go champion Ke Jie just a few months ago.[129] This partly explains AI research; we saturate the AI capabilities at our current level of computing power (computer processing and storage speeds) and then wait for Moore's Law to catch up with us again.

Artificial Intelligence is Everywhere

We now live in the age of “big data”, an age in which we can collect large amounts of information that cannot be processed by a single person. The application of artificial intelligence in this area has been quite successful in several industries such as technology, banking, marketing, and entertainment.[130] We've seen that while algorithms don't get much better, big data and big computers only allow artificial intelligence to learn using brute force. There may be some evidence that Moore's Law is slowing down a bit, but data growth is certainly not losing momentum. Breakthroughs in computer science, math, or neuroscience all serve as a potential basis for breaking the ceiling of Moore's Law.[131] storage for the future? In the short term, the language of AI is like the next big thing. It is already underway. These days, machines even call me! One can imagine interacting with an expert system in a fluent conversation or having a conversation in two different languages translated in real-time. We can also expect to see driverless cars on the road in the next twenty years (and that's quite conservative). In the long run, the goal is general intelligence, i.e., a machine that surpasses human cognitive ability in every task. It matches the kind of sentient robot we often see in the movies. It seems unthinkable to me that this will be achieved in the next 50 years.[132] Even if the possibility is there, ethical issues will act as a powerful barrier against actualization. When that time comes (but better before it does), we'll have to have a serious conversation about machine politics and ethics (ironically, two topics that are fundamentally human.). Genetic. His current project employs the use of machine learning to model animal behavior. In his free time, Rockwell enjoys playing soccer and debating mundane topics.

---

[129] Cp. *Ibid*.
[130] Cp. *Weiyu Wang and Keng Siau*., Machine Learning, 2019, pp. 65.
[131] Cp. *Chris A. MacK*., Moore’s Law, 2011, pp. 203.
[132] Cp. *Weiyu Wang and Keng Siau*., Automation, 2019, pp. 64.

2.5 Definition of Agriculture

Over the centuries, the development of agriculture has contributed to the rise of civilizations.

Before agriculture became widespread, humans spent most of their lives foraging for food, and hunting animals. wild and pick wild plants.[133] About 11,500 years ago, people gradually learned to grow cereals and root crops and settled into a life based on agriculture. 2,000 years ago, most of the Earth's population became dependent on agriculture. When people started farming, they also started keeping and raising wildlife.[134] The adaptation of wild plants and animals to their use by humans is called domestication. The first plants to be domesticated were probably rice or maize in 7500 BC. The first domesticated animal was the dog, used for hunting. Finally, humans use livestock such as oxen for poling, pulling, and transporting. Agriculture helps humans produce surplus food. Surplus allows people to devote themselves to other tasks unrelated to agriculture.[135]

Agriculture kept the ancient nomads close to their fields and led to the development of permanent villages. They are linked by trade. The new economies have been very successful in several areas where cities have developed, and civilizations have developed. The first civilizations based on intensive agriculture appeared near Tigris and the Euphrates in Mesopotamia (now Iraq and Iran) and along the Nile in Egypt.[136]

Improved Technology

For thousands of years, the development of agriculture was very slow. One of the first agricultural tools was the fire. The Native Americans used fire to control the growth of berry-producing plants, which they knew grew rapidly after a wildfire. The farmer plowed the soil by hand, used an ax to uproot trees, and used a digging stick to break and pow the soil.[137] Over time, improved agricultural tools of bone, stone, copper, and iron were developed. New storage methods have evolved for use in times of scarcity. They also

[133] Cp. *No Author.,* Agriculture, 2014, no page number.
[134] Cp. *Dennis Keeney.*, Sustainable Agriculture, 1990, pp. 281.
[135] Cp. *Ibid*.
[136] Cp. *Hassan R. El-Ramady, Samia M. El-Marsafawy, and Lowell N. Lewis.,* Climate Changes, 2013, pp. 46.
[137] Cp. *Jean Paul Chavas and Céline Nauges.*, Technology Adoption, 2020, pp. 47.

began making clay pots and other containers for transporting and cooking food. Around 5500 BC, farmers in Mesopotamia developed simple irrigation systems. By channeling water from streams to their fields, farmers were able to settle in areas that were once considered unsuitable for agriculture. In Mesopotamia, then Egypt, and China, people organized and worked together to build and maintain better irrigation systems. Farmers have also developed improved crop varieties. For example, around 6000 BC, a new type of wheat appeared in South Asia and Egypt. Its husks were easier to remove and could be made into bread. As the Romans expanded their empire, they adapted the best agricultural methods from the people they conquered.[138] They wrote manuals on the agricultural techniques they observed in Africa and Asia and adapted the Chinese also adapted agricultural tools and methods from neighboring empires. A Vietnamese rice variety that ripens quickly and allows farmers to harvest multiple crops in one growing season. This type of rice quickly became popular throughout China. Many European farmers in the Middle Ages used a field-farming system. One field was planted in the spring, another in the fall, and another that remained uncultivated or fallow. This system conserved nutrients in the soil, increasing agricultural yields. The rulers of the Islamic Golden Age (which reached its peak around 1000) in North Africa and the Middle East considered this to be a science.[139] Farmers in the Islamic Golden Age learned about crop rotation.

During the 15th and 16th centuries, explorers introduced new crop varieties and agricultural products to Europe. From Asia, they brought home coffee, tea, and indigo, a plant used to make blue dye. From the Americas, they obtained crops such as potatoes, tomatoes, maize (maize), beans, peanuts, and tobacco. Some of them have become staple foods and expanded people's diets.

2.5.1. Artificial Intelligence in Agriculture

Global spending on smart and connected agricultural technologies and systems, including AI and machine learning, is expected to triple revenue by 2025, to $15.3 billion according to BI Intelligence Research.[140]

---

[138] Cp. *Ibid*.
[139] Cp. *Ibid*.
[140] Cp. *Indrajit Ghosh et al.*, Crop Recommendation, 2018, no page number.

Dedicated to AI technologies and solutions in agriculture is expected to grow from $1 billion in 2020 to $4 billion in 2026, achieving a compound annual growth rate/ year (CAGR) is 25.5%, according to Markets & Markets. IoT-enabled agricultural monitoring (IoT) is smart, the fastest growing technology segment of connected agriculture is expected to reach $0.5 billion by 2025, according to PwC.

AI, machine learning (ML), and IoT sensors provide real-time data to algorithms that help increase agricultural efficiency, improve crop yields, and reduce the daily cost of food production. According to United Nations population and hunger projections, the world's population will increase by 2 billion by 2050, which will require a 60% increase in food production to feed them. U.S.A. according to the U.S. Department of Agriculture's Bureau of Economic Research, the growing, processing, and distribution of food alone is a $1.7 trillion business.[141] AI and ML have shown the potential to help close the gap in food needs projected for an additional 2 billion people worldwide by 2050.

Uses AI and machine learning-based surveillance systems to monitor real-time video feeds from each crop field, identify the animal or human violations, and send alerts instantly. AI and machine learning reduce the chances of wildlife and domestic animals accidentally damaging a crop or a remote farming site. With rapid advances in AI-powered video analytics and machine learning algorithms, everyone involved in agriculture can protect their fields and perimeters. AI and machine learning CCTV systems scale as easily to a large-scale farm operation as it does to an individual farm. Machine learning-based surveillance systems can be programmed or trained over time to identify personnel from vehicles and use machine learning to identify employees who are working on-site.[142]

AI and machine learning improve crop yield predictions with real-time sensor data and visual analytics data from the aircraft driver. The amount of data collected by smart sensors and drones that deliver real-time video streaming gives agriculture professionals a whole new set of data they've never had access to before. Machine learning is the ideal technology to combine huge data sets and provide limit-based advice to optimize crop

[141] Cp. *N. N. Misra et al*., IoT, 2020, pp. 1.
[142] Cp. *Ibid*.

yields.[143] Here's an example of how AI, machine learning, buried sensors, infrared imaging, and real-time video analytics combine to provide farmers with new insights into how they can improve their health and safety.

Crop yields: Yield mapping is an agricultural technique based on supervised machine learning algorithms to find patterns in large-scale data sets and understand their orthogonality in real-time, this is invaluable for crop planning. It is possible to know the potential yield rate of a given field before the crop cycle begins. Using a combination of machine learning techniques to analyze 3D maps, social condition data from sensors, and drone-based soil color data, agricultural professionals can now predict yields of potential land for a given crop.[144] A series of flights are performed to obtain as much of the exact data set as possible.

2.6 Definition of Smart Farming

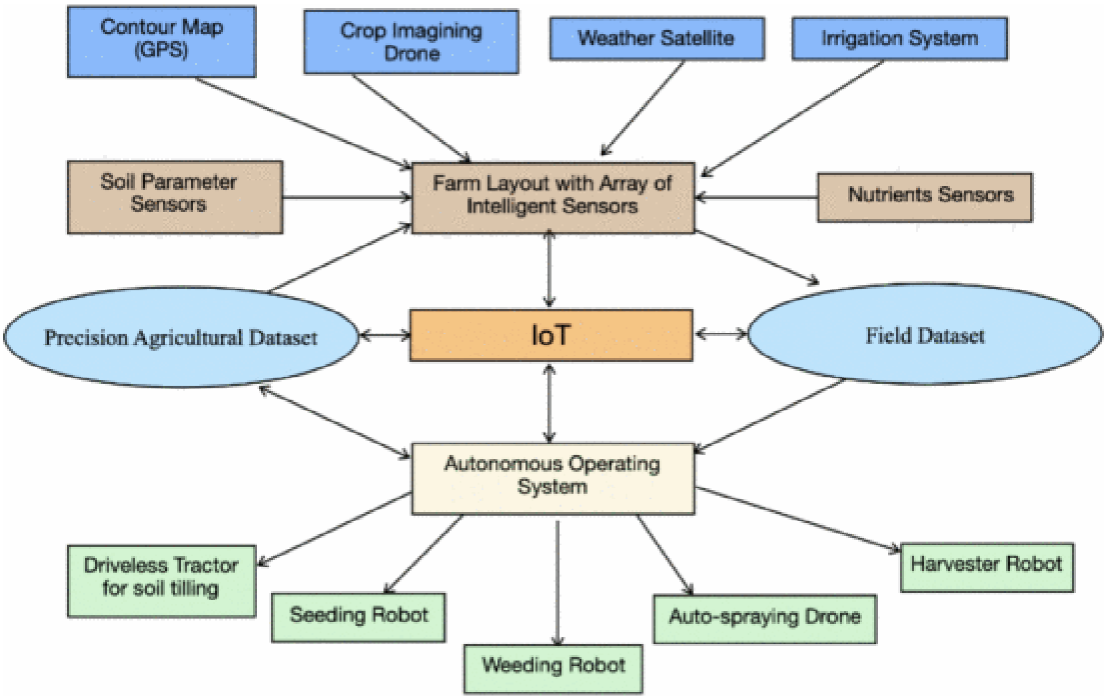
Smart Farming represents the application of modern Information and Communication Technologies (ICT) into agriculture, leading to what can be called the Third Green Revolution.
Following the plant breeding and genetics revolutions, this Third Green Revolution is taking over the agricultural world based upon the combined application of ICT solutions such as precision equipment, the Internet of Things (IoT), sensors and actuators, geo-positioning systems, Big Data, Unmanned Aerial Vehicles (UAVs, drones), robotics, etc.[145]

Smart Farming has a real potential to deliver a more productive and sustainable agricultural production, based on a more precise and resource-efficient approach. However, while in the USA possibly up to 80% of farmers use SFT, in Europe it is no more than 24%.[146] From the farmer's point of view, Smart Farming should provide the farmer with added value in the form of better decision-making or more efficient exploitation operations and management. In this sense, smart farming is strongly related

[143] Cp. *Poorna Shankar, Urvi Patel et al.*, Crop Prediction, 2022, pp. 128.
[144] Cp. *Ibid*.
[145] Cp. *Sjaak Wolfert et al*., Smart Farming, 2017, pp. 73.
[146] Cp. *Vasileios Moysiadis et al*., Smart Farming, 2021, pp. 100345.

to three interconnected technology fields addressed by Smart AKIS Network.[147]

Figure 11: Smart Farming



Source: https://ieeexplore.ieee.org/abstract/document/9297856

Management Information Systems: Planned systems for collecting, processing, storing, and disseminating data in the form needed to carry out a farm's operations and functions.

Precision Agriculture: Management of spatial and temporal variability to improve economic returns following the use of inputs and reduce environmental impact. It includes Decision Support Systems (DSS) for whole-farm management to optimize returns on inputs while preserving resources, enabled by the widespread use of GPS, GNSS, aerial images by drones, and the latest generation of hyperspectral images provided by Sentinel satellites, allowing the creation of maps of the spatial variability of as many variables as can be measured (crop yield, terrain features/topography, organic matter content, moisture levels, nitrogen levels, etc.).[148]

Farm Automation and Robotics: The process of applying robotics, automatic control, and artificial intelligence to all levels of agricultural production, including agricultural

[147] Cp. *Ibid*.
[148] Cp. *P. C. Robert*., Precision Agriculture, 2002, pp. 143.

robots and agricultural drones. Smart Farming applications are not only targeted at conventional large farms but can also be new leverage to drive other current or growing trends in farms, such as family farming. (Small or complex spaces, specific crops and/or livestock, high quality, or special seed preservation, etc.), organic farming and promotion of transparent and highly respected agriculture according to the conscience of consumers, society, and the European market.[149] Smart farming can also bring huge benefits to environmental issues, such as through more efficient use of water or the optimization of treatment methods and inputs.

---

[149] Cp. *Akshay Krishnan, Shashank Swarna, and H. S. Balasubramanya*., Robotics, 2020, no page number.

3. Research Design

All academic research depends on the methodology and the clarity of reporting.[150] Due to this the guidelines are to be specified on how and what is to be done and the set of rules that is to be followed.

3.1 Structured/Systematic Literature Review

A structured literature review, also known as a systematic review, is a sort of review that use an iterative analytical technique to gather secondary data before beginning to analyze it.[151] These systematic reviews build a synthesis that further formulates research questions with a broader scope, then identifies and synthesizes the data to link to the review questions.[152]

For a variety of reasons, systematic literature reviews are done. They assist in providing a theoretical foundation for ongoing study and in-depth learning of the research topic. It provides a clearer and more thorough summary of the collected information for a topic we're working on, for example. It also aids us in identifying research gaps between our present knowledge of the topic and our intended outcomes. The primary goal of researchers who use and use systematic literature reviews is to ensure that the results and conclusions are all founded on facts and evidence.[153]

Analyzed each document and split it according to the needs for selecting distinct review methodologies. These methods are important because they help you understand the nature of the literature review. The most noticeable quality of literature reviews is their concision, which is critical when dealing with vast amounts of material. As previously said, these are the points that assist us in selecting acceptable literature studies that we may combine into a "mixed method design" that is appropriate for these paper project working on and thus expect desirable outcomes.[154]

In a nutshell, the quantitative technique creates numerical data that is then employed in studies, or even if non-numeric data is acquired, it is only transformed to numbers,

[150] Cp. *David Moher et al*., Systematic Reviews, 2009, no page number.
[151] Cp. *No Author*., Systematic Reviews, no page number.
[152] Cp. *No Author*., Literature Review, no page number.
[153] Cp. *Hannah Snyder*., Research Methodology, pp. 336.
[154] Cp. *Jack Daniel Rittelmeyer and Kurt Sandkuhl*., Effects of Artificial Intelligence, 2021, pp. 133.

whereas the qualitative approach generates textual data.[155]

Keywords are words that can sum up your research or review, and any reader can deduce what to expect from the literature review based on the keywords you choose. Prior to the data and information, this paper will associate with the research issue, relevant keywords are needed first to be searched. Keywords also play a significant effect in the outcome.

The following are the keywords that were used in this study:

- Crop Leaf Disease
- EfficinetNetB3
- Smart Farming
- Digital Transformation
- Agriculture Industry
- Big Data
- Artificial Intelligence
- Convolution Neural Network
- Descriptive Analytics
- Prescriptive Analysis
- Structured Literature
- Quantitative data analysis
- Qualitative data analysis
- Predictive Analysis
- Importance of literature review
- Types of research

[155] Cp. *Feyisa Mulisa*., Quantitative, 2021, pp. 120.

Selection criteria:

- For our research on the review for study queries, we predicted the criteria for employing the selection of literatures. We employed two types of criteria: first, inclusion criteria, in which he included distinct definite reception, included that the learning is investigative, and included that the learning must in the previous ten years, a great deal of information has been published and generated.
- In the second approach, exclusion criteria, research is used as an empirical method, qualitative technique is used, the research was published more than 10 years ago, and it may be published in languages other than English.[156]
- There are various questions to ask when selecting and reviewing literature for your study, such as:
- Do you know what your research topic's scope and purpose are?
- Have you correctly cited the source of the literature?
- Have you incorporated studies that are current and relevant?
- Have you employed primary and secondary sources that are relevant to your topic?
- Is the connection to the problem statement obvious?
- Are the research limitations, including methodology and design, adequate?

3.2 Justification of Research Approach

1. Quantitative Method

Quantitative as the name suggests talks about quantity viz. the numbers. This method usually deals with numbers and statistics. This method usually deals with finding relationships between variables and how they change with respect to each other (i.e., their correlations).[157]

There are two types of statistics namely descriptive and inferential statistics.

Descriptive Statistics: This kind of statistics is usually performed on a sample size which

---

[156]Cp. *No Author*., Literature Review, no page number.
[157] Cp. *Oberiri Destiny Apuke*., Quantitative Research Methods , 2017, pp. 45.

is a subset of an entire population. A sample size usually represents the whole population accurately. A population can be defined as the whole set of sample data that is collected. Additionally, while collecting the data there are a few things that are to be kept in mind like the mode of data collection, if the data collected is whole and these collected data is represented in the form of graphs and infographics which provides insights into the data that is to be dealt with. The visualization of the data helps to interpret the data in the form of story.[158]

Inferential Statistics: This kind of statistics in used to draw inferences and make predictions from a given sample data set. If this inferential statistic is to be performed for a larger sample size, we could draw inferences to an entire population as well.[159]

Inferential Statistics can be performed in the following ways:

1. **Hypothesis Testing**: Hypothesis testing is where an assumption is made called the 'Null-Hypothesis'(H0) which is tested against the data set. If the null hypothesis holds good it is accepted and if it, doesn't it is rejected, and the Alternative Hypothesis (H1) is accepted.[160] In our case point we have chosen the null hypothesis to be if a transaction is fraudulent or not.

2. **Confidence Intervals:** Instead of taking data points the data is converted to intervals and the statistics is applied to it. And tested against the hypothesis.

3. **Regression Analysis:** This method is usually used for prediction. This method discusses the significance of the independent variable on dependent variable. There are several types like univariant and multivariant regression analysis. Univariant is usually where several dependent variables depend on one single independent variable. Whereas, in multivariant several independent variables determine the value of a single dependent variable.[161]

---

[158] Cp. *George Argyrous and George Argyrous*., Statistics for Research, 2005, pp. 401.
[159] Cp. *Shane Allua, Laura M. Stapleton, and S. Natasha Beretvas*., Mixture Models, 2008, pp. 360.
[160] Cp. *David R Anderson, Kenneth P Burnham, and William L Thompson*., Null Hypothesis, 2000, pp. 921.
[161] Cp. *Gülden Kaya Uyanık and Neşe Güler*., Multiple Linear Regression Analysis, 2013, pp. 235.

2. Qualitative Method

The Qualitative methods refer to the descriptive part of the data collected.[162] They have more depth to the research. There are several methods to qualitative research, they are:

1. **Focus Groups:** This method refers to a group of people that accurately represents the population. They are usually given a task, or a group of tasks and they are observed. The tasks are usually like filling a questionnaire etc. This is usually carried out during the launch of a product. They are asked what they like and don't like? What can be changed? Etc. And the changes are made to the products based on the results observed.
2. **Interviews:** Interviews are a common method of qualitative research where an interviewer asks a specifically framed few questions and the answers are recorded for research purposes and inferences are drawn based on the answers.
3. **Case Studies:** This type of research focuses on the problems that have already occurred and research is done on various aspects about what could have been done? And what could be done to avoid the problem in the future? Etc.

3. Mixed Method

This is the most effective and most used methodology in research. This method comprises of a combination of both qualitative and quantitative research method- ology.[163]

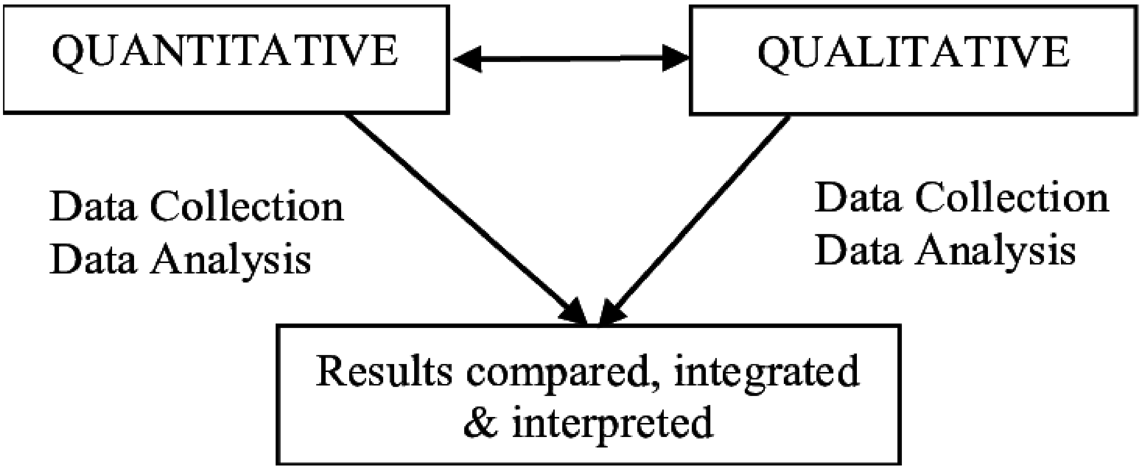
Some of the benefits of using mixed methods are:

1. It offsets the disadvantages of the individual research methods and provides robust research results.
2. It provides better interpretation and understanding of the data.

This method uses mixed method of research as we do hypothesis testing and visualization of data to draw inferences with the case study of credit card fraud detection using predictive models.

[162] Cp. *Haradhan Mohajan and Haradhan Kumar Mohajan*., Qualitative Research Method, 2018, pp. 1.
[163] Cp. *Gabriela López-Aymes et al*., Mixed Methods Research, 2021, pp. 680.

Figure 12: Representation of Mixed Method of Research.



Source: https://www.researchgate.net/figure/Concurrent-Triangula- tion-Design-Source-Creswell-2009_fig1_317032837accessed on 27.01.2022

Therefore, this paper uses quantitative approach for the study, as we collect data first and then use the statistical method and technological tools. The data is collected from a secondary data source Kaggle website where it is saved.

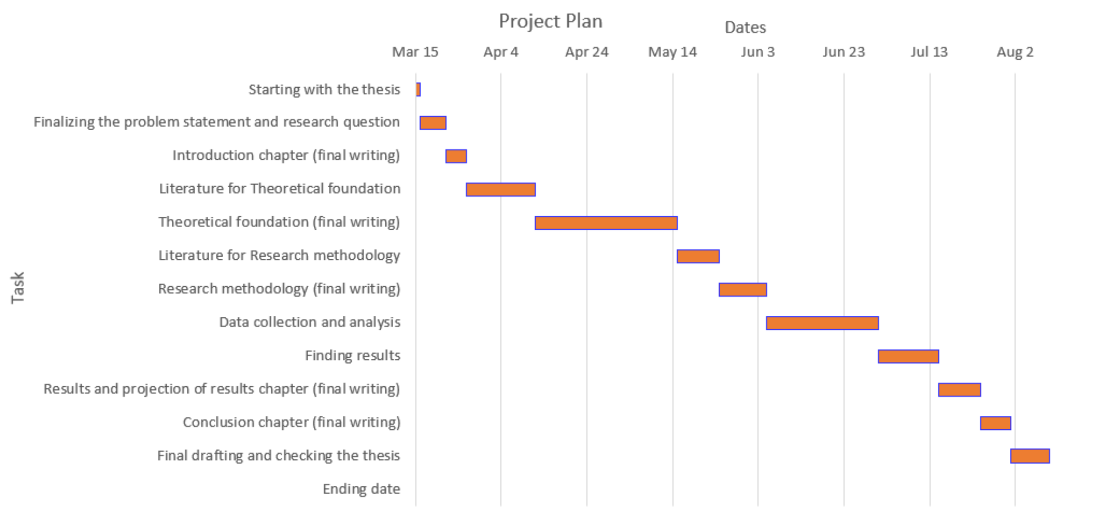
3.3 Project Plan

This project's plan is as follows:

- On the 15th of March 2022, this project was started.
- The problem statement and research question for this project was finalized on 16th March 2022.
- On the 22nd of March 2022, started writing introduction chapter for this project.
- The source for the literature review were started on 27th of March and on the 12nd of April 2022, started writing theoretical foundation chapter for this project.
- The source for the literature review were started on 15th of May and on the 25th of May 2022, started writing research methodology chapter for this project.
- Data collection and analysis started on the 5th of June 2022.
- Started finding results on the 1st of July 2022.

- On the 15th of July 2022, started writing projection of results chapter for this project.
- On the 25th of July 2022, started writing conclusion for this project.
- From 1st of August 2022, started final drafting and checking this project.
- On the 16th of August 2022, final submitted this project.

Figure 13: Project Plan



	Ending date	Final drafting and checking the thesis	Conclusion chapter (final writing)	Results and projection of results chapter (final writing)	Finding results	Data collection and analysis	Research methodology (final writing)	Literature for Research methodology	Theoretical foundation (final writing)	Literature for Theoretical foundation	Introduction chapter (final writing)	Finalizing the problem statement and research question	Starting with the thesis
Date	16-08-2022	01-08-2022	25-07-2022	15-07-2022	01-07-2022	05-06-2022	25-05-2022	15-05-2022	12-04-2022	27-03-2022	22-03-2022	16-03-2022	15-03-2022
Duration	0	9	7	10	14	26	11	10	33	16	5	6	1

Source: Author Own Represent

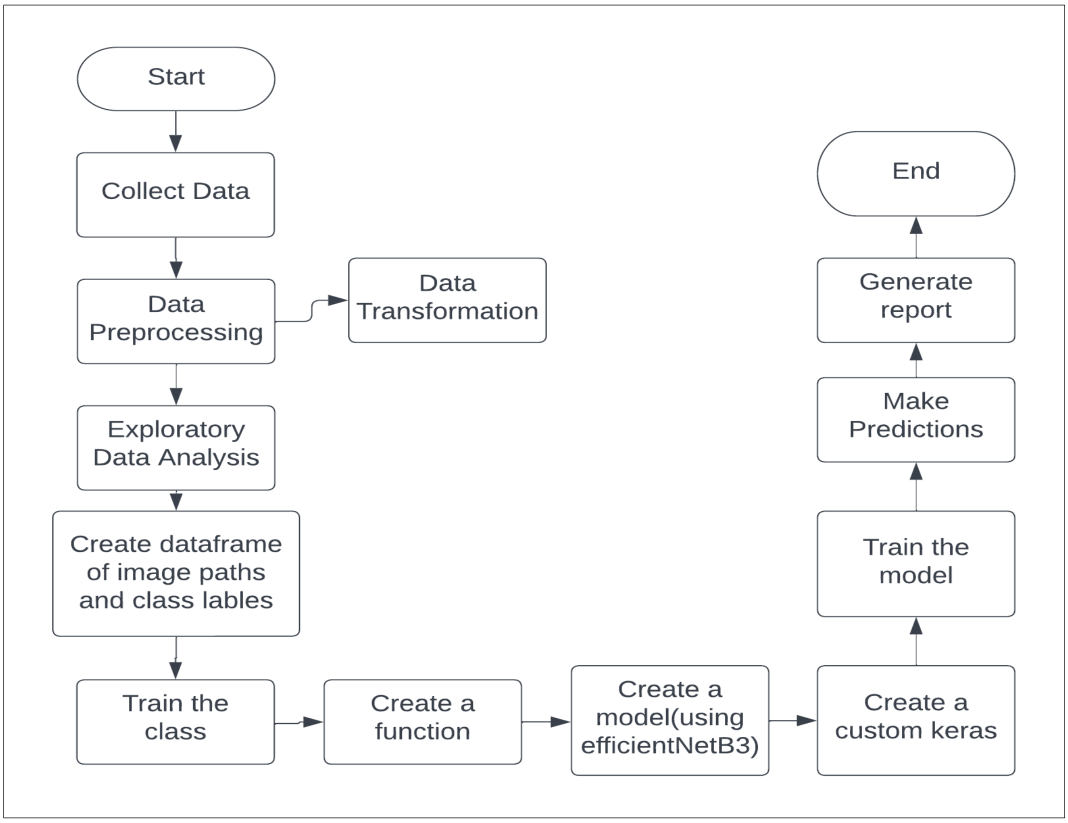
3.4 Project Flow

The project flow of this study is as follows:

- After deciding upon the Research questions, the data collection was started for this thesis.
- After the data collection, the focus was on data preprocessing, in data processing transforming the data into different class.
- After which, Exploratory data analysis was started for this thesis to learn what the data says, i.e., for finding out some useful insights from the data.

- Perform read operation in images and create a data frame of image paths and class labels.
- After that train the class for this thesis and create a function to show training image samples.
- Then, started building model and created custom Keras for this thesis.
- After that trained the model and make predictions for this thesis.
- At the end, generated report for detected disease in crop and predict crop yield to an overview of this thesis.

Figure 14: Project Flow



Source: Author Own Representation

3.5 Data Collection & Description

The data for this thesis has been collected from different sources like Kaggle and PlantVillage dataset. In this dataset available total number of classes is 14 and total number of images is 13,024.

The following are the class labels that were used in this thesis:

- Corn_Common_Rust
- Corn_Gray_Leaf_Spot
- Corn_Healthy
- Corn_Northern_Leaf_Blight
- Potato_Early_Blight
- Potato_Healthy
- Potato_Late_Blight
- Rice_Brown_Spot
- Rice_Healthy
- Rice_Leaf_Blast
- Rice_Neck_Blast
- Wheat_Brown_Rust
- Wheat_Healthy
- Wheat_Yellow_Rust

Corn Plant Species: There are 4 types of corn plant species into this dataset.

1. Common Rust (1192 images)

2. Gray Leaf Spot (513 images)

3. Healthy (1162 images)

4. Northern Leaf Blight (985 images)

Total images: 3852 images and 52,229,657 bytes

All the images of the Corn species have been collected from the PlantVillage dataset.[164] PlantVillage is a standard and most popular leaf image dataset for plant disease detection.

Potato Plant Species: There are 3 types of potato plant species into this dataset.

[164] Cp. *David. P. Hughes and Marcel Salathe*., Disease Diagnostics, 2015, no page number.

1. Early Blight (1000 images)

2. Healthy (152 images)

3. Late Blight (1000 images)

Total images: 2152 images and 39,441,201 bytes

All the images of the Potato species have been collected from the PlantVillage dataset.[165]

Rice Plant Species: There are 4 types of rice plant species into this dataset.

1. Brown Spot (613 images)

2. Healthy (1488 images)

3. Leaf Blast (977 images)

4. Neck Blast (1000 images)

Total images: 4078 images and 2,105,364,423 bytes

Rice images were collected from the datasets:

1. Dhan-Shomadhan: A Dataset of Rice Leaf Disease Classification for Bangladeshi Local Rice.[166]

The images from the Leaf background images of the classes Brown Spot and Rice Blast were added to the classes Rice_Brown_Spot and Rice_Leaf_Blast to increase the number of training samples. This dataset has the license CC BY 4.0. The images were not changed or edited.

2. "Rice Leafs" dataset from Kaggle.[167]

The images of this dataset were added to the classes Rice_Brown_Spot, Rice_Healthy, and Rice_Leaf_Blast.

Wheat Plant Species: There are 3 types of wheat plant species into this dataset.

1. Brown Rust (902 images)

---

[165] Cp. *Ibid*.
[166] Cp. *Md Rayhan Chowdhury and Shahadat Hossain*., Analysis of Genetic Variability, 2021, no page number.
[167] Cp. *No Author*., https://www.kaggle.com/datasets/shayanriyaz/riceleafs accessed August 4, 2022,.

2. Healthy (1116 images)

3. Yellow Rust (924 images)

Total images: 2942 images and 405,695,274 bytes

Wheat images were collected from the datasets: "Wheat Disease Detection" dataset from Kaggle.[168]

3.6 Algorithm Description

1. Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is a form of artificial neural network that is specifically made to process pixel input and is used in image recognition and processing.[169]

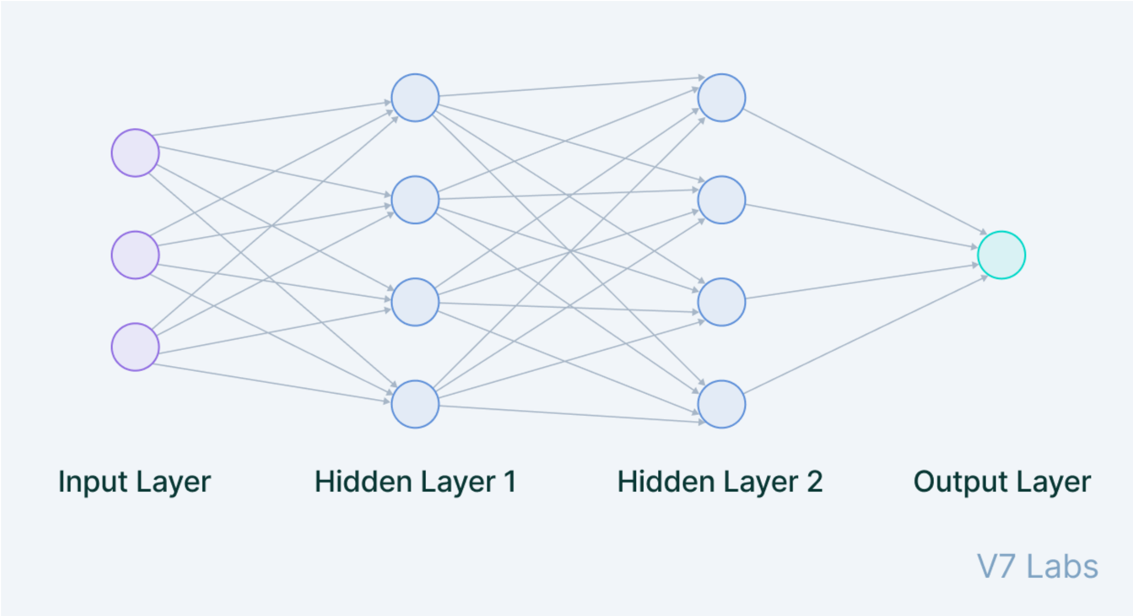
CNNs are effective artificial intelligence (AI) systems for image processing that employ deep learning to carry out both generative and descriptive tasks. They frequently use machine vision, which includes image and video identification, recommender systems, and natural language processing (NLP).[170]

CNN is composed of an input layer, an output layer, and many hidden layers in between. Tens or even hundreds of layers can be present in a convolutional neural network, and each layer can be trained to recognize various aspects of an image. Each training image is subjected to filters at various resolutions, and the result of each convolved image is utilized as the input to the following layer. Beginning with relatively basic criteria like brightness and borders, the filters can get more complex until they reach features that specifically identify the object.

---

[168] Cp. *No Author*., https://www.kaggle.com/datasets/sinadunk23/behzad-safari-jalal accessed August 4, 2022.
[169] Cp. *No Author*., https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network accessed August 7, 2022.
[170] Cp. *Ibid*.

Figure 15: CNN Architecture



Source: https://www.v7labs.com/blog/convolutional-neural-networks-guide

These layers carry out operations on the data to discover characteristics unique to the data. Convolution, activation or ReLU, and pooling are three of the most used layers.

- Convolution runs a series of convolutional filters through the input images, activating different aspects of the images with each filter.
- Rectified linear unit (ReLU), which maintains positive values while translating negative values to zero, enables quicker and more efficient training. Since only the activated features are carried over to the following layer, this is frequently referred to as activation.
- By conducting nonlinear downsampling on the output, pooling reduces the number of parameters the network needs to learn.
- Next, by applying the nonlinearity or an activation function to the convolution layer output, we obtain the following:[171]

$$h^k = f(W^k * x + b^k)$$

[171] Cp. *Laith Alzubaidi et al*., Deep Learning, 2021, pp. 11.

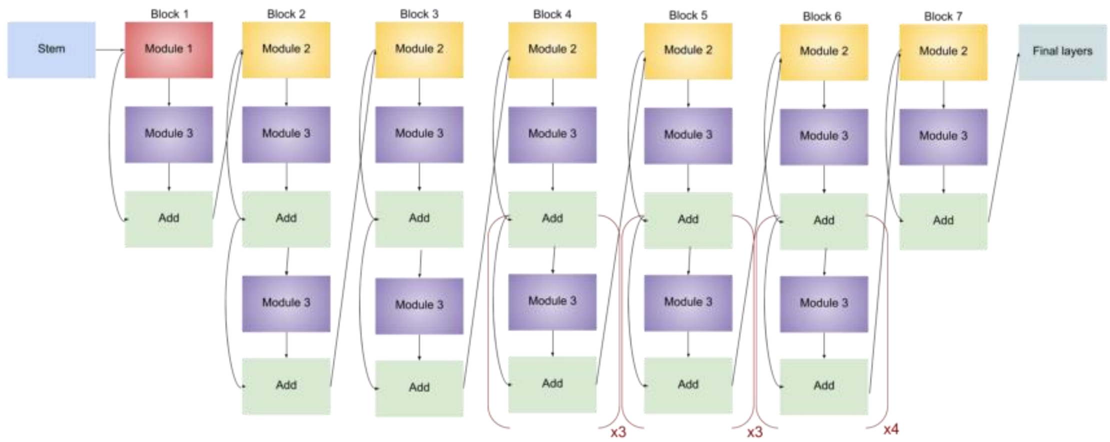
2. EfficientNet Model

EfficientNet is a convolutional neural network design and scaling technique that uses a compound coefficient to consistently scale all depth, width, and resolution dimensions. The EfficientNet scaling method evenly scales network width, depth, and resolution using a set of preset scaling coefficients, in contrast to standard practice, which scales these variables arbitrarily.[172]

According to Tan and Le, EfficientNet is a family of ConvNets that is accurate and effective and is based on a scaled-up version of the standard Neural Architecture Search (NAS) model.[173] EfficientNet was created utilizing the concept of uniformly scaling the model in all dimensions (depth, width, and resolution) using a simple compound coefficient. Better model accuracy was obtained when scaling up ConvNets by balancing all dimensions using a constant ratio. Based on this concept, if it is intended to use 2^n times more computational power, model can be scaled up in depth by a^n, in width by b^n and in resolution by c^n, where a, b, c represent constants.

In EfficientNet model there are different models from EfficientNet B0 to B7. In this thesis EfficientNetB3 model was used.

Figure 16: Architecture for EfficientNetB3



Source: https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142

---

[172] Cp. *Mingxing Tan and Quoc V. Le*., EfficientNet, 2019, pp. 10691.
[173] Cp. *Ibid.*

EfficientNetB3 function:[174]

```
tf.keras.applications.EfficientNetB3(
    include_top=True,
    weights="imagenet",
    input_tensor=None,
    input_shape=None,
    pooling=None,
    classes=1000,
    classifier_activation="softmax",
    **kwargs
)
```

This function returns a Keras image classification model, optionally loaded with weights pre-trained on ImageNet.

Each Keras application expects a specific kind of input preprocessing. For EfficientNet, input preprocessing is included as part of the model (as a Rescaling layer), and thus and thus tf.keras.applications.efficientnet.preprocess_input is actually a pass-through function. EfficientNet models expect their inputs to be float tensors of pixels with values in the [0-255] range.

3.7 Tools Used in Study

We utilized the following tools in this research:

- Python
- Jupyter notebook
- Tensorflow

Python:

Python is a high-level coding language that is easy to learn, cooperate with, and is object-oriented. Python is intended to be an easy language to learn. It employs English terms instead of punctuation frequently, and it has less syntactical structures than other languages.

Python is one of the few programming languages that allows the user to think like a coder and focus on the work at hand rather than on the syntax. This allows the student to go quickly into topics that demand a deeper grasp. The novice will also become acquainted

[174] Cp. *No Author.,* https://keras.io/api/applications/efficientnet/#efficientnetb3-function accessed August 7, 2022,.

with programming paradigms and principles that are important to any coder.[175]

Features of python are:

- Easy-to-learn: Python has limited keywords, a straightforward structure, and a clearly defined syntax, making it easy to learn. This allows the pupil to quickly learn the language.
- Easy-to-read: Python code that is easy to read is more clearly defined and obvious to the eyes.
- Easy-to-maintain: Simple to maintain Python's source code is also simple to maintain.
- A large standard library: The majority of Python's library is portable and cross-platform compatible on UNIX, Windows, and Macintosh systems.
- Interactive Mode: Python includes an interactive mode that enables for collaborative testing and debugging of code snippets.
- Portable: Python is portable, which means it can run on a broad range of hardware platforms while maintaining a consistent user interface.
- Extendable: The Python interpreter can be enhanced with low-level modules. These modules enable programmers to make their tools more efficient by improving or modifying them.
- Databases: Python has interfaces for all of the most popular commercial databases.
- GUI Programming: Python allows you to create graphical user interfaces that can be molded and converted to a variety of system calls, libraries, and operating systems, including Windows MFC, Macintosh, and Unix's X Window system.
- Scalable: Python is more scalable than shell scripting in terms of structure and support for large programs.

[175] Cp. *Ramachandran Trichur Narayanan*., Python, 2019, no page number.

Jupyter notebook:

The most extensively used solution for expressing literate programming is Jupyter Notebook. It was designed with the goal of making data analysis more easily documented, distributed, and duplicated. Since its inception in 2013, the system has grown to include over 9 million notes on GitHub.[176] Jupyter is based on IPython and supports a few programming languages in addition to Python, including Julia, R, JavaScript, and C. It also allows not just code and text to be inserted, but also various types of rich media, such as image, video, and even interactive widgets that combine HTML and JavaScript.[177]

The following are some of the benefits of utilizing a Jupyter notebook:

- They are useful for displaying your work. The code as well as the results are both visible. The notebooks on Kaggle (an online community of data scientists and machine learning practitioners) are a fantastic example.
- As a starting point, applying for other people's employment is simple. You can execute the code cell by cell to see what the code generates.
- Extremely simple to host server-side, which is beneficial for security. A great deal of information is sensitive and must be safeguarded. This is provided for free by a server-side Jupyter Notebook configuration.

Tensorflow:

TensorFlow is an open-source library developed by Google primarily for deep learning applications. Traditional machine learning is also supported. TensorFlow was first created without having deep learning in mind to handle huge numerical computations. However, it turned out to be quite helpful for the development of deep learning as well, so Google made it open source.

Tensors, which are multi-dimensional arrays with more dimensions, are the only type of data that TensorFlow takes. When handling a lot of data, multi-dimensional arrays come in quite helpful.

Data flow graphs with nodes and edges serve as the foundation for TensorFlow's

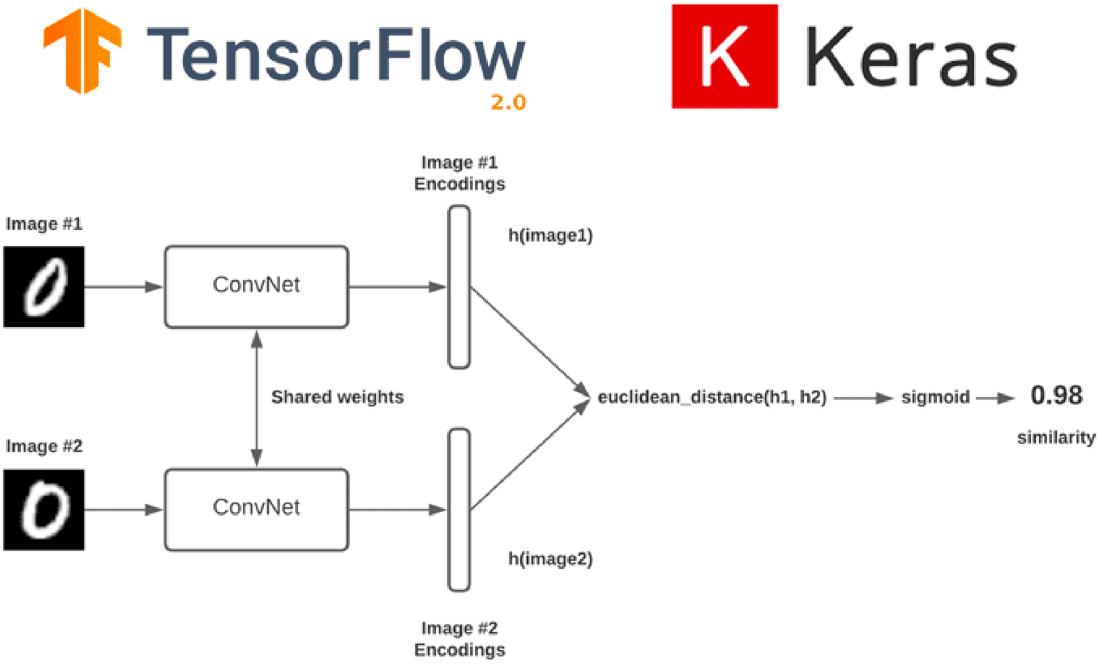
---

[176] Cp. *João Felipe Pimentel et al.*, Jupyter Notebooks, 2021, no page number.
[177] Cp. *Joao Felipe Pimentel et al.*, Quality and Reproducibility, 2019, pp. 510.

operation. It is considerably simpler to distribute the execution of TensorFlow code using GPUs across a cluster of computers because the execution mechanism takes the form of graphs.[178]

Tensorflow Architecture:

Figure 17: Tensorflow Architecture



Source: https://pyimagesearch.com/2020/11/30/siamese-networks-with-keras-tensorflow-and-deep-learning/

Tensorflow architecture works in three significant steps:

- Data pre-processing - structure the data and brings it under one limiting value.
- Building the model - build the model for the data.
- Training and estimating the model - use the data to train the model and test it with unknown data.

---

[178] Cp. *Bo Pang, Erik Nijkamp, and Ying Nian Wu.*, TensorFlow, 2019, pp. 231.

4. Projection of Results

This section discusses the application of the adaptation of AI and ML in Agriculture industry, followed by a practical example of crop disease detection by using EfficientNetB3 model.

4.1. Machine Learning and Artificial Intelligence in Transformation of the Agricultural Industry.

This paper uses the mixed research approach and integrative literature review to study the state of Machine Learning and Artificial Intelligence in the Agriculture industry with a detailed code review of crop disease detection system with the application of Deep Learning Algorithms. The world has experienced tremendous growth in technology because of exponential growth and revolutions in various industry. AI is one of the consequences of the result of such exponential growth of technological industry.

The world has seen many significant revolutions throughout history which altered the whole face of the earth. One of the significant revolutions was introduction of information technology in 1969. And the present revolution in technology is the introduction of Artificial intelligence which looks to automate and autonomize everything we use and most of the things around us.[179]

The AI portrays the evanescing of the thin line that exist between the physical, biological and the computerized world. It looks to bring forth computers that have human-like thinking created using complex algorithms and heavy research.

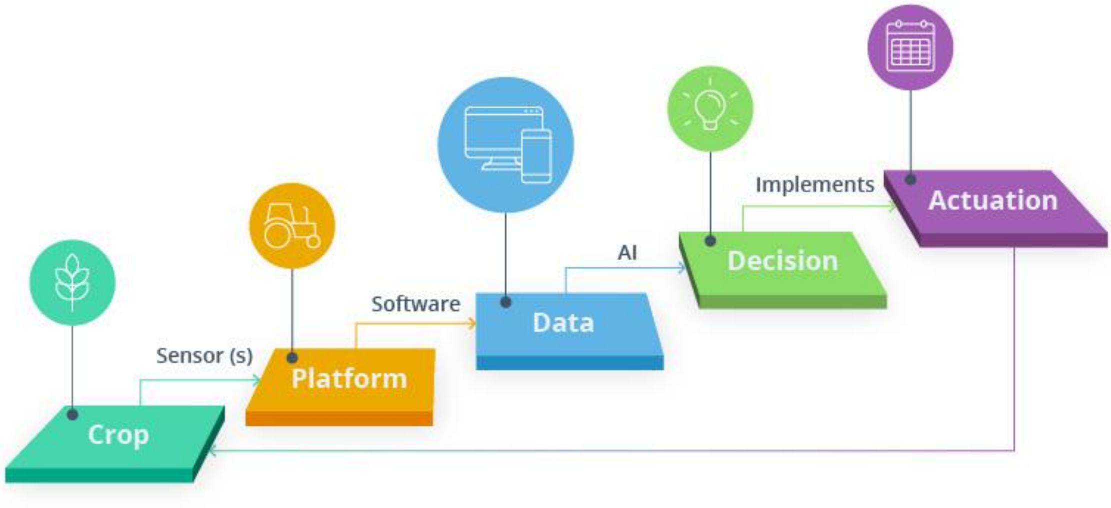
The AI revolution dictates the way of life it changes how we go to work, how we socialize with each other etc. just because of the introduction of Internet of things and Internet of systems. Implementation of AI throughout human technology could introduce many wonderous systems like autonomous cars, unbiased trials at the court etc.[180]

---

[179] Cp. *No Author*., https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/ accessed August 7, 2022.

[180] Cp. *No Author*., https://medium.datadriveninvestor.com/evolution-of-ai-past-present-future-6f995d5f964a accessed August 7, 2022.

Figure 18: The role of AI in the agriculture information management cycle



Source: https://www.mdpi.com/2073-4395/10/2/207

Agriculture Industry is one of the many industries to contribute and implement Artificial Intelligence. The use of this technology has been growing rapidly. As a result, AI is being deployed in a variety of agriculture sectors. Some of the sectors where AI is transforming the agriculture industry are the following:

1. **Crop and Soil Monitoring**

The kind of soil and nutrition of the soil have a significant impact on the crops that are grown and their quality. The quality of the soil is deteriorating as a result of growing deforestation, making it difficult to assess.[181]

A German-based tech start-up PEAT has developed an AI-based application called Plantix that can identify the nutrient deficiencies in soil including plant pests and diseases by which farmers can also get an idea to use fertilizer which helps to improve harvest quality. This app uses image recognition-based technology.

By saving time and enabling farmers to grow the greatest crop for the season, AI technologies assist farmers in analyzing land, soil, crop health, and other factors. Vertical farming may save water, use land more effectively, and grow crops inside of buildings in metropolitan settings. It can lessen the issues with labor shortages. predicts crop seasons,

[181] Cp. *SMT M SAMPTKUMARI BANDAGAR*., CROP AND SOIL MONITORING, 2022, pp. 2456.

weather, climate, and other factors for the following year. AI-based forecasts make it possible to recommend the right pesticides, crops, and locations at the right time before a disease spreads widely.[182]

2. Weed Management

Weed control, which is the main problem in pest control, is one of a farmer's greatest challenges. The use of herbicides outpaces the usage of all other types of pesticides, including insecticides. Herbicide application is something no farmer loves to do, but neither does seeing the weeds consume all the water and nutrients meant for the crops. The use of herbicides poses a serious threat to both the environment and human health.

Numerous businesses and initiatives aspire to use computer visualization, robotics, and machine learning. Advanced AI techniques have been developed to manage weeds precisely and effectively, reducing the need for herbicides. Pasqual created a sophisticated system for identifying and eliminating undesirable plants in crops including wheat, barley, and oats.[183] Burks also used the same set of contributions from the previous article to match three different neural networks, primarily backpropagation, counter dissemination, and radial based function models, and found that the backpropagation setup performs at its best with 97% accuracy.[184]

3. Pest Management

The most unpleasant farming issue that ultimately results in significant financial loss is infested pest infestation.[185] Through the development of computer systems that can classify aggressive pests and so recommend management strategies, scientists have been working for decades to minimize this menace.

Artificial intelligence is used extensively by pest treatment businesses to program and improve anything from the pest route plan to pest prediction. With the aid of drone technology, pest control businesses and farmers can virtually inspect all of their crops and provide nearly constant monitoring in an effort to discover pests, diseases, dead soil, or

---

[182] Cp. *Ibid*.
[183] Cp. *G. M. Pasqual*., Control of Weeds, 1994, pp. 125.
[184] Cp. *T. F. Burks et al*., Neural-Network Classifiers, 2005, pp. 296.
[185] Cp. *Krishnayan Dasgupta and Indrajit Ghosh.*, AI in Agriculture, no page number.

unusual crop deterioration.[186] Based on this knowledge, the farmer can collect data from any specific agricultural region and so prevent the disease from spreading further.

4. Disease Detection

Plant diseases pose a significant threat to the global economy, the environment, consumers, and farmers. Pests and diseases alone kill 35% of the crops in India, causing a significant cost to the farmers. The use of pesticides without selection poses a risk to human health because some of them are poisonous and biomagnified. By keeping an eye on the crop, spotting the disease, and administering the appropriate treatments, these impacts can be avoided.[187] To identify an indisposed plant and then take the necessary steps for its recovery, you need to have a lot of experience and competence. Worldwide, computerized tools are used to analyze the condition and then suggest ways to control it.

Sensing and image analysis are done to make sure that images of the leaves are divided into external sections like the non-diseased area, background, and diseased portion of the leaf in order to detect disease. The affected leaf section is then harvested and taken to the lab for further examination. This aids in the identification of pests and the subsequent detection of nutritional insufficiency.

5. Yield Prediction

For agricultural cost estimation as well as marketing strategies, crop yield forecast is quite helpful. Additionally, in the era of precision agronomy, prediction models can be used to examine relevant features that definitely affect the yield. A new ecosystem for sustainable, effective, and precision farming is evolving with the rise of innovative techniques such as artificial intelligence, satellite imagery, cloud machine learning, and predictive analysis. The combination of these cutting-edge technologies is assisting farmers in achieving the highest average yield possible as well as improved control over the quantity of food grains, ensuring they continue to make a profit. To predict harvest from topsoil restrictions, Liu and Minzan used an AI neural network model involving a backpropagation learning algorithm.[188]

---

[186] Cp. *KMICE'06 Proceedings Committee. and Universiti Utara Malaysia*., Proceedings of Knowledge, 2006, pp. 334.
[187] Cp. *Sukhvir Kaur, Shreelekha Pandey, and Shivani Goel*., Plants Disease, 2019, pp. 512.
[188] Cp. *Gang Liu, Xuehong Yang, and Minzan Li*., Crop Yield, 2005, pp. 1017.

Statistical climate information, the best time to plant in each season, and current time Sufficient Moisture the AI algorithms can be used to identify data from routine raindrop statistics and soil moisture to provide forecasts and give farmers suggestions on the ideal sowing window.[189]

6. Agricultural Robots

High-tech farming is not a contradiction. With applications that regulate irrigation, GPS systems that drive tractors, and RFID-chipped ear tags that track cattle, a modern agricultural enterprise is more like Silicon Valley than American Gothic.

And a vital component of that technological foundation is robots.

Robots harvest lettuce, gather strawberries, pick apples, and pull weeds. Drones collect airborne photographs that enable farmers to immediately evaluate the health of their crops. And in the backyards of densely populated metropolitan markets, robotic greenhouses are emerging thousands of kilometers from traditional farming zones, growing produce.

All of this occurs at a time when producers are dealing with an expensive, long-term labor scarcity and when food consumption is anticipated to increase dramatically due to the projected increase in global population from 7.7 billion to 9.7 billion people in less than 30 years.[190]

Types of agricultural robots:

1. Crop-Harvesting Robots
2. Weeding Robots
3. Robotics Greenhouse/Robotics Farming
4. Aerial Imagery Drones and Seed Planting Drones

[189] Cp. *Javad Ansarifar, Lizhi Wang, and Sotirios V. Archontoulis*., Regression Model, 2021, pp. 4.
[190] Cp. *No Author*., https://builtin.com/robotics/farming-agricultural-robots accessed August 7, 2022.

4.2. Case Point in Crop Disease Detection using EfficientNetB3 Model.

4.2.1 Model Architecture

The crop leaf disease detection and classification model architecture consist of 5 phases for detecting crop leaf diseases and classifying them based on their visual symptoms.[191]

1. **Image acquisition**

   This is the initial stage of identifying and classifying crop leaf diseases. This stage's goal is to gather and arrange the image dataset that will be used in the following steps. This is accomplished by taking pictures using cameras on mobile phones, digital cameras, drones, and UAVs either in real time (on location) or under controlled circumstances.

2. **Image preprocessing**

   To get better results, image preparation is essential. Color modifications were applied to reduce noise. Resizing techniques were employed to minimize the size of the image captured by a digital camera. It also aids in memory size reduction. In this literature, cutting the leaves from the captured image, color modifications, rescaling, background removal, image enhancing, flipping, rotating, Shear, and image smoothing are some of the often-used image preprocessing methods.

3. **Image segmentation**

   For the identification and categorization of crop leaf diseases, image segmentation is crucial. It divides the image into different regions or portions. To gather useful data for feature extraction, it examines the image data. There are two methods for segmenting images: one uses similarity and the other uses discontinuities.

4. **Feature extraction**

   The process of extracting features from an image's constituent parts is known as feature extraction. Shape, color, and texture are the most typical characteristics used in the classification and diagnosis of plant diseases. Due to many classes, crop diseases can have different visual appearances. The crop leaf disease system

---

[191] Cp. *Pallepati Vasavi, Arumugam Punitha, and T. Venkat Narayana Rao*., Crop Leaf Disease Detection, 2022, pp. 2079.

can quickly identify illnesses based on the crop leaf image's shape. The importance of color is the second feature. It sets apart the various crop leaf diseases. The final characteristic, texture, shows how different color patterns can be seen in cropped photos of leaves. Energy, entropy, contrast, correlation, sum of squares, sum entropy, cluster shadow, cluster prominence, and homogeneity are frequent texture characteristics.

5. **Classification**

   Crop leaf diseases were categorized using the ML and DL classification algorithm, respectively. The key distinction between deep learning techniques and conventional machine learning is made via feature extraction. Traditional machine learning (ML) does not calculate features automatically, although in deep learning (DL), features are extracted automatically and are used as learning weights. Therefore, with DL, the system itself picks up the necessary traits by receiving enough data. KNN, SVM, DT, RF, BPNN, NN, NB, and ensemble learning are the most frequently applied machine learning algorithms for classifying plant diseases. In the research, CNN and CNN models that incorporated transfer learning and were pre-trained on ImageNet were the most widely used deep learning algorithms.

4.2.2 Data Exploration

After loading all the necessary packages (refer to Appendix 1). The data injection was done with the following code:

Figure 19: Data Injection

```
directory = 'Crop___Disease'
```

Source: Author Own Representation

After the injection of database, we need to get the glimpse of data, the following code was used:

Figure 20: List of all the Directory

```
import os
print(os.listdir(directory))
```

```
['Corn', 'Potato', 'Rice', 'Wheat']
```

Source: Author Own Representation

After the code was run the output says that the data set has 4 different directories. Next, to get the gist of the data the following code was used for to know number of files and classes are there in each directory.

Figure 21: Glimpse of the data set.

```
BATCH_SIZE = 32
IMG_SIZE = (256, 256)
```

Source: Author Own Representation

After the code was run in corn directory found 3852 files belonging to 4 classes, potato directory found 2152 files belonging to 3 classes, rice directory found 4078 files belonging to 4 classes and wheat directory found 2942 files belonging to 3 classes.

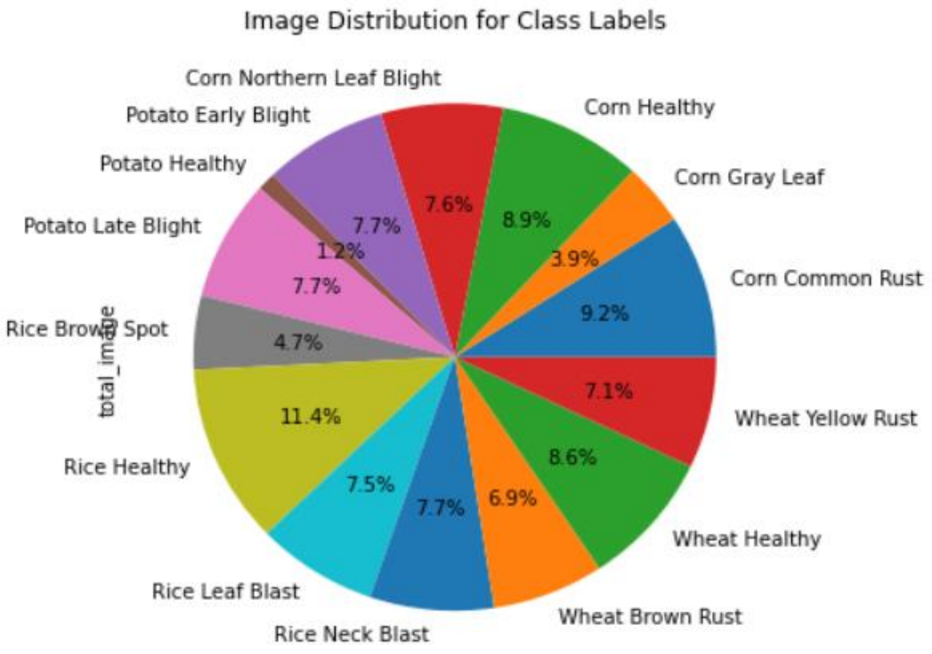
Figure 22: List of Category

Category	total_image
Corn Common Rust	1192
Corn Gray Leaf	513
Corn Healthy	1162
Corn Northern Leaf Blight	985
Potato Early Blight	1000
Potato Healthy	152
Potato Late Blight	1000
Rice Brown Spot	613
Rice Healthy	1488
Rice Leaf Blast	977
Rice Neck Blast	1000
Wheat Brown Rust	902
Wheat Healthy	1116
Wheat Yellow Rust	924

Source: Author Own Representation

The above diagram represents that there are 14 types of different class labels available into this dataset.

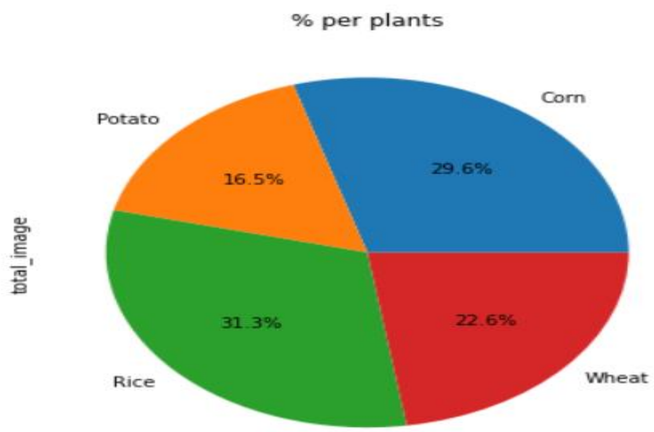
Figure 23: Image Distribution for Class Labels



Source: Author Own Representation

The above graph shows that the image distribution for class labels with total number of percentages. So, in the above graph potato healthy images was very less compared to other crops healthy image.

Figure 24: Image Percentage per Crops

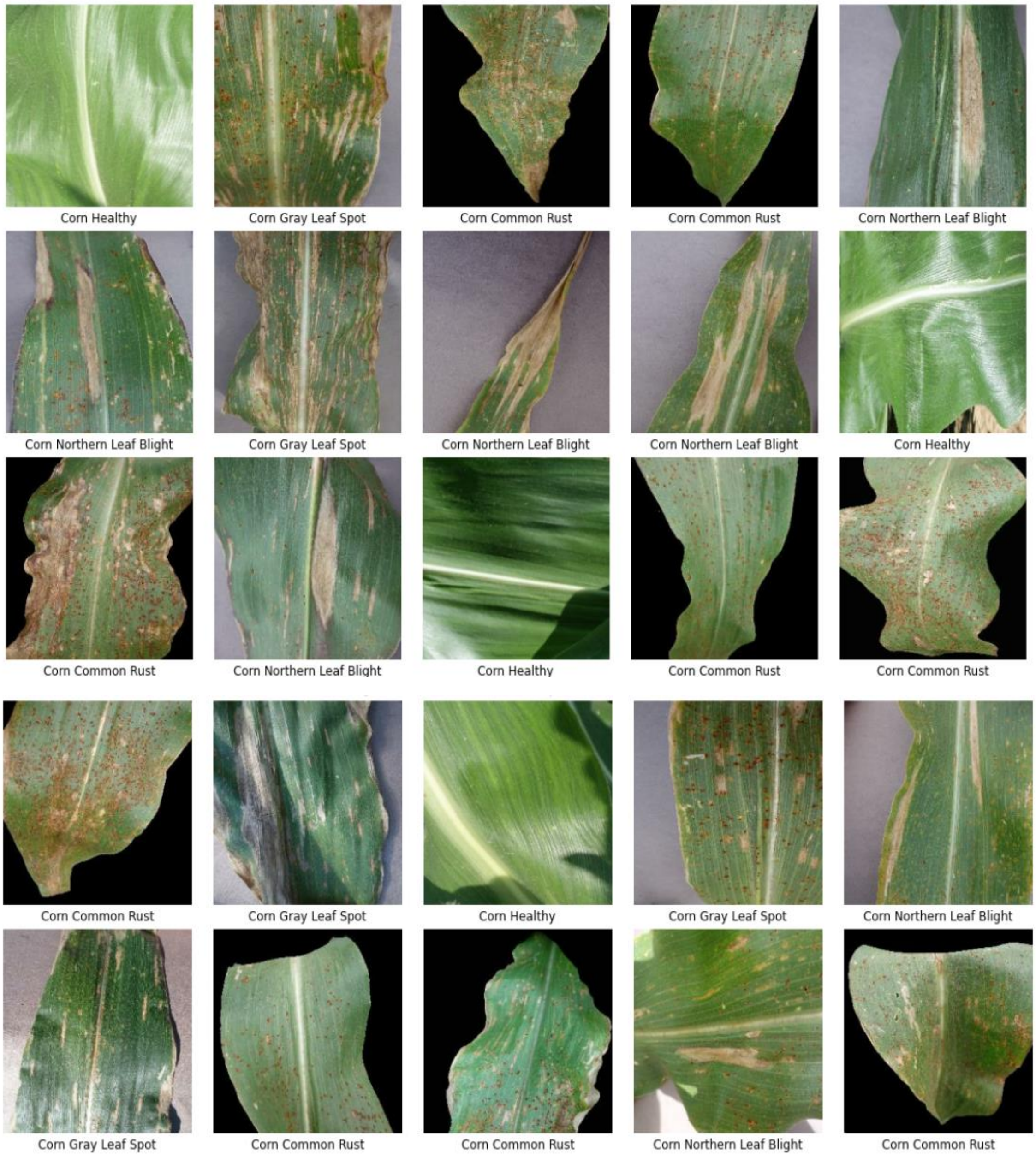


Source: Author Own Representation

The above graph shows that a total number of percentages per plants. So, in that graph rice image percentage was high so we can accurate more disease for rice plant compared to other plants.

Next, run code for display all the class labels images for all the different crops.

Figure 25: Corn Leaf Images



Source: Author Own Representation

The above display images shows that a different 4 types class labels(Corn Common Rust, Corn Gray Leaf Spot, Corn Healthy, Corn Northern Leaf Blight) of corn leaf images.

Figure 26: Potato Leaf Images

Source: Author Own Representation

The above display images shows that a different 3 types class labels(Potato Early Blight, Potato Healthy, Potato Late Blight) of potato leaf images.

Figure 27: Rice Leaf Images

Source: Author Own Representation

The above display images shows that a different 4 types class labels(Rice Brown Spot, Rice Healthy, Rice Leaf Blast, Rice Neck Blast) of rice leaf images.

Figure 28: Wheat Leaf Images

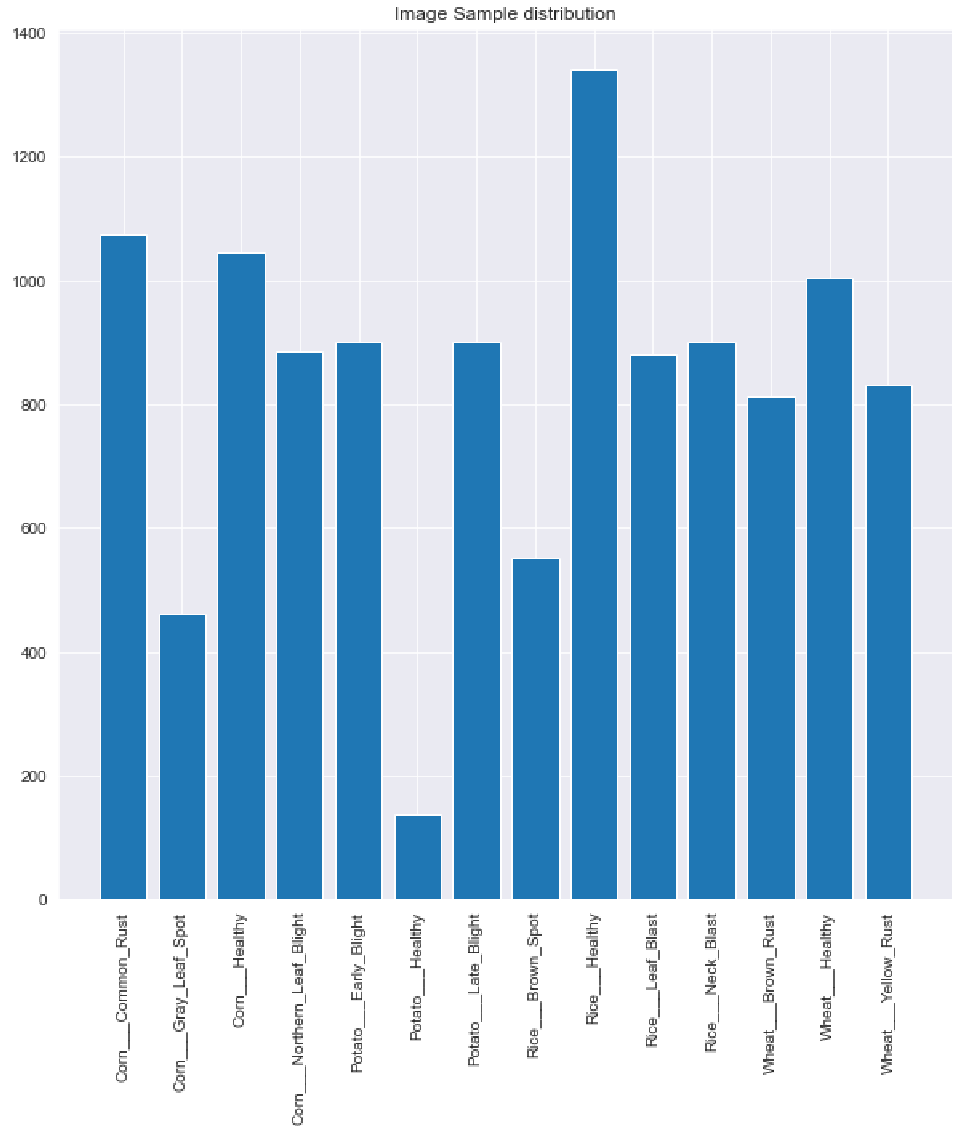
Source: Author Own Representation

The above display images shows that a different 3 types class labels(Wheat Brown Rust, Wheat Healthy, Wheat Yellow Rust) of wheat leaf images.

4.2.3 Data Modelling, Evaluation and Deployment

After performing the data exploratory analysis continue the process for data modelling, evaluation, and deployment. For that first read in images create a dataframe of image paths and class labels.

Figure 29: Image Sample Distribution



Source: Author Own Representation

The above graph shows that Rice_Healthy has the most images is1339 and Potato_Healthy has the least images is 137 so average height is 578, average width is 566 and aspect ratio is1.0210814148363765.

Next, each class has more than 150 images so used trim function for train_df dataframe so no class has more than 200 image samples.

Figure 30: Trim Function

```
def trim(df, max_samples, min_samples, column):
    df=df.copy()
    groups=df.groupby(column)
    trimmed_df = pd.DataFrame(columns = df.columns)
    groups=df.groupby(column)
    for label in df[column].unique():
        group=groups.get_group(label)
        count=len(group)
        if count > max_samples:
            sampled_group=group.sample(n=max_samples, random_state=123,axis=0)
            trimmed_df=pd.concat([trimmed_df, sampled_group], axis=0)
        else:
            if count>=min_samples:
                sampled_group=group
                trimmed_df=pd.concat([trimmed_df, sampled_group], axis=0)
    print('after trimming, the maximum samples in any class is now ',max_samples, ' and the minimum samples in any class is ', mi
    return trimmed_df
```

Source: Author Own Representation

After running the above code output was after trimming the maximum sample in any class is 200 images and the minimum sample in any class is 137 images. After the apply trim function, we need to balance the train_df dataframe (Refer to Appendix 1 for the full code). The balance function takes a dataframe as an input. It generates and saves augmented images for each class so that each class will have the same number of images specified by the value n. The input dataframe is concatenated with the augmented images filepaths to provide a balanced dataset.

Figure 31: Output of balance() function

```
Initial length of dataframe is  2737
 Augmented images already exist. To delete these and create new images enter D, else enter U to use these images
 D
Found 137 validated image filenames.     for class        Potato___Healthy        creating  63    augmented images
Total Augmented images created=  63
Length of augmented dataframe is now  2800
```

Source: Author Own Representation

After running the balance() function output was the above image. It's shows that the initial length of dataframe is 2737. And some of augmented images is already exist in dataset so enter D for delete those images and create new augmented images or else enter U for those images. So, total created augmented images was 63 so now length of augmented dataframe is 2800.

Next, create the 4 generators train_gen, test_gen, final_test_gen and valid_gen.

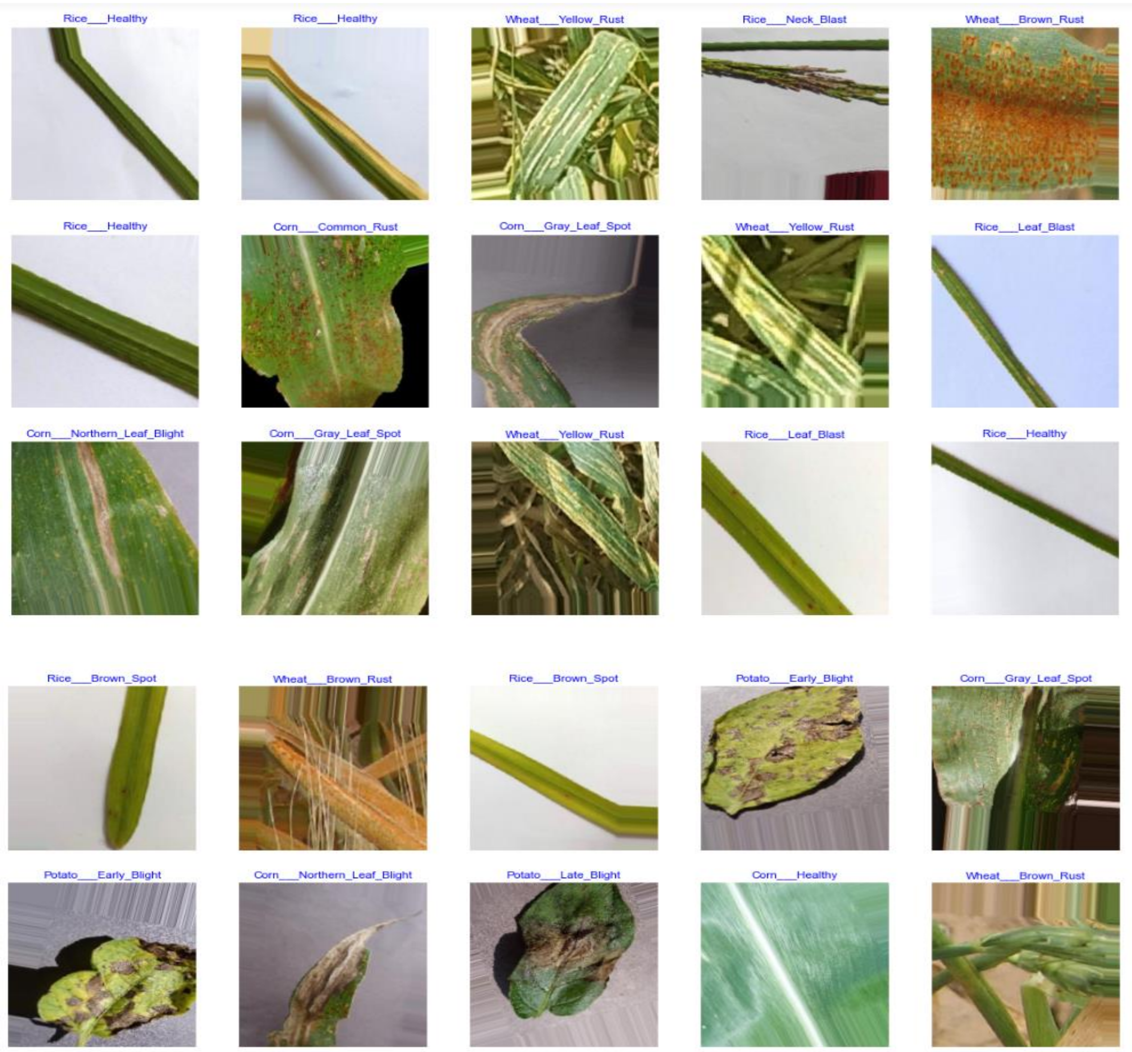
Figure 32: Output for generator

```
Found 2800 validated image filenames belonging to 14 classes.        for train generator
Found 651 validated image filenames belonging to 14 classes.         for valid generator
Found 652 validated image filenames belonging to 14 classes.         for test generator
test batch size:  4   test steps:  163  number of classes :  14
```

Source: Author Own Representation

After running the code above output came it's shows found 2800 validated image filenames for train generator, 651validated image filenames for valid generator and 652 validated image filenames for test generator all files belonging to 14 classes. For that the test batch size was 4 and test steps was 163. After that create a function to show example training images.

Figure 33: Sample images for train generator



Source: Author Own Representation

The above image shows a sample image for train generator with all classes. Now create a model using transfer learning with EfficientNetB3 (Refer to Appendix 1 for the full code).

After that create a custom Keras callback to continue and optionally set LR or halt training. The LR_ASK callback is a convenient callback that allows you to continue training for ask_epoch more epochs or to halt training. If you select to continue training for more epochs, you are given the option to retain the current learning rate (LR) or to enter a new value for the learning rate. The form of use is: ask=LR_ASK(model, epochs, ask_epoch)

Where:

- Model is a string which is the name of your compiled model.
- Epochs is an integer which is the number of epochs to run specified in model.fit
- ask_epoch is an integer. If ask_epoch is set to a value say 5 then the model will train for 5 epochs. Then the user is asked to enter H to halt training or enter an inter value. For example, if you enter 4 training will continue for 4 more epochs to epoch 9 then you will be queried again. Once you enter an integer value you are prompted to press ENTER to continue training using the current learning rate or to enter a new value for the learning rate.

  At the end of training the model weights are set to the weights for the epoch that achieved the lowest validation loss.

Figure 34: Instantiate custom callback

```
epochs=40
ask_epoch=5
ask=LR_ASK(model, epochs,  ask_epoch)
#rlronp=tf.keras.callbacks.ReduceLROnPlateau(monitor="val_loss", factor=0.5, patience=2,verbose=1)
#callbacks=[rlronp, ask]
callbacks=[ask]
```

Source: Author Own Representation

After that like above image instantiate custom callback. Next, train the model (Refer to Appendix for the full code).

Figure 35: Training Model Result

```
Training will proceed until epoch 5  then you will be asked to
 enter H to halt training or enter an integer for how many more epochs to run then be asked again
Epoch 1/40
94/94 [==============================] - 127s 1s/step - loss: 8.0532 - accuracy: 0.7004 - val_loss: 7.0905 - val_accuracy:
0.8203

 validation loss reduced, saving weights from epoch  1  as best weights
Epoch 2/40
94/94 [==============================] - 90s 955ms/step - loss: 6.0073 - accuracy: 0.8575 - val_loss: 5.3451 - val_accurac
y: 0.8986

 validation loss reduced, saving weights from epoch  2  as best weights
Epoch 3/40
94/94 [==============================] - 88s 942ms/step - loss: 4.7348 - accuracy: 0.9129 - val_loss: 4.2196 - val_accurac
y: 0.9232

 validation loss reduced, saving weights from epoch  3  as best weights
Epoch 4/40
94/94 [==============================] - 88s 932ms/step - loss: 3.7879 - accuracy: 0.9204 - val_loss: 3.3057 - val_accurac
y: 0.9401

 validation loss reduced, saving weights from epoch  4  as best weights
Epoch 5/40
94/94 [==============================] - 88s 936ms/step - loss: 3.0175 - accuracy: 0.9314 - val_loss: 2.6296 - val_accurac
y: 0.9447

 validation loss reduced, saving weights from epoch  5  as best weights
```

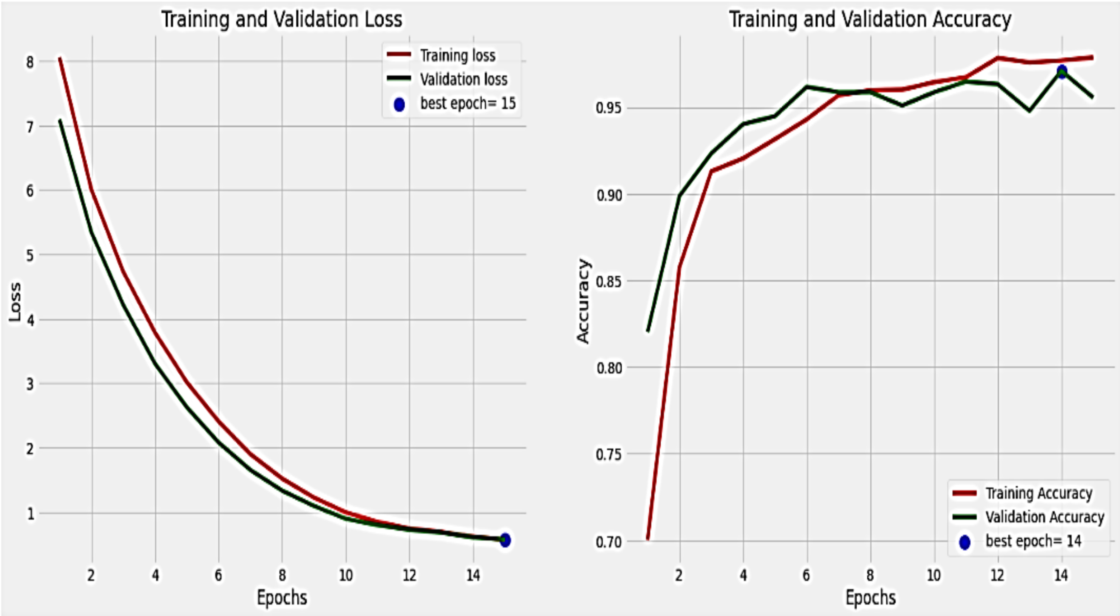
Source: Author Own Representation

The above image shows the results of training model. Here, training will proceed until epoch 5 after that you will be asked to enter H to halt training or enter an integer for how many more epochs to run then be asked again. After running up to 5 epoch model accuracy is 0.9447 and loss is 2.6296. So, as training progresses in epoch, the accuracy of the model increases and the loss decreases. So, at last model accuracy was 95%.

4.2.4 Results and Findings

Upon investigating the data set, the data set seemed to be unbalanced which was confirmed by running it for data exploration. After visualizing the features and gaining an understanding of the relationship between the features one predictive model were investigated. The data was split into training set, test set and validation set. Now, define a function (Refer to Appendix for the full code) to plot the training data.

Figure 36: Plot the training data



Source: Author Own Representation

The training data is shown in the graph above. In the first graph, it is shown that when 15 epochs are run, the training loss is 0.5726 and validation loss is 0.5773 become less, and training accuracy of the model is 0.9786 and the validation accuracy of the model is 0.9555.

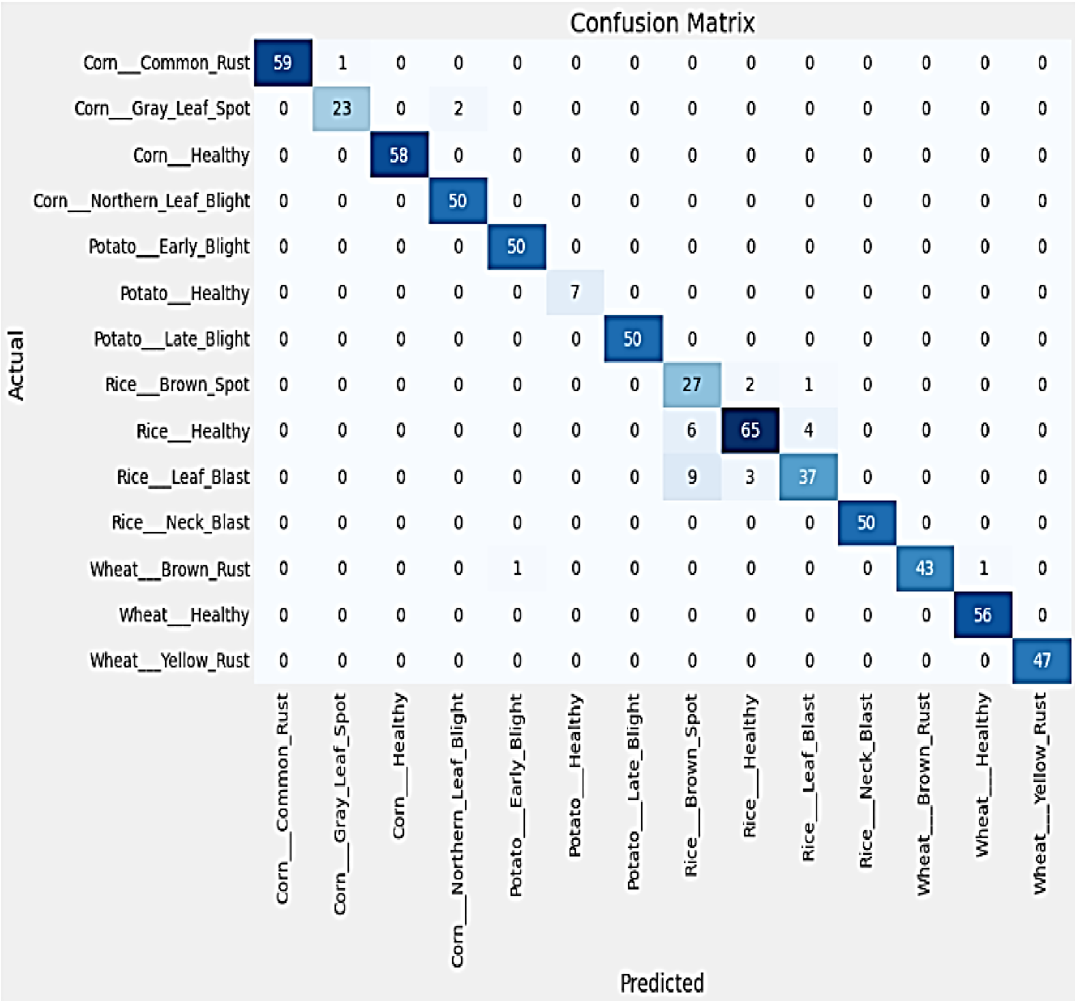
In the second graph, it is shown that when 14 epochs are run, the training accuracy is 0.9768 and validation accuracy is 0.9708 become high, and training loss of the model is 0.6192 and the validation loss of the model is 0.6106.

So, model accuracy is 95% because in 15 epoch loss was decreased.

Next, make prediction on the test set. Define a function (Refer to Appendix for the full code) which takes in a test generator and an integer test_steps and generates predictions on the test set including a confusion matrix and a classification report.

Figure 37: Confusion Matrix

```
163/163 [==============================] - 17s 90ms/step
there were 30 errors in 652 tests for an accuracy of  95.40
```



Source: Author Own Representation

With the help of above confusion matrix, we can know the error rate of the model. In the above confusion matrix, there are 30 errors in 652 tests so because of this error rate the model accuracy is 95%. In this matrix it's shows that this model predicted 9 images in rice category which is Rice_Leaf_Blast as Rice_Brown_Spot and 6 images in rice category which is Rice_Healthy as Rice_Brown_Spot.

Figure 38: Classification Report

Classification Report:

----

	precision	recall	f1-score	support
Corn___Common_Rust	1.0000	0.9833	0.9916	60
Corn___Gray_Leaf_Spot	0.9583	0.9200	0.9388	25
Corn___Healthy	1.0000	1.0000	1.0000	58
Corn___Northern_Leaf_Blight	0.9615	1.0000	0.9804	50
Potato___Early_Blight	0.9804	1.0000	0.9901	50
Potato___Healthy	1.0000	1.0000	1.0000	7
Potato___Late_Blight	1.0000	1.0000	1.0000	50
Rice___Brown_Spot	0.6429	0.9000	0.7500	30
Rice___Healthy	0.9286	0.8667	0.8966	75
Rice___Leaf_Blast	0.8810	0.7551	0.8132	49
Rice___Neck_Blast	1.0000	1.0000	1.0000	50
Wheat___Brown_Rust	1.0000	0.9556	0.9773	45
Wheat___Healthy	0.9825	1.0000	0.9912	56
Wheat___Yellow_Rust	1.0000	1.0000	1.0000	47
accuracy			0.9540	652
macro avg	0.9525	0.9558	0.9521	652
weighted avg	0.9588	0.9540	0.9548	652

Source: Author Own Representation

With the help of above classification report, we obtain the value of precision, recall, f1-score and accuracy. In this report precision, recall and f1-score value is high for 14 class labels which means that this model is very accurate. And our model accuracy is 0.9540 which means model was 95% accurate.

5. Research Results

The below sections explain how this paper answers the individual given research questions:

RQ1. How can we detect crop disease using AI?

We are loading the dataset which contains images for training purposes for the Efficient B3 convolution Network. In the training phase, we will only train the top layer of images and freeze the pre-trained models. Kera's Application deep learning models have pretrained weights done in ImageNet, which is already trained, this gives us better feature extraction and fine-tuning of the model. Since the pre-trained model was trained on ImageNet and not on crop images, there are some limitations to this approach. We countered it by adding a second training phase.

After each epoch, we save the model if it is better than the previous one, according to the Quadratic Weighted Kappa score on the validation set. We also monitor the Mean Squared Error and stop training if it does not improve for 2 epochs. This way we can counter overfitting by monitoring Quadratic Weighted Kappa and Mean Squared Error.

After using the training set of images, we used test images to check the disease classification. In this training phase, we are unfreezing all the layers to get fine tune model if the images have spots which that model spotted, then it will be considered an unhealthy image and if the image has no spots, then the image is healthy, upon this we also gave feature extraction for the model. This feature extraction will provide us with the name tag for the respective image whether it is healthy or not.

RQ2. How can we predict the yield of crops based on the disease?

The test data contains 652 images, and it contains Corn, Potato, Rice, and wheat images. In the entire data set, 452 images are unhealthy, which was detected by the algorithm.

Name	Healthy	Unhealthy
Corn	58	135
Potato	7	100
Rice	75	129
Wheat	56	92

- population size: 65200
- confidence limit: 95%
- the margin of error: 4%
- sample size 652 around 600

In this dataset there are total 652 images so in those 452 images are unhealthy. so, 69.9% of leaves or pictures in the sample size is unhealthy, which shows that 70% of the yield will be reduced from last year.

Apart from these images, we need soil data, yield performance data, and weather data, which contains precipitation and solar radiation with maximum and minimum temperature for the crop time. We also need soil variables which include dry bulk density, percentage of clay, pH of the soil, drained upper limit, saturated soil hydraulic conductivity, witling point, soil organic matter, sand percentage, and water content in soil at different depths.

If we compute disease formed on images according to the data mentioned above points, then we will get an accurate yield prediction for the crops. Only with the disease pictures of crops can we predict how much crop is damaged but not the yield growth of the crop.

6. Limitation & Conclusion

Conclusion:

The main goal of the current study is to identify behavioral factors influencing the transformation of AI system in agriculture based on a theoretical framework.

The objective of this paper was the AI can assist farmers in increasing productivity while lowering production costs and labor-intensive tasks. It goes without saying that as AI becomes more prevalent across all application domains, our current approach to agricultural research and development will undergo a desirable change. Precision farming is becoming more common in traditional agriculture because of AI's movement toward greater automation and accuracy in real-time management. The farming community must be able to use and benefit from the AI solution. AI companies should offer an open-source platform by making their solutions more affordable to promote quicker adoption and deeper insight among farmers.

Real-time data monitoring has demonstrated the value of artificial intelligence. This has been put in place to manage yield and crops. Machines inspect crop leaves for illness and determine whether they are healthy or not. Based on that and some other additional soil variables which include dry bulk density, percentage of clay, pH of the soil, drained upper limit, saturated soil hydraulic conductivity, witling point, soil organic matter, sand percentage, and water content in soil farmers can predict yield of the crop. The effective techniques can guarantee farmers proper field management and healthier harvests. The AI delivers timely information through the appropriate channels, which might help users become more resilient.

This paper explores how artificial intelligence has been applied to farming over the years, commencing in 1983. The purpose of this paper was to present as much information as possible regarding the various AI techniques applied to agriculture. From the 1980s through the 1990s, expert systems based on rules were widely used, but starting in 1990, fuzzy inference systems and artificial neural networks assumed a dominant position. Artificial neural networks are currently used in combination with hybrid systems, such as neuro-fuzzy or image processing.

By making better decisions in the field, the use of AI will help farmers achieve their goal of a healthier harvest. Since food is the most important human need, the power of data may be used more creatively to anticipate risk, analyze possibilities, and act before hunger increases into a humanitarian crisis. This will aid in the global development of the entire planet.

Every business in the agriculture industry could bring in AI and ML into its workflow and improve its business efficiency and bring in new ways of adapting these new technologies into the work culture. To conclude, every business as an organization should learn to adapt and work to make sure they are using the latest technologies to gain an upper hand in the competitive market. Although, adopting AI could reduce work stress and bring in worklife balance into the working environment one should also keep themselves up to date and adapt to newer technologies and learn them faster to retain their employment status.

Limitation:

Although, there are several advantages to the AI in Agriculture Industry, there also is a downside to it. Some of them are,

- The most common limitations of crop leaf disease prediction system based on visual symptoms are Lack of sufficient datasets mentioned in the papers since the PlantVillage is the only public dataset that is generated under controlled conditions.
- Some of the authors developed their own dataset but they are not giving the access to others to compare the results and PlantVillage dataset could not provide the images for the commercial crops like chili with abundant number of diseases.
- Fourthly, the prediction system needs enormous resources if the prediction was based on deep learning methodologies.
- So, there is a significance to develop squeeze models to run the application in mobile phones, drones, UAVs, and robots.

Future Scope:

Establishing real-time, enormous photos and classes of plant diseases will be a common task for future development. In order to evaluate crop and yield monitoring in support of intelligent agriculture, crop disease datasets can be combined to include location, weather, and soil data of the afflicted plant. For the purpose of identifying plant illnesses in expansive horticultural fields, the crop disease prediction system can be improved.

Bibliography

Abdullah, Mohammad Fikry, and Kamsuriah Ahmad. "The Mapping Process of Unstructured Data to Structured Data." *International Conference on Research and Innovation in Information Systems, ICRIIS*, 2013, 151–55. https://doi.org/10.1109/ICRIIS.2013.6716700.

"Agriculture: Definition and Overview," 2014. https://doi.org/10.1007/978-1-4419-0465-2_64.

Allua, Shane, Laura M. Stapleton, and S. Natasha Beretvas. "Testing Latent Mean Differences between Observed and Unobserved Groups Using Multilevel Factor Mixture Models." *Educational and Psychological Measurement* 68, no. 3 (2008): 357–78. https://doi.org/10.1177/0013164407312600.

Alzubaidi, Laith, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions." *Journal of Big Data 2021 8:1* 8, no. 1 (March 31, 2021): 1–74. https://doi.org/10.1186/S40537-021-00444-8.

Anderson, David R, Kenneth P Burnham, and William L Thompson. "Null Hypothesis Testing: Problems, Prevalence, and an Alternative." *The Journal of Wildlife Management* 64, no. 4 (2000): 912–23.

Ansarifar, Javad, Lizhi Wang, and Sotirios V. Archontoulis. "An Interaction Regression Model for Crop Yield Prediction." *Scientific Reports 2021 11:1* 11, no. 1 (September 7, 2021): 1–14. https://doi.org/10.1038/s41598-021-97221-7.

Apuke, Oberiri Destiny. "Quantitative Research Methods : A Synopsis Approach." *Kuwait Chapter of Arabian Journal of Business and Management Review* 6, no. 11 (September 2017): 40–47. https://doi.org/10.12816/0040336.

Arasu, Arvind, and Hector Garcia-Molina. "Extracting Structured Data from Web Pages," 2003.

Argyrous, George, and George Argyrous. "Statistics for Research : With a Guide to SPSS," 2005, 401.

Asemi, Adeleh, Asefeh Asemi, Andrea Ko, and Ali Alibeigi. "An Integrated Model for Evaluation of Big Data Challenges and Analytical Methods in Recommender Systems." *Journal of Big Data* 9, no. 1 (December 1, 2022): 1–26. https://doi.org/10.1186/S40537-022-00560-Z/FIGURES/15.

Auer, Sören, Simon Scerri, Aad Versteden, Erika Pauwels, Angelos Charalambidis, Stasinos Konstantopoulos, Jens Lehmann, et al. "The BigDataEurope Platform – Supporting the Variety Dimension of Big Data." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10360 LNCS (2017): 41–59. https://doi.org/10.1007/978-3-319-60131-1_3.

Ben-Kiki, Oren, Clark Evans, Brian Ingerson, and by Oren Ben-Kiki. "YAML Ain't Markup Language (YAML™) Version 1.1 Working Draft 2004-12-28 XSL • FO RenderX YAML Ain't Markup Language (YAML™) Version 1.1 Working Draft 2004-12-28," 2001. http://www.unicode.org/],.

Bourhis, Pierre, Juan L. Reutter, Fernando Suárez, and Domagoj Vrgoč. "JSON: Data Model, Query Languages and Schema Specification." *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems* Part F127745 (May 9, 2017): 123–35. https://doi.org/10.1145/3034786.3056120.

Buneman, Peter, Susan Davidson, Mary Fernandez, and Dan Suciu. "Adding Structure to Unstructured Data." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 1186 (1996): 336–50. https://doi.org/10.1007/3-540-62222-5_55.

Burks, T. F., S. A. Shearer, J. R. Heath, and K. D. Donohue. "Evaluation of Neural-Network Classifiers for Weed Species Discrimination." *Biosystems Engineering* 91, no. 3 (July 1, 2005): 293–304. https://doi.org/10.1016/J.BIOSYSTEMSENG.2004.12.012.

Caffery, Liam J., David Clunie, Clara Curiel-Lewandrowski, Josep Malvehy, H. Peter Soyer, and Allan C. Halpern. "Transforming Dermatologic Imaging for the Digital Era: Metadata and Standards." *Journal of Digital Imaging* 31, no. 4 (August 1, 2018): 568–77. https://doi.org/10.1007/S10278-017-0045-8/FIGURES/2.

Canton-Croda, Rosa Maria, and Damian Emilio Gibaja-Romero. “Metamodels and Category Theory in the Transformation of Semi-Structured Data.” *Proceedings - 2020 8th Edition of the International Conference in Software Engineering Research and Innovation, CONISOFT 2020*, November 1, 2020, 117–23. https://doi.org/10.1109/CONISOFT50191.2020.00026.

Chardonnens, Thibaud, Philippe Cudre-Mauroux, Martin Grund, and Benoit Perroud. “Big Data Analytics on High Velocity Streams: A Case Study.” *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013*, 2013, 784–87. https://doi.org/10.1109/BIGDATA.2013.6691653.

Chavas, Jean Paul, and Céline Nauges. “Uncertainty, Learning, and Technology Adoption in Agriculture.” *Applied Economic Perspectives and Policy* 42, no. 1 (March 1, 2020): 42–53. https://doi.org/10.1002/AEPP.13003.

Chowdhury, Md Rayhan, and Shahadat Hossain. “Analysis of Genetic Variability, Heritability and Correlation among Bangladeshi Local Rice Cultivars (Oryza Sativa L.) Project Paper View Project Molecular Detection and Biological Control of Human Hair Dandruff View Project.” *Article in Pakistan Journal of Agricultural Research* 34, no. 2 (2021). https://doi.org/10.17582/journal.pjar/2021/34.2.356.361.

“Criteria for a Good Literature Review - Writing a Literature Review - Guides at Worcester Polytechnic Institute,” n.d.

Cuzzocrea, Alfredo, and Ernesto Damiani. “Privacy-Preserving Big Data Exchange: Models, Issues, Future Research Directions.” *Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021*, 2021, 5081–84. https://doi.org/10.1109/BIGDATA52589.2021.9671686.

Dasgupta, Krishnayan, and Indrajit Ghosh. “SEMINAR PROCEEDINGS OF UGC SPONSORED NATIONAL LEVEL SEMINAR ON AI & ITS IMPACT ON MODERN IT WORLD AI in Agriculture JUTE INSECT PEST IDENTIFICATION: AN EXPERT SYSTEM BASED APPROACH,” n.d.

Demirkan, Haluk, James C. Spohrer, and Jeffrey J. Welser. “Digital Innovation and Strategic Transformation.” *IT Professional* 18, no. 6 (November 1, 2016): 14–18. https://doi.org/10.1109/MITP.2016.115.

Dobrev, Dimiter. “A Definition of Artificial Intelligence,” October 3, 2012. https://doi.org/10.48550/arxiv.1210.1568.

Eberendu, Adanma Cecilia. “Unstructured Data: An Overview of the Data of Big Data Encouraging Female Students to Program View Project Unstructured Data: An Overview of the Data of Big Data.” *Article in International Journal of Emerging Trends & Technology in Computer Science* 38, no. 1 (2016). https://doi.org/10.14445/22312803/IJCTT-V38P109.

Ebert, Christof, Carlos Henrique, and Cabral Duarte. “Digital Transformation The Influence of Regulatory Requirements on ICT Businesses View Project,” 2018. https://doi.org/10.1109/MS.2018.2801537.

Eddelbuettel, Dirk, Murray Stokely, and Jeroen Ooms. “RProtoBuf: Efficient Cross-Language Data Serialization in R.” *Journal of Statistical Software* 71 (July 11, 2016): 1–24. https://doi.org/10.18637/JSS.V071.I02.

El-Ramady, Hassan R., Samia M. El-Marsafawy, and Lowell N. Lewis. “Sustainable Agriculture and Climate Changes in Egypt,” 2013, 41–95. https://doi.org/10.1007/978-94-007-5961-9_2.

Faroukhi, Abou Zakaria, Imane El Alaoui, Youssef Gahi, and Aouatif Amine. “Big Data Monetization throughout Big Data Value Chain: A Comprehensive Review.” Accessed May 28, 2022. https://doi.org/10.1186/s40537-019-0281-5.

Favaretto, Maddalena, Eva de Clercq, Christophe Olivier Schneble, and Bernice Simone Elger. “What Is Your Definition of Big Data? Researchers’ Understanding of the Phenomenon of the Decade.” *PLOS ONE* 15, no. 2 (2020): e0228987. https://doi.org/10.1371/JOURNAL.PONE.0228987.

Ghosh, Indrajit, Uditendu Sarkar, Gouravmoy Bannerjee, and Swarup Das. “Artificial Intelligence in Agriculture: A Literature Survey Artificial Intelligence in Agriculture: A Literature Survey View Project Site Specific Crop Recommendation View Project Artificial Intelligence in Agriculture: A Literature Survey.” *International Journal of Scientific Research in Computer Science Applications and Management Studies IJSRCSAMS* 7, no. 3 (2018). www.ijsrcsams.com.

Gorodov, Evgeniy Yur Evich, and Vasiliy Vasil Evich Gubarev. “Analytical Review of

Data Visualization Methods in Application to Big Data.” *Journal of Electrical and Computer Engineering* 2013 (January 1, 2013). https://doi.org/10.1155/2013/969458.

Gupta, Yogesh Kumar, and Surbhi Kumari. “A Study of Big Data Analytics Using Apache Spark with Python and Scala.” *Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020*, December 3, 2020, 471–78. https://doi.org/10.1109/ICISS49785.2020.9315863.

Haenlein, Michael, and Andreas Kaplan. “A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence:” *Https://Doi.Org/10.1177/0008125619864925* 61, no. 4 (July 17, 2019): 5–14. https://doi.org/10.1177/0008125619864925.

Hagberg, Johan, Malin Sundstrom, and Niklas Egels-Zandén. “The Digitalization of Retailing: An Exploratory Framework.” *International Journal of Retail and Distribution Management* 44, no. 7 (July 11, 2016): 694–712. https://doi.org/10.1108/IJRDM-09-2015-0140/FULL/PDF.

Hartley, Janet L., and William J. Sawaya. “Tortoise, Not the Hare: Digital Transformation of Supply Chain Business Processes.” *Business Horizons* 62, no. 6 (November 1, 2019): 707–15. https://doi.org/10.1016/J.BUSHOR.2019.07.006.

Hsu, Feng Hsiung. “IBM’s Deep Blue Chess Grandmaster Chips.” *IEEE Micro* 19, no. 2 (March 1999): 70–81. https://doi.org/10.1109/40.755469.

Hughes, David. P., and Marcel Salathe. “An Open Access Repository of Images on Plant Health to Enable the Development of Mobile Disease Diagnostics,” November 25, 2015. https://doi.org/10.48550/arxiv.1511.08060.

Jagadish, H. V., Johannes Gehrke, Alexandros Labrinidis, Yannis Papakonstantinou, Jignesh M. Patel, Raghu Ramakrishnan, and Cyrus Shahabi. “Big Data and Its Technical Challenges.” *Communications of the ACM* 57, no. 7 (2014): 86–94. https://doi.org/10.1145/2611567.

Jiang, Lin, and Zhijia Zhao. “JSONSki: Streaming Semi-Structured Data with Bit-Parallel Fast-Forwarding,” 2022, 12. https://doi.org/10.1145/3503222.3507719.

Kaul, Vivek, Sarah Enslin, and Seth A. Gross. “History of Artificial Intelligence in

Medicine." *Gastrointestinal Endoscopy* 92, no. 4 (October 1, 2020): 807–12. https://doi.org/10.1016/J.GIE.2020.06.040.

Kaur, Harkiran, and Aanchal Phutela. "Commentary upon Descriptive Data Analytics." *Proceedings of the 2nd International Conference on Inventive Systems and Control, ICISC 2018*, June 27, 2018, 678–83. https://doi.org/10.1109/ICISC.2018.8398884.

Kaur, Sukhvir, Shreelekha Pandey, and Shivani Goel. "Plants Disease Identification and Classification Through Leaf Images: A Survey." *Archives of Computational Methods in Engineering* 26, no. 2 (2019): 507–30. https://doi.org/10.1007/s11831-018-9255-6.

Keeney, Dennis. "Sustainable Agriculture: Definition and Concepts." *Journal of Production Agriculture* 3, no. 3 (July 1, 1990): 281–85. https://doi.org/10.2134/JPA1990.0281.

Khan, Nawsher, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani. "Big Data: Survey, Technologies, Opportunities, and Challenges." *Scientific World Journal* 2014 (2014). https://doi.org/10.1155/2014/712826.

Kitchin, Rob. "Big Data and Human Geography: Opportunities, Challenges and Risks." *Https://Doi.Org/10.1177/2043820613513388* 3, no. 3 (December 10, 2013): 262–67. https://doi.org/10.1177/2043820613513388.

KMICE'06 Proceedings Committee., and Universiti Utara Malaysia. "Proceedings of Knowledge Management International Conferencs & Exhibition (KMICE) 2006. Volume 1," 2006, 334.

Krishnan, Akshay, Shashank Swarna, and H. S. Balasubramanya. "Robotics, IoT, and AI in the Automation of Agricultural Industry: A Review." *Proceedings of B-HTC 2020 - 1st IEEE Bangalore Humanitarian Technology Conference*, October 8, 2020. https://doi.org/10.1109/B-HTC50970.2020.9297856.

Liu, Gang, Xuehong Yang, and Minzan Li. "An Artificial Neural Network Model for Crop Yield Responding to Soil Parameters." *Lecture Notes in Computer Science* 3498, no. III (2005): 1017–21. https://doi.org/10.1007/11427469_161/COVER.

López-Aymes, Gabriela, María de los Dolores Valadez, Elena Rodríguez-Naveiras, Doris

Castellanos-Simons, Triana Aguirre, and África Borges. “A Mixed Methods Research Study of Parental Perception of Physical Activity and Quality of Life of Children Under Home Lock Down in the COVID-19 Pandemic.” *Frontiers in Psychology* 12 (March 15, 2021): 680. https://doi.org/10.3389/FPSYG.2021.649481/BIBTEX.

Lu, Baochuan, and Pam Smallwood. “Front Range Community College Westminster, CO” 36 (2020).

MacK, Chris A. “Fifty Years of Moore’s Law.” *IEEE Transactions on Semiconductor Manufacturing* 24, no. 2 (May 2011): 202–7. https://doi.org/10.1109/TSM.2010.2096437.

Mauro, Andrea De, Marco Greco, and Michele Grimaldi. “A Formal Definition of Big Data Based on Its Essential Features.” *Library Review* 65, no. 3 (April 4, 2016): 122–35. https://doi.org/10.1108/LR-06-2015-0061/FULL/PDF.

———. “What Is Big Data? A Consensual Definition and a Review of Key Research Topics.” *AIP Conference Proceedings* 1644, no. 1 (February 17, 2015): 97. https://doi.org/10.1063/1.4907823.

Mayer-Schönberger, Viktor, and Kenneth Cukier. “Big Data: A Revolution That Will Transform How We Live, Work, and Think.” *American Journal of Epidemiology* 179, no. 9 (May 1, 2014): 1143–44. https://doi.org/10.1093/AJE/KWU085.

McNulty, Terry, and Ewan Ferlie. “Process Transformation: Limitations to Radical Organizational Change within Public Service Organizations:” *Http://Dx.Doi.Org/10.1177/0170840604046349* 25, no. 8 (June 30, 2016): 1389–1412. https://doi.org/10.1177/0170840604046349.

Minsky, Marvin. “Steps Toward Artificial Intelligence.” *Proceedings of the IRE* 49, no. 1 (1961): 8–30. https://doi.org/10.1109/JRPROC.1961.287775.

Misra, N. N., Yash Dixit, Ahmad Al-Mallahi, Manreet Singh Bhullar, Rohit Upadhyay, and Alex Martynenko. “IoT, Big Data and Artificial Intelligence in Agriculture and Food Industry.” *IEEE Internet of Things Journal*, May 30, 2020, 1–1. https://doi.org/10.1109/JIOT.2020.2998584.

Mohajan, Haradhan, and Haradhan Kumar Mohajan. “Munich Personal RePEc Archive

Qualitative Research Methodology in Social Sciences and Related Subjects Qualitative Research Methodology in Social Sciences and Related Subjects." *Journal of Economic Development, Environment and People* 7 (2018): 1.

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, Doug Altman, Gerd Antes, David Atkins, et al. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *PLoS Medicine* 6, no. 7 (July 2009). https://doi.org/10.1371/JOURNAL.PMED.1000097.

Mohr, Svenja, and Rainer Kühl. "Acceptance of Artificial Intelligence in German Agriculture: An Application of the Technology Acceptance Model and the Theory of Planned Behavior." *Precision Agriculture* 22, no. 6 (2021): 1816–44. https://doi.org/10.1007/s11119-021-09814-x.

Morakanyane, Resego, and Audrey A Grace. "Conceptualizing Digital Transformation in Business Organizations: A Systematic Review of Literature," 2017. https://doi.org/10.18690/978-961-286-043-1.30.

Moysiadis, Vasileios, Panagiotis Sarigiannidis, Vasileios Vitsas, and Adel Khelifi. "Smart Farming in Europe." *Computer Science Review* 39 (February 1, 2021): 100345. https://doi.org/10.1016/J.COSREV.2020.100345.

Mulisa, Feyisa. "When Does a Researcher Choose a Quantitative, Qualitative, or Mixed Research Approach?" *Interchange* 53, no. 1 (2021): 113–31. https://doi.org/10.1007/s10780-021-09447-z.

Myers, J Paul. "The Japanese Fifth Generation Computing Project: Curricular Applications *," n.d. https://doi.org/10.5555/3447307.3447314.

Narayanan, Ramachandran Trichur. "Novice Programmer to New-Age Application Developer: What Makes Python Their First Choice?" *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*, July 1, 2019. https://doi.org/10.1109/ICCCNT45670.2019.8944583.

Newmarch, Jan. "Data Serialization." *Network Programming with Go*, 2017, 57–86. https://doi.org/10.1007/978-1-4842-2692-6_4.

Pang, Bo, Erik Nijkamp, and Ying Nian Wu. "Deep Learning With TensorFlow: A Review:" *Https://Doi.Org/10.3102/1076998619872761* 45, no. 2 (September 10, 2019): 227–48. https://doi.org/10.3102/1076998619872761.

Pasqual, G. M. "Development of an Expert System for the Identification and Control of Weeds in Wheat, Triticale, Barley and Oat Crops." *Computers and Electronics in Agriculture* 10, no. 2 (March 1, 1994): 117–34. https://doi.org/10.1016/0168-1699(94)90016-7.

Pavlovskiy, Evgeniy N. "Problems and Prospectives of Big Data Storage and Processing Standartization." *SIBIRCON 2019 - International Multi-Conference on Engineering, Computer and Information Sciences, Proceedings*, October 1, 2019, 995–98. https://doi.org/10.1109/SIBIRCON48586.2019.8958046.

Pezoa, Felipe, Juan L. Reutter, Fernando Suarez, Martín Ugarte, and Domagoj Vrgoč. "Foundations of JSON Schema." *25th International World Wide Web Conference, WWW 2016*, 2016, 263–73. https://doi.org/10.1145/2872427.2883029.

Pimentel, Joao Felipe, Leonardo Murta, Vanessa Braganholo, and Juliana Freire. "A Large-Scale Study about Quality and Reproducibility of Jupyter Notebooks." *IEEE International Working Conference on Mining Software Repositories* 2019-May (2019): 507–17. https://doi.org/10.1109/MSR.2019.00077.

Pimentel, João Felipe, Leonardo Murta, Vanessa Braganholo, and Juliana Freire. "Understanding and Improving the Quality and Reproducibility of Jupyter Notebooks." *Empirical Software Engineering* 26, no. 4 (2021). https://doi.org/10.1007/s10664-021-09961-9.

Poltavtseva, Maria A., Pavel V. Semyanov, and Elizaveta A. Zaitzeva. "Heterogeneous Semi-Structured Data Analysis in Information Security." *2020 International Conference Engineering and Telecommunication, En and T 2020*, November 25, 2020. https://doi.org/10.1109/ENT50437.2020.9431309.

Rittelmeyer, Jack Daniel, and Kurt Sandkuhl. "Effects of Artificial Intelligence on Enterprise Architectures - A Structured Literature Review." *Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOCW*, 2021, 130–37. https://doi.org/10.1109/EDOCW52865.2021.00042.

Robert, P. C. "Precision Agriculture: A Challenge for Crop Nutrition Management." *Progress in Plant Nutrition: Plenary Lectures of the XIV International Plant*

Nutrition Colloquium, 2002, 143–49. https://doi.org/10.1007/978-94-017-2789-1_11.

Roedder, Nico, David Dauer, Kevin Laubis, Paul Karaenke, and Christof Weinhardt. "The Digital Transformation and Smart Data Analytics: An Overview of Enabling Developments and Application Areas." *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 2016, 2795–2802. https://doi.org/10.1109/BIGDATA.2016.7840927.

SAMPTKUMARI BANDAGAR Asst Professor, SMT M. "CROP AND SOIL MONITORING BY ARTIFICIAL INTELLIGENCE TECHNIQUE." *International Journal for Research Trends and Innovation* 7 (2022): 2456–3315. www.ijrti.org.

Schuett, Jonas. "A Legal Definition of AI." *Article in SSRN Electronic Journal*, 2019. https://doi.org/10.2139/ssrn.3453632.

Schwertner, K. "DIGITAL TRANSFORMATION OF BUSINESS." *Trakia Journal of Sciences* 15 (2017): 388–93. https://doi.org/10.15547/tjs.2017.s.01.065.

Shah, Syed Iftikhar Hussain, Vassilios Peristeras, and Ioannis Magnisalis. "Government Big Data Ecosystem: Definitions, Types of Data, Actors, and Roles and the Impact in Public Administrations." *ACM Journal of Data and Information Quality* 13, no. 2 (May 6, 2021). https://doi.org/10.1145/3425709.

Shankar, Poorna, Prashant Pareek, Urvi Patel, and Canny Sen. "A Peer Reviewed Open Access International Journal." *Journal of Development Economics and Management Research Studies (JDMS)* 2582, no. 11: 127–37. Accessed June 1, 2022. https://www.cdes.org.in/https://www.cdes.org.

Sharma, Robin. "Artificial Intelligence in Agriculture: A Review." *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, no. Iciccs (2021): 937–42. https://doi.org/10.1109/ICICCS51141.2021.9432187.

Shmueli, Galit, and Otto R. Koppius. "Predictive Analytics in Information Systems Research." *MIS Quarterly: Management Information Systems* 35, no. 3 (2011): 553–72. https://doi.org/10.2307/23042796.

Simitsis, Alkis, Panos Vassiliadis, and Timos Sellis. "Optimizing ETL Processes in Data

Warehouses." *Proceedings - International Conference on Data Engineering*, 2005, 564–75. https://doi.org/10.1109/ICDE.2005.103.

Sinha, V., F. Doucet, C. Siska, R. Gupta, S. Liao, and A. Ghosh. "YAML: A Tool for Hardware Design Visualization and Capture." *Proceedings of the International Symposium on System Synthesis* 2000-January (2000): 9–14. https://doi.org/10.1109/ISSS.2000.874023.

Snyder, Hannah. "Literature Review as a Research Methodology: An Overview and Guidelines." *Journal of Business Research* 104 (November 2019): 333–39. https://doi.org/10.1016/J.JBUSRES.2019.07.039.

Soltanpoor, Reza, and Timos Sellis. "Prescriptive Analytics for Big Data." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9877 LNCS (2016): 245–56. https://doi.org/10.1007/978-3-319-46922-5_19.

Song, Sa Kwang, Donald J. Kim, Myunggwon Hwang, Jangwon Kim, Do Heon Jeong, Seungwoo Lee, Hanmin Jung, and Wonkyung Sung. "Prescriptive Analytics System for Improving Research Power." *Proceedings - 16th IEEE International Conference on Computational Science and Engineering, CSE 2013*, 2013, 1144–45. https://doi.org/10.1109/CSE.2013.169.

Suseendran, G., D. Balaganesh, D. Akila, and Souvik Pal. "Deep Learning Frequent Pattern Mining on Static Semi Structured Data Streams for Improving Fast Speed and Complex Data Streams." *2021 International Conference on Optimization and Applications, ICOA 2021*, May 19, 2021. https://doi.org/10.1109/ICOA51614.2021.9442621.

Tan, Mingxing, and Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." *36th International Conference on Machine Learning, ICML 2019* 2019-June (May 28, 2019): 10691–700. https://doi.org/10.48550/arxiv.1905.11946.

"The Difference between a Systematic Review and a Literature Review - Covidence," n.d.

Thorpe, Jane Hyatt, and Elizabeth Alexandra Gray. "Big Data and Public Health:

Navigating Privacy Laws to Maximize Potential.” *Public Health Reports* 130, no. 2 (March 1, 2015): 171–75. https://doi.org/10.1177/003335491513000211.

Uyanık, Gülden Kaya, and Neşe Güler. “A Study on Multiple Linear Regression Analysis.” *Procedia - Social and Behavioral Sciences* 106 (December 2013): 234–40. https://doi.org/10.1016/J.SBSPRO.2013.12.027.

Vasavi, Pallepati, Arumugam Punitha, and T. Venkat Narayana Rao. “Crop Leaf Disease Detection and Classification Using Machine Learning and Deep Learning Algorithms by Visual Symptoms: A Review.” *International Journal of Electrical and Computer Engineering* 12, no. 2 (2022): 2079–86. https://doi.org/10.11591/ijece.v12i2.pp2079-2086.

———. “Crop Leaf Disease Detection and Classification Using Machine Learning and Deep Learning Algorithms by Visual Symptoms: A Review.” *International Journal of Electrical and Computer Engineering* 12, no. 2 (April 1, 2022): 2079–86. https://doi.org/10.11591/IJECE.V12I2.PP2079-2086.

Vial, Gregory. “JOURNAL OF STRATEGIC INFORMATION SYSTEMS REVIEW Manuscript Title: Understanding Digital Transformation: A Review and a Research Agenda,” 2019.

Wang, Guanhua. “Improving Data Transmission in Web Applications via the Translation between XML and JSON.” *Proceedings - 2011 3rd International Conference on Communications and Mobile Computing, CMC 2011*, 2011, 182–85. https://doi.org/10.1109/CMC.2011.25.

Wang, Weiyu, and Keng Siau. “Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda.” *Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/JDM.2019010104* 30, no. 1 (January 1, 1AD): 61–79. https://doi.org/10.4018/JDM.2019010104.

“Why Systematic Reviews Matter,” n.d.

Wiktorski, Tomasz, Yuri Demchenko, and Oleg Chertov. “Data Science Model Curriculum Implementation for Various Types of Big Data Infrastructure Courses.” *Proceedings - IEEE 15th International Conference on EScience, EScience 2019*, September 1, 2019, 541–47. https://doi.org/10.1109/ESCIENCE.2019.00074.

Wolfert, Sjaak, Lan Ge, Cor Verdouw, and Marc Jeroen Bogaardt. "Big Data in Smart Farming – A Review." *Agricultural Systems* 153 (May 1, 2017): 69–80. https://doi.org/10.1016/J.AGSY.2017.01.023.

Wong, Ho Ting, Vico Chung Lim Chiang, Kup Sze Choi, and Alice Yuen Loke. "The Need for a Definition of Big Data for Nursing Science: A Case Study of Disaster Preparedness." *International Journal of Environmental Research and Public Health 2016, Vol. 13, Page 1015* 13, no. 10 (October 17, 2016): 1015. https://doi.org/10.3390/IJERPH13101015.

Wu, Xindong, Xingquan Zhu, Gong Qing Wu, and Wei Ding. "Data Mining with Big Data." *IEEE Transactions on Knowledge and Data Engineering* 26, no. 1 (January 2014): 97–107. https://doi.org/10.1109/TKDE.2013.109.

Xu, Zongben, and Yong Shi. "Exploring Big Data Analysis: Fundamental Scientific Problems." *Annals of Data Science* 2, no. 4 (December 1, 2015): 363–72. https://doi.org/10.1007/S40745-015-0063-7/FIGURES/1.

Yu, Seon-Cheol, Dong-Bin Shin, and Jong-Wook Ahn. "A Study on Concepts and Utilization of Geo-Spatial Big Data in South Korea." *KSCE Journal of Civil Engineering* 20, no. 7 (2016): 2893–2901. https://doi.org/10.1007/s12205-016-0504-7.

Zolnowski, Andreas, and Research Zolnowski. "Business Model Transformation Patterns of Data-Driven Innovations Service Dominant Architecture View Project Service Dominant Architecture View Project BUSINESS MODEL TRANSFORMATION PATTERNS OF DATA-DRIVEN INNOVATIONS," 2016. https://www.researchgate.net/publication/304492522.

Internet Resources:

"11 Agricultural Robots You Should Know | Built In." Accessed August 7, 2022. https://builtin.com/robotics/farming-agricultural-robots.

"Big Data: How Data Analytics Is Transforming the... (PDF)." Accessed May 19, 2022. https://pdfroom.com/books/big-data-how-data-analytics-is-transforming-the-world/0YpgQAqVgNz.

Contributor, Jason Bloomberg. “Digitization, Digitalization, And Digital Transformation: Confuse Them At Your Peril I Write and Consult on Digital Transformation in the Enterprise.” Accessed May 29, 2022. https://www.forbes.com/sites/jasonbloomberg/2018/04/29/digitization-digitalization-and-digital-transformation-confuse-them-at-your-peril/#78e677fd2f2c.

“Data Collection and Analysis - Google Books.” Accessed May 28, 2022. https://books.google.de/books?hl=en&lr=&id=BEDTrvUH8NcC&oi=fnd&pg=PA243&dq=unstructured+data&ots=WySzfTLIK8&sig=17kARegGpQeTZAgdjXg4E56vj88&redir_esc=y#v=onepage&q=unstructured data&f=false.

“Digital Transformation: Key Types to Understand - BairesDev.” Accessed May 29, 2022. https://www.bairesdev.com/blog/digital-transformation-types-to-understand/.

“EfficientNet B0 to B7.” Accessed August 7, 2022. https://keras.io/api/applications/efficientnet/#efficientnetb3-function.

“Evolution of AI: Past, Present, Future | by Christina Aguis | DataDrivenInvestor.” Accessed August 7, 2022. https://medium.datadriveninvestor.com/evolution-of-ai-past-present-future-6f995d5f964a.

“Extensible Markup Language (XML) 1.0.” Accessed May 28, 2022. http://www.w3.org/TR/1998/REC-xml-19980210http://www.w3.org/TR/1998/REC-xml-19980210.xmlhttp://www.w3.org/TR/1998/REC-xml-19980210.htmlhttp://www.w3.org/TR/1998/REC-xml-19980210.pdfhttp://www.w3.org/TR/1998/REC-xml-19980210.pshttp://www.w3.org/TR/REC-xml.

Fernandez, Mary, Dan Suciu, and Chun-Nan Hsu. “XML-QL: A Query Language for XML A Query Language for XML Alin Deut Sch Semist Ruct Ured Dat a and XML Induct Ion of Int Egrat Ed View for XML Dat a Wit h Het Erogeneous DT Ds XML-QL: A Query Language for XML.” Accessed May 28, 2022. http://www.w3.org/TR/1998/NOTE-xml-ql-19980819http://www.w3.org/TR/1998/NOTE-xml-ql-19980819.htmlLatestversion:http://www.w3.org/TR/NOTE-xml-ql.

"How Can We Discover the Most Valuable Types of Big Data and Artificial Inte...: FOM Hochschule Online-Literatursuche." Accessed May 28, 2022. https://eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=3&sid=aaf1d93e-7777-44f8-b739-f40b767a2011%40redis.

"Importance And Benefits Of Artificial Intelligence." Accessed June 1, 2022. https://www.hdatasystems.com/blog/importance-and-benefits-of-artificial-intelligence.

Nyce, Charles. "Predictive Analytics White Paper." Accessed May 29, 2022. www.aicpcu.org.

"Rice Leafs | Kaggle." Accessed August 4, 2022. https://www.kaggle.com/datasets/shayanriyaz/riceleafs.

"The 4 Main Areas of Digital Transformation | With Examples | Poppulo." Accessed May 29, 2022. https://www.poppulo.com/blog/what-are-the-4-main-areas-of-digital-transformation.

"The 4 Types of Digital Transformation and 3 Approaches." Accessed May 29, 2022. https://www.chaione.com/blog/4-digital-transformation-types.

"The 4 Types of Digital Transformation Explained." Accessed May 29, 2022. https://www.servicesbytechdata.com/resources/the-4-types-of-digital-transformation-explained.

"The History of Artificial Intelligence - Science in the News." Accessed August 7, 2022. https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

"The Opportunity of Big Data Research for Health Geographers : A Case Study of Disaster Management — The Hong Kong Polytechnic University." Accessed May 19, 2022. https://research.polyu.edu.hk/en/publications/the-opportunity-of-big-data-research-for-health-geographers-a-cas.

"What Is Business Process Transformation? | A Complete Guide - Kissflow." Accessed May 29, 2022. https://kissflow.com/workflow/bpm/business-process-transformation/.

"What Is Convolutional Neural Network? - Definition from WhatIs.Com." Accessed

August 7, 2022. https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network.

Declaration in Lieu of Oath

I hereby declare that I produced the submitted paper with no assistance from any other party and without the use of any unauthorized aids and, in particular, that I have marked as quotations all passages which are reproduced verbatim or near-verbatim from publications. Also, I declare that the submitted print version of this thesis is identical with its digital version. Further, I declare that this thesis has never been submitted before to any examination board in either its present form or in any other similar version. I herewith agree/disagree that this thesis may be published. I herewith consent that this thesis may be uploaded to the server of external contractors for the purpose of submitting it to the contractors' plagiarism detection systems. Uploading this thesis for the purpose of submitting it to plagiarism detection systems is not a form of publication.

Date: 16/08/2022 Urvi Patel