Cape Peninsula
University of Technology

Comparative analysis of explainable deep learning models for identification of plant nutrient deficiencies

by

Junior Mkhatshwa

Thesis submitted in partial fulfilment of the requirements for the degree

Master of Technology: Information Technology

in the Faculty of Informatics and Design

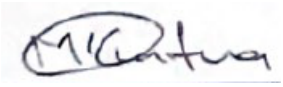
at the Cape Peninsula University of Technology

Supervisor: Prof Justine Olawande Daramola
Co-supervisor: Dr Tatenda Kavu

Cape Town
September 2023

DECLARATION

I, **Junior Mkhatshwa**, declare that the contents of this thesis represent my own unaided work and that the thesis/dissertation has not previously been submitted for academic examination towards any qualification. Furthermore, it represents my own opinions and not necessarily those of the Cape Peninsula University of Technology.



20 September 2023

Signed **Date**

ABSTRACT

Essential mineral nutrients play a crucial role in the growth and survival of plants. The lack of nutrients in plants threatens global food security and affects farmers who solely depend on producing healthy crops. Traditionally the identification of nutrient deficiencies in a crop is done manually by experienced farmers. Deep learning (DL) has shown promise in image classification. However, the lack of understanding regarding the accuracy and explainability of specific DL models for the identification of plant nutrient deficiencies is a hindrance to making informed decisions about the suitability of these algorithms for practical implementation.

This study aimed to assess the performance and explainability of these models to facilitate better decision-making in agriculture. To achieve this, the study formulated four objectives: 1) identify the features that are essential to determine plant nutrient deficiencies; 2) determine the requirements of explainable DL for nutrient deficiency identification; 3) explore how explainable DL could be applied to a plant image dataset to identify plant nutrient deficiencies; and 4) determine the performance and explainability of selected DL algorithms when used for plant nutrient deficiencies.

The study used a deductive approach to achieve the aforementioned objectives, using a quantitative research methodology and an experimental research design to investigate the performance and interpretability of three machine learning (ML) models based on two plant datasets: rice and banana. The three DL models were a Convolutional Neural Network (CNN), and two pre-trained models: Inception-V3 and Visual Geometry Group (VGG-16). For the explainability of the models, the study used two XAI techniques: Shapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM). The study found that the choice of DL models has a significant impact on the performance of nutrient deficiency identification in different plant datasets. Inception-V3 achieved a *very good* F1-Score of 92% for the banana dataset. VGG-16 follows with a *good* F1-Score of 81% and the CNN, while not as strong as the other models, achieves an acceptable F1-Score of 68%.

Based on these findings, Inception-V3 is effective in detecting nutrient deficiency in banana plants. Regarding explainability using SHAP, the CNN and VGG-16 models were found to rely on a limited set of prominent features. However, Inception-V3 appears to rely on a broader range of features, with many features making significant contributions to the final prediction. When Grad-CAM was used to assess explainability, for the banana and rice datasets, it was noted that the Grad-CAM heatmap of the CNN model highlights the contours of the plant leaf, while the other two models (Inception V-3, and VGG-16) focus on the leaf itself. VGG-16 accurate localisation of the affected regions proved to be more reliable due to the quality of its

heatmap. The result of this study shows that Inception-V3 is the most accurate model, but it may not be the most explainable model due to its complex architecture. On the other hand, the VGG-16 has a simpler architecture that tends to offer a better explanation. Therefore, balancing accuracy and explication when selecting a model for a particular task is essential.

The study contributes to the literature by incorporating explainable deep learning in the context of plant nutrient deficiency identification. Moreover, unlike prior research that primarily evaluated accuracy without considering explainability, the study addressed this gap by comparing the explainability of GRAD-CAM and SHAP techniques, shedding light on how these models arrive at their predictions. The research successfully addressed its objectives, providing valuable insights into both the theoretical and practical aspects of this domain. The study's holistic approach and valuable findings pave the way for the integration of XAI techniques in agriculture, adding value to the field and opening avenues for future research and innovation.

Keywords: explainable artificial intelligence; Machine learning; SHAP; nutrient deficiency; deep learning; transfer learning; convolutional neural network; Grad-CAM

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following individuals:

- I am deeply grateful to Prof Justine Olawande Daramola, my supervisor for the guidance, support, and direction provided since my first encounter with him as a professor of Artificial intelligence and Data Science (IDS580S). Thank you for giving me intellectual freedom and opportunities to grow.
- Dr Tatenda Kavu, thank you for all your helpful research insight and input.
- I would like to express my gratitude to my parents, Oters and Florah Mkhatshwa, for their sacrifices, understanding and encouragement throughout my life. Thank you for always being there for me and helping me pursue my dreams.
- Laura Makhubedu for her words of encouragement and loving support.
- My Siblings and friends for unknowingly bringing much-needed humour into my life.

This thesis has been a long, exciting, and rewarding journey. I am grateful for the support I received, and I am glad I did not have to go through it alone.

DEDICATION

I dedicate this thesis to my parents, Oters and Florah Mkhatshwa. I am forever grateful for your love, prayers, unwavering support, and constant encouragement.

PUBLICATION FROM THE THESIS

Junior Mkhatshwa, Tatenda Kavu, Olawande Daramola (2023). Analyzing the Performance and Interpretability of CNN-based architectures for plant nutrient deficiencies identification, *Multimedia Tools and Application* (Under Review)

Junior Mkhatshwa & Olawande Daramola (2023). Image-based Identification of Plant Diseases and Nutrient Deficiency Using Interpretable Machine Learning. *Digital Data Processing* (Under Review)

TABLE OF CONTENTS

List of Figures

List of Tables

GLOSSARY

Abbreviation/Acronym	Definition
ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
BN	Batch Normalization
Ca	Calcium
CNN	Convolutional Neural Network
CV	Cross-validation
CPU	Central Processing Unit
CV	Computer Vision
DCNN	Deep convolutional neural network
DNN	Deep Neural Network
Fe	Ferrum
GAN	Generative Adversarial Network
Grad- CAM	Gradient-weighted Class Activation Mapping
GPU	Graphics Processing Unit
JPEG	Joint Photographic Experts Group
K	Potassium
KNN	K nearest neighbour
LR	Learning Rate
Mg	Magnesium
ML	Machine Learning
Mn	Manganese
N	Nitrogen
NLP	Natural Language Processing
NN	Neural Network
P	Phosphorus
ReLu	Rectified Linear Unit
RF	Random Forest
RGB	Red-green-blue
S	Sulphur
SVM	Support vector machine
XAI	eXplainable Artificial Intelligence
YOLO	You only look once
Zn	Zinc

CHAPTER ONE

INTRODUCTION

1.1 Motivation for the Study

Essential mineral nutrients play a crucial role in the growth and survival of plants, and they are required in varying quantities as macronutrients and micronutrients. Minni & Rehna (2016) state that macronutrients are nutrients that are needed in significant amounts, while micronutrients are required in smaller quantities. Potassium (K), Phosphorus (P), Sulphur (S), Nitrogen (N), Magnesium (Mg), and Calcium (Ca) are the six macronutrients that are essential for plant growth and development (Shete, Gosavi, Bulbule, Patil & Pawar, 2020). Micronutrients consist of Ferrum (Fe), Manganese (Mn), Zinc (Zn), and Cuprum (Cu) (also known as copper). In the absence of the aforementioned nutrients, plants cannot complete their lifecycle and accomplish their physiological functions (Yi, Krusenbaum, Unger, Hüging, Seidel, Schaaf & Gall, 2020).

Traditionally the identification of nutrient deficiencies in a crop is done manually by experienced farmers. According to Singh (2017), the process is expensive, time-consuming, and laborious. Moreover, identifying nutrient deficiencies is difficult without the right facilities (Singh & Misra, 2017). To understand the deficiencies in plants, non-invasive methods like computer vision and sensors have been created. According to Inácio (2018), machine learning (ML) has been used to monitor plant nutrient deficiencies accurately.

Recent advancements in deep learning (DL) have seen significant progress in the field of Artificial Intelligence (AI) (Lecun, Bengio & Hinton, 2015). DL has advanced well beyond traditional machine learning (ML) methods in domains such as natural language processing (NLP) (Mccann, Bradbury, Xiong & Socher, 2017), image classification (Krizhevsky, Sutskever & Hinton, 2012), and audio recognition (Chiu, Sainath, Wu, Prabhavalkar, Nguyen, Chen, Kannan, Weiss & Rao, 2018). The use of DL models in agriculture is vast and varied. These models have been applied to plant identification, as demonstrated by Ghazi, Yanikoglu & Aptoula (2017). Additionally, DL models have been utilised for yield forecasting, as evidenced by Wolanin, Mateo-Garciá, Camps-Valls, Gómez-Chova, Meroni, Duveiller, Liangzhi & Guanter (2020). Furthermore, they have been used for disease detection, as shown by Mohanty, Hughes & Salathé (2016).

According to Liu, Wang, Liu & Alsaadi (2016), DL is a novel advancement in ML that generates ground-breaking outcomes in various fields such as computer vision (CV). Due to its deep architecture, DL can manage much more challenging tasks (Wang, 2017). DL has the

advantages of automation, speed, and low cost (Lu, Han, Chen, Yu & Xue, 2021). Moreover, recent studies such as Alzubaidi, Zhang, Humaidi, Al-Dujaili, Duan, Al-Shamma, Santamaría & Fadhel (2021), have shown that DL's performance has surpassed human performance on tasks such as image classification.

According to Gerlings, Shollo & Constantiou (2021), diverse types of algorithms that people may not completely understand are being used increasingly to help or entirely automate important decisions. The term used to describe this is the "AI black box". As black box models have critical implications, the right to explanation becomes crucial for the acceptance of AI in society. Providing explanations that make the AI algorithm more expressive can enhance human comprehension, boost confidence in decision-making, and foster impartial and fair decisions (Das & Rad, 2020).

This study presents a novel approach in the field of agriculture by focusing on the comparative analysis of the performance and explainability of DL models in the identification of plant nutrient deficiencies. The intricacies of the agricultural domain, including diverse plant types, seasonal variations, and environmental factors, demand transparent and understandable models.

To address this need, the study used two different plant datasets to explore the performance and explainability of DL models in identifying nutrient deficiencies in plants. By leveraging these models, the researcher aimed to uncover crucial insights into how these essential features in the data influence the prediction of nutrient deficiencies. As highlighted by Schmidt, Jensen & Husted (2016), Nutrient deficiencies can significantly impact plant growth and yield, and timely identification and correction are crucial for optimal plant performance.

1.2 Background

Human nutrition is affected by the nutritional value of plant products, whether directly or indirectly, as stated by Kihara, Bolo, Kinyua, Rurinda & Piikki (2020). Agriculture has seen the automation of various tasks through the application of image analysis techniques, including plant species identification, nutrient deficiency detection, and image-based plant disease detection, as reported by Watchareeruetai, Noinongyao, Wattanapaiboonsuk, Khantiviriya & Duangsrisai (2018), and Mohanty, Hughes & Salathé (2016).

Visual signs of nutrient deficiencies can be seen in several sections of the plant, particularly the leaves. Visual symptoms include altered leaf colour, necrosis, and chlorosis (Leena, 2018). The early-stage diagnosis of nutrient deficiencies is an important task. The final plant produced can be affected by nutrient shortages in plants. According to Mohanty et al. (2016), the lack of nutrients in plants threatens global food security and affects smallholder farmers who solely

depend on producing healthy crops. Identifying nutrient deficiencies is crucial to ensuring that all plants are developed according to specifications for nutrients like nitrogen.

Kali (2006), published a paper in the International Encyclopedia of Education journal that first introduced the concept of DL. The fundamental idea behind DL is to use "deeper" neural networks that apply various convolutions to provide a hierarchical representation of the input. DL's automatic feature extraction from raw data is a significant advantage. Kamilaris (2018), states that DL is highly efficient and fast in solving more complex problems.

In general, DL models are complex and not easily understandable by humans. Barredo (2020) notes that the demand for transparency among AI stakeholders is increasing, particularly as black-box models are employed to make important predictions in critical situations. The absence of transparency and accountability surrounding prediction models can have disastrous results in some circumstances. XAI refers to techniques for obtaining the intelligible output of an ML model and presenting it in a manner that is understandable to humans (Gerlings, Shollo & Constantiou, 2021). Páez (2019), highlights that one of the crucial objectives of XAI is to establish trust in the model. ML models need to provide not only output but also understandable explanations behind the rationale of the machine.

The identification of plant nutrient deficiencies through the implementation of ML techniques has been explored in the existing literature such as Tran, Choi, Le & Kim (2019). However, there is still a significant research gap in the comparative analysis of the explainability of DL models in identifying nutrient deficiencies in plants. While some studies have looked into the use of ML for this purpose, there has not been much research done on the effectiveness and potential benefits of using explainable DL. Addressing this gap is essential to gaining a comprehensive understanding of the most effective and explainable approach for nutrient deficiency identification in plants.

1.3 Research Problem

Deep learning (DL) models' inability to provide sufficient justification is what leads to the models being called black boxes. According to Rudin (2019), several ML model predictions cannot be understood by humans due to their nature of being a black box. According to Sparrow, Howard & Degeling (2021), we need to understand why the machine-learning model performs the way it does.

According to Barredo (2020), practical implementation of explainability is imperative. However, it is currently one of the main barriers to AI. In some cases, not understanding model predictions may result in financial losses. The lack of understanding regarding the accuracy

and explainability of specific DL models hinders researchers from selecting the appropriate algorithm for detecting nutrient deficiencies in plants.

The lack of understanding regarding the accuracy, and explainability of specific DL models for the identification of plant nutrient deficiencies is a hindrance to making informed decisions about the suitability of these algorithms for practical implementation.

1.4 Aim, Objectives, and Research Questions

The study aims to compare the performance of deep learning algorithms when applied to plant nutrient deficiency and to ascertain the level of their explainability.

1.4.1 Research Objectives

To achieve the aim of this study, the following objectives were formulated:

i. To identify the features that are essential to determine plant nutrient deficiencies.
ii. To determine the requirements of explainable deep learning for nutrient deficiency identification.
iii. To explore how explainable deep learning could be applied to a plant image dataset to identify plant nutrient deficiencies.
iv. To determine the performance of selected deep learning algorithms when used for plant nutrient deficiencies and their explainability.

1.4.2 Research Questions

The main research question is:

RQ1: What is the relative performance and explainability of deep learning algorithms when applied to plant nutrient deficiency?

Some sub-research questions were formulated to enable the answering of the main research question. These are:

RSQ1: What are the features essential to determine plant nutrient deficiencies?

RSQ2: What are the requirements of explainable deep learning for plant nutrient deficiencies identification?

RSQ3: How can explainable deep learning be applied to a plant image dataset for plant nutrient deficiency identification?

RSQ4: What is the rating of the performance and explainability of selected deep learning algorithms for plant nutrient deficiency identification?

1.5 Significance of the Study

The research has multiple benefits for various stakeholders. Firstly, it can aid plant health professionals in identifying nutrient deficiencies in plants, simplifying their work. Secondly, farmers can use the research findings to identify nutrient deficiencies that may be affecting their crop yield and quality, enabling them to make informed decisions based on the data. Additionally, the study's findings can contribute to the body of literature, benefiting academics and upcoming researchers. The research offers new methods of using XAI in agriculture, adding value to the existing literature.

1.6 Delineation of Study

The study compared the performance of deep learning algorithms when applied to plant nutrient deficiency and to ascertain their explainability. This study focussed on three deep learning models: standard CNN, VGG-16, and Inception-v3. The research was limited to two plant types: rice and banana.

1.7 Structure of the Thesis

The organisation of the rest of this thesis is described as follows.

Chapter 2 presents a review of the literature. The chapter is divided into four sections, which include the challenges of identifying nutrient deficiencies, deep learning techniques, XAI, and a summary of related work. Chapter 3 covers the research approach, methodological approach, research design, and research strategy. Additionally, the study's ethical considerations are discussed in detail. Chapter 4 presents the implementation details for the model and the datasets used in this research. It provides sufficient information for other researchers to replicate the findings of the study. Chapter 5 presents the experimental results. The final chapter, Chapter 6, provides a summary of the findings and proposes possible future work.

1.8 Chapter Summary

In this chapter, the research problem is outlined, along with the study’s aim and objectives, as well as the derived research questions aimed at addressing the problem. Additionally, the chapter introduces the topic and offers a comprehensive background of the study. Moreover, the author substantiates the significance of the research.

CHAPTER TWO

LITERATURE REVIEW

This chapter presents a review of the literature. The chapter is divided into four sections, which include the challenges of identifying nutrient deficiencies, deep learning techniques, XAI, and a summary of related work.

2.1 Relationship between plant nutrient deficiency and plant diseases

McCauley (2013) states that nutrient deficiency occurs when a growing plant's needs are not met due to the absence of a crucial nutrient in sufficient amounts. Spann and Schumann (2010) state that the way each nutrient affects a plant's response to a disease, be it positive or negative, is unique to each complex of plant diseases. Different nutrients affect plant diseases differently (Bhaduri, Rakshit & Chakraborty, 2014). Moreover, crops that receive appropriate nutrients become more tolerant or resistant to disease, and plants that lack nutrients are more sensitive to diseases. According to Timothy and Arnold (2009), the two organisms' genotypes, the age of the plant, and environmental changes all play a role in resistance.

According to Bhaduri et al. (2014), it is challenging to discern disease symptoms since they typically reflect the plant's changing nutritional status. When there is a lack of nutrients, bacteria and fungi create more enzymes that dissolve some of the plant tissue (Pandey, Shamim, Srivastava, Dwivedi, Awasthi & Singh, 2015). Plants are more vulnerable to bacterial attacks if N levels are low, but K and Ca are essential for creating an efficient defence against diseases (Bhaduri et al., 2014). This resistance is lowered by potassium shortages brought on by excessive dolomite or magnesium application. On the other hand, a lack of zinc can make plants more vulnerable to powdery mildew, a fungus that damages the leaves and stems of plants. The following subsection provides an overview of how vital nutrients affect disease resistance or severity:

i. **Potassium (K)**: When K is added, plants' sensitivity to diseases is often reduced to the level necessary for healthy growth. Proteins, starches, and cellulose are produced by plants using K. According to Bhaduri et al. (2014), since cellulose makes up the majority of cell walls, a K deficiency results in leaky cell walls, which leads the leaf apoplast to have large quantities of sucrose (a precursor to starch) and amino acids (protein building blocks). Common scabs in potatoes and Phytophthora root rot in citrus are two diseases that are made worse by K's unfavourable interactions with Ca (Pandey et al.,

2015). Moreover, it has frequently been noted that K lowers the prevalence of several diseases including bacterial leaf blight and black rust in rice (Pandey et al., 2015).

ii. **Nitrogen (N)**: For plant growth, nitrogen is the most crucial nutrient. It is so simple to explain how nitrogen contributes to disease resistance, there exists an extensive collection of literature that discusses the influence of nitrogen on diseases (Timothy & Arnold, 2009). N is a crucial ingredient in amino acids. According to Timothy and Arnold (2009), the way that the plant assimilates the two types of nitrogen NO3 – and NH4+ can have a significant impact on how diseases develop. However, the first one is more easily susceptible to leaching. Bhaduri et al. (2014) state that some disease-causing fungal species, including Fusarium and Phytophthora, are poisonous to nitrate (NO2-), which are created in the soil by beneficial soil bacteria from ammonium nitrogen throughout the N cycle. N makes diseases including powdery mildew of wheat, rusts, and grey leaf spots of corn as well as seed infection in cereals more severe. Early blight is more likely to affect potatoes with low N levels (Pandey et al., 2015).

iii. **Phosphorus (P)**: According to Bhaduri et al. (2014), P is a crucial nutrient that ranks second in importance after N. It is important for a plant's development, including energy storage and root growth. It is present in various organic compounds in cells, such as DNA, ribonucleic acids, and phospholipids, and plays a vital role in metabolic activities in both plants and pathogens. Moreover, phytic acid is the form in which P is stored in seeds. Plant enzymes are activated as a result (Pandey et al., 2015). Aye and Masih (2023), state that the plants may display signs including stunted growth, and purple discolouration on leaves when they do not have enough P. According to Timothy and Arnold (2009), when used to treat seedlings for fungal illnesses, P has been proven to be most effective because rapid root growth enables plants to fend off sickness. Scab incidence rises as a result of O, but diseases like powdery mildew, and smut decrease in wheat (Pandey et al., 2015).

iv. **Calcium (Ca)**: Cell walls and membranes include calcium, which is present in the form of Ca polygalacturonates. Ca deficiency consequently leads to cell structure disturbance, which makes plants less able to fend off infection by the pathogens that cause several diseases (Bhaduri et al., 2014). According to Pandey et al. (2015), Ca is believed to provide defence against certain diseases through mechanisms such as binding to oxalic acid or thickening the cell wall.

2.2 Nutrient deficiency symptoms in plants

Plant nutrients can be classified as micronutrients and macronutrients. Kamelia, Rahman, Saragih & Haerani (2020) state that three methods of analysing deficiencies exist: visual observation, plant analysis, and soil analysis. Plants exhibit signs of nutrient inadequacy in their leaves as a result of shortage (Goyal, Amit, Gupta, Piuri, Ganzha & Paprzycki, 2021). Using information from the leaf, the agriculturist primarily assesses the state of the plant's nutrition. The human eye often fails to detect these nutrients until the plant has already suffered damage. Therefore, the development of technology is crucial to enable farmers and experts to identify these deficiencies at an early stage.

Plants are traditionally mashed and chemically treated to detect nutrient deficiencies (Sharma, Nath, Kumar & Chaudhary 2022). However, completing this task is expensive, requires a lot of patience, and must be repeated frequently. As a result, the annual nutrient deficiency might not be accurate. Moreover, the plant's condition may deteriorate, and the plant test results will not be known for days or weeks (Kamelia et al., 2020).

Ibrahim (2022) states that the initial indications of nutrient deficiency manifest in the leaves, such as yellowing, stunted roots, and reduced leaf size (Table 2.1). Furthermore, Jose, Nandagopalan, Ubalanka & Viswanath (2021) state that these symptoms result in stunted growth and inadequate flowering.

Table 2.1: List of minerals that are important for plant growth.

Deficient Nutrient	**Symptoms of Deficiency**	**Example**
Primary Nutrients		
Nitrogen (N)	General light green to yellow (chlorotic) or light brown especially on older leaves (Billericay Fertiliser Services, 2015).	 (Nutifafa, 2020)

Phosphorous (P)	The leaves develop brown spots and necrosis. Plants' growth is stunted (Sawyer, 2004).	 (Nutifafa, 2020)
Potassium (K)	Tip burn and inter-venial marginal necrosis (Jose et al., 2021).	 (Nutifafa, 2020)
Secondary Nutrients		
Calcium (Ca)	The form or distortion of new leaves.	
Magnesium (Mg)	Within the veins of young leaves, yellowing occurs. Leaves may also drop off.	 (Jeyalakshmi, 2017)

Sulfur (S)	Leaves turn yellow first. Symptoms resemble those of nitrogen deficiency	 (Nutifafa, 2020)
Micronutrients		
Boron (B)	Leaf tips turn yellow followed by necrosis (Billericay Fertiliser Services, 2015).	 (Nutifafa, 2020)
Iron (Fe)	Bronzing of leaves with tiny brown spots.	
Manganese (Mn)	Yellowing on young leaves and brown spots on older leaves	 (Billericay Fertiliser Services, 2015)

Zinc (Zn)	Yellowing occurs between the veins of the new leaves. Symptoms appear on older leaves first	 (Nutifafa, 2020)

2.3 Deep Learning

Deep learning (DL) is a subfield of machine learning (ML) that involves the use of artificial neural networks (ANN) to learn and make predictions (Kamilaris, 2018). The structure of DL is shown in Figure 2.2. DL relies on algorithms to learn from data, and the availability of high-quality labelled datasets has contributed to its popularity (Chandra, Desai, Guo & Balasubramanian, 2020). The term "deep" in DL refers to the deep architecture of learning, which involves multiple layers of neural networks (Chandra et al., 2020). DL models such as VGG16, Inception V3, and DenseNet169 are equipped with pre-trained weights, signifying that their networks have been trained on different datasets, particularly Imagenet. Transfer learning technique is the term used to describe these pre-trained weights procedures, as illustrated in Figure 2.1 (Moujahid, Cherradi, AL-Sarem, Behatti, Alsaeedi & Saeed, 2022).

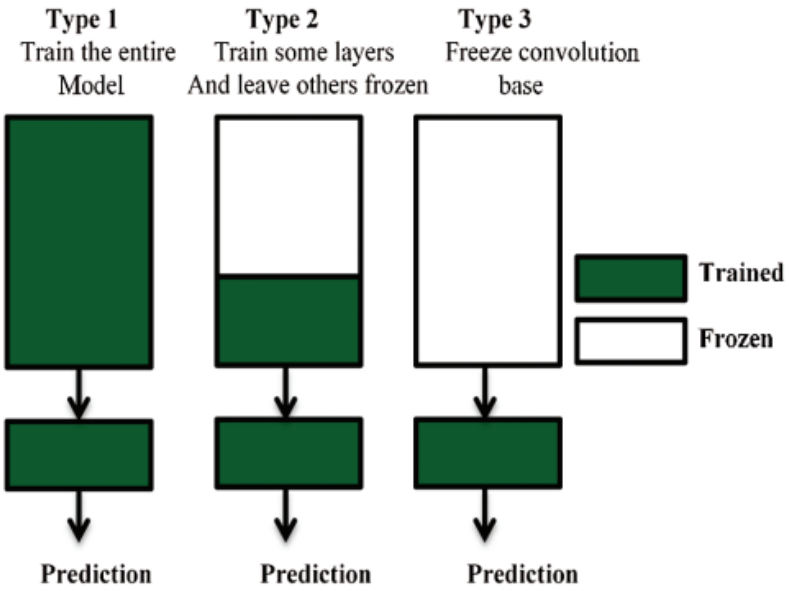


Figure 2.1: Transfer learning process (Moujahid et al., 2022)

According to Alzubaidi, Zhang, Humaidi, Duan, Santamaria, Fadhel & Farhan (2021), DL techniques may be categorised as unsupervised, partially supervised (semi-supervised) and supervised. Liu, Zhang, Hou, Wang, Mian, Zhang & Tang (2020) state that “self-supervised learning” was first introduced in robotics. The aim of supervised learning is to reduce the total classification error of the model (O’Shea & Nash, 2015).

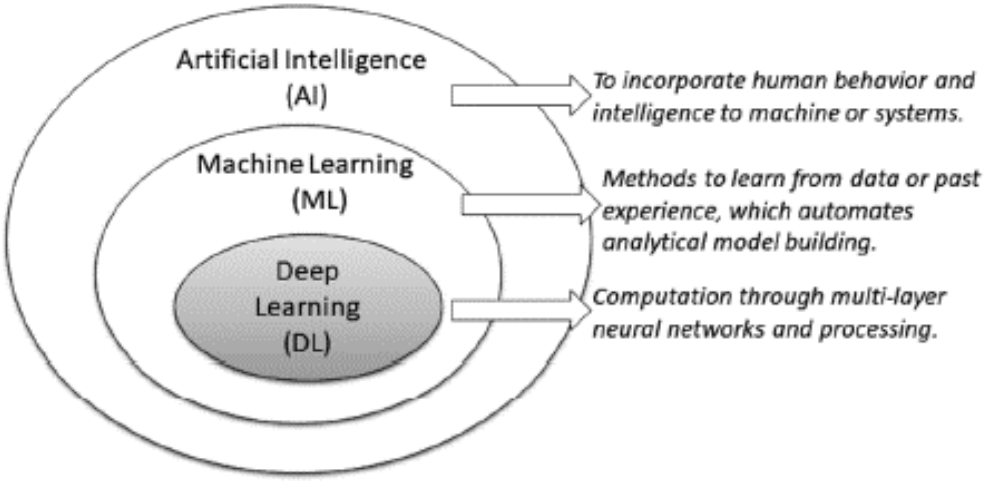


Figure 2.2: Machine learning structure (Sarker, 2021)

2.3.1 Importance of deep learning

Due to its distinct qualities and capabilities, DL has become important in many domains, including agriculture. This section emphasises the significance of DL by highlighting its universal learning approach, robustness, generalisation, and scalability.

i. **Universal learning approach**: The term “universal learning” is occasionally used to describe DL because it can function in all application fields (Alzubaidi et al., 2021).
ii. **Robustness:** Model robustness describes how much a model performs differently when trained on new data compared to being used on training data.
iii. **Generalisation**: When a DL model generalises, it can accurately predict the pattern of previously unseen data that comes from the same distribution that was used to build the model (Alzubaidi et al., 2021).
iv. **Scalability**: Scalability is the ability of DL applications to manage large amounts of data and conduct numerous computations efficiently and quickly.

2.3.2 Classification of deep learning approaches

DL approaches can be categorised into three groups: supervised, semi-supervised or partially supervised, and unsupervised. According to Zahangir, Tarek, Christopher, Stefan & Paheding (2018), another learning method is deep reinforcement learning (DRL), sometimes referred to

as Reinforcement Learning (RL), which is typically seen as falling under the category of partially supervised learning.

i. Supervised Learning

Supervised learning is the one that is frequently used form of machine learning. It is trained using well-labelled data. According to Alzubaidi et al. (2021), a set of inputs and their matching outputs $(xt, yt) \sim p$ are present in the environment when using supervised learning. There are various supervised learning methods for DL, including DNNs, RNNs and CNNS. The need for a large number of datasets with labels for training is a significant disadvantage of supervised learning. Several uses for supervised learning include text recognition, Image classification, and speech recognition.

ii. Semi-supervised Learning

This approach uses semi-labelled datasets as the foundation for the learning process. Compared to supervised learning, which can only use labelled data, semi-supervised learning can enhance learning performance by using more unlabelled examples. According to Zahangir et al. (2018), generative adversarial networks (GANs), another type of generative network, can be used as a semi-supervised learning technique.

iii. Unsupervised Learning

Without access to labelled data, unsupervised learning enables the implementation of the learning process. Frequently, unsupervised learning methods include dimensionality reduction, generative, and clustering techniques. According to Wang (2017), clustering is among the most frequently used techniques for unsupervised learning. Several uses for unsupervised learning include recommender systems, customer segmentation and product segmentation.

iv. Deep Reinforcement Learning

AI and reinforcement learning frameworks are combined in DRL to assist software agents in learning how to accomplish their objectives. DL has been identified by Wang (2017) as a solution to address problems associated with high-dimensional state spaces. DL's ability to learn multiple levels of abstraction from data allows reinforcement learning to tackle more complex problems with less prior information. According to Zahangir et al. (2018), the success of DeepMind in teaching the system to play Atari games without prior knowledge of the rules has led to the widespread adoption of DRL.

2.3.3 Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a special type of multilayer neural network, which is one of the different DL algorithms (Ghosh, Sufian, Sultana, Chakrabarti & De, 2019). It is recognised as one of the most important algorithms in the field of computer vision. According to Wang (2017), the three concepts that make up CNN architectures are local receptive field, shared weights, and spatial subsampling. CNNs recognise visual patterns from images with minimal processing. When it comes to different visual problems, CNN has been utilised successfully (Kamilaris, 2018).

Canziani, Paszke & Culurciello (2016), explain that CNN is composed of three different types of layers: convolutional, pooling, and fully connected layers, as illustrated in Figure 2.3. The convolutional layer is responsible for extracting input images and determining the output of neurons.

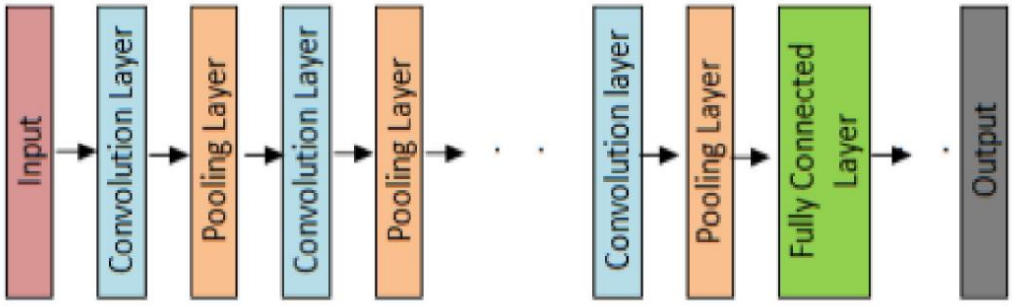


Figure 2.3 Conceptual model of CNN (Ghosh et al., 2019)

Ghosh et al. (2019), state that the feature maps generated after convolutional operations are sub-sampled using pooling layers. Additionally, learnable features are reduced. There are two primary types of pooling: max pooling and average pooling. The maximum number from each submatrix of the action map is taken by max pooling and forms a separate matrix from it (Ghosh et al., 2019). Some information is lost in the pooling layer; however, they help limit the risk of overfitting.

2.3.3.1 CNN Layers

The layered design of CNN is one of the main factors in its strength. The various types of layers that make up the fundamental components of a CNN will be examined in this section along with their functions.

i. Convolution Layers

According to Alzubaidi et al. (2021), the collection of convolutional filters that make up the convolutional layer is the most crucial component. Convolution is used to condense images into a format that is simpler for the computer to manage while simultaneously ensuring that no essential elements are lost for accurate prediction. According to Chandra, Desai, Guo & Balasubramanian (2020), CNN convolves the whole image with different kernels, generating various feature maps. According to Bhatt, Patel, Talsania, Patel, Vaghela, Pandya, Modi & Ghayvat (2021), the convolutional layer utilises a 3x3 or 5x5 filter to process an image input, as illustrated in Figure 2.4. To extract features, a filter that is configured on the input layer is used in the convolution process, producing a feature map.

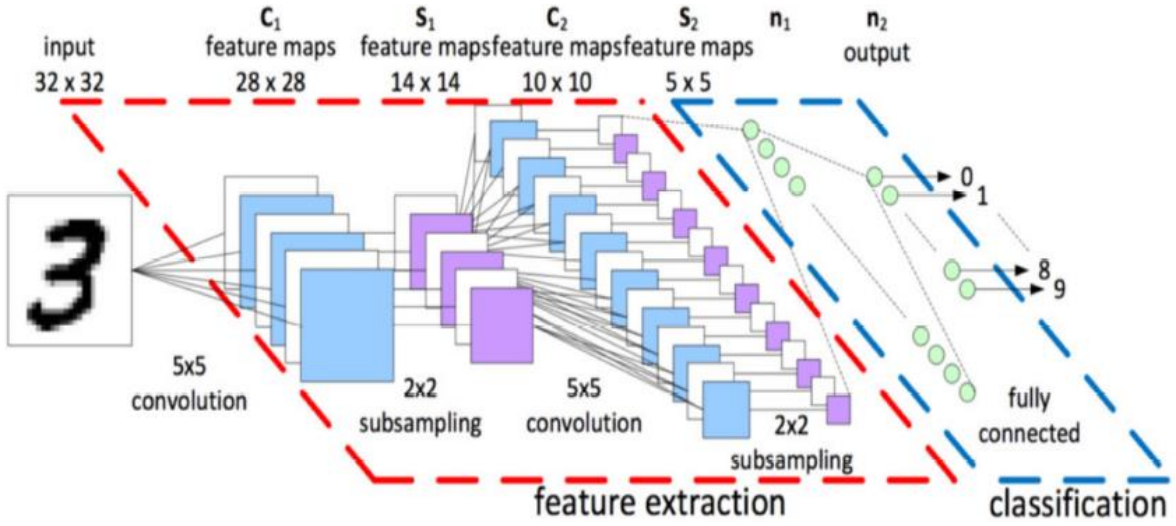


Figure 2.4: Convolution layer (Bhatt et al., 2021)

ii. Pooling Layers

The pooling layers aim to slowly shrink the representation (O'Shea & Nash, 2015). According to Bhatt, Patel, Talsania, Vaghela, Pandya, Modi & Ghayvat (2021), dimensionality reduction through pooling, either maximum or average, reduces the computer power needed to process data. Additionally, pooling helps prevent over-fitting and reduces training time (Bhatt et al., 2021).

Pooling layers are a type of layer commonly used in CNNs, which are a class of deep learning algorithms designed for processing grid-like data such as images. CNN pooling layers are used to down-sample the input data and reduce its dimensionality, which has several advantages.

Pooling layers in a CNN, offer several benefits, including a reduction in the network's computational complexity. By decreasing the number of parameters that the network needs to

learn, training becomes more efficient and faster. Moreover, pooling layers can help to prevent overfitting, which happens when a model performs well on the training data but poorly on unseen data.

iii. Activation Function

Activation functions are essential in mapping inputs to outputs in neural networks (Alzubaidi et al., 2021). These functions can be classified into two types: linear and non-linear activation functions. The ability to differentiate is a crucial requirement for an activation function (Alzubaidi et al., 2021). Non-linear activation functions are preferred as they enable the model to create complex mappings between the network's inputs and outputs (Bhatt et al., 2021). The most commonly used activation functions in CNNs and other deep neural networks is the one listed below (Alzubaidi et al., 2021).

- **Sigmoid**: The sigmoid function has a limited input domain of real numbers and an output range of zero to one. Alzubaidi et al. (2021), described the s-shaped curve of the sigmoid function using the following equation:

$$f(x)sigm = \frac{1}{1+ e^{-x}} \quad (2.1)$$

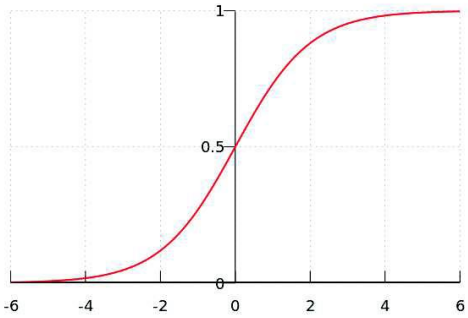


Figure 2.5: Sigmoid function (Airola et al., 2017)

- **Tanh**: The Tanh function is similar to the sigmoid function in that it accepts a real number as input, even though its output is limited to the range between -1 and 1 (Alzubaidi et al., 2021). The following equation can be used to represent the Tanh function:

$$f(x)\tanh = \frac{e^{x}- e^{-x}}{e^{x}+ e^{-x}} \quad (2.2)$$

- **ReLU:** The ReLU function converts all input values to positive numbers and applies this process to each of the filtered images. This function requires less calculation

compared to other functions, as noted by Alzubaidi et al. (2021). The following equation can be used to represent the ReLU function:

$$f(x)ReLu = \max(0, x) \quad (2.3)$$

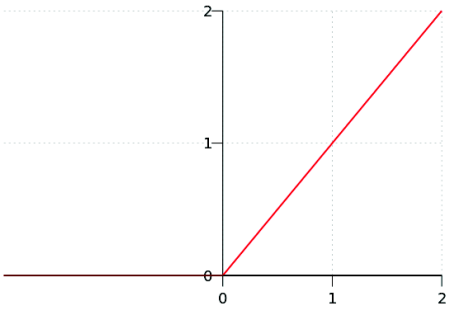


Figure 2.6: ReLu function (Airola et al., 2017)

iv. Fully Connected Layers

The layer of neurons is connected to the neurons in the two adjacent layers, as stated by O'Shea and Nash, (2015). Bhatt et al. (2021) explain that these fully connected layers are located at the bottom of the network. This layer is responsible for converting the 2D feature maps into a 1D feature vector (Chandra et al., 2020).

2.3.3.2 CNN architectures

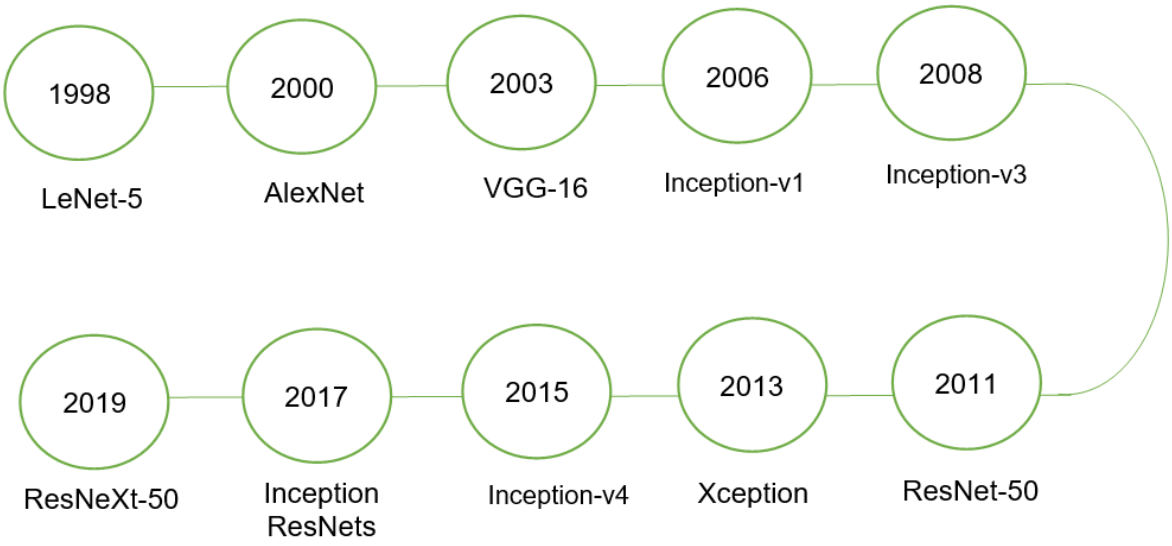


Figure 2.7: CNN architectures timeline (Source: Researcher)

i. LeNet-5

LeNet-5 is one of the earliest CNN architectures as depicted in Figure 2.7, introduced by Bengio (1998). At that time, it trained with the backpropagation algorithm for handwritten digit

recognition tasks. Khan, Anabia, Umme & Aqsa (2020) reported that LeNet-5 comprises seven trainable layers, which include two convolutional layers, two fully connected layers, and an output layer with a Gaussian connection. CNN may be said to have started with LeNet-5, which details its basic building blocks. Zahangir et al. (2018) claim that CNNs gained popularity as computer hardware's capability began to advance.

ii. **AlexNet**

In the past years, various CNNs have been suggested, including AlexNet, which was created by Krizhevsky, Sutskever & Hinton (2012), and emerged as the winner of the ImageNet large-scale visual recognition challenge in the same year. It gained an advantage because the top-5 error fell from 26% to 15.3%. According to Wang (2017), AlexNet's triumph can be attributed to its unique architectural design as well as its innovative training mechanism. Krizhevsky et al. (2012), reported that the architecture comprises 5 Conv layers and 3 fully connected layers, containing a total of 60 million parameters.

AlexNet underwent training on two NVIDIA GTX 580 GPUs during the ILSVRC (ImageNet Large Scale Visual Recognition Competition), and the process took 5 to 6 days to complete. To enhance the applicability of CNN, AlexNet increased the number of feature extraction stages from five (as in LeNet) to seven and also incorporated a 1000-way softmax classifier as noted by (Alzubaidi et al., 2021)

iii. VGGNets

VGGNet by Simonyan (2015) was proposed by the Visual Geometry Group (VGG). It is one of the popular CNN architectures because it uses small-sized filters (Ghosh et al., 2019). Between each of the 33 convolutional layers, Simonyan and Zisserman (2015) used maximum pooling (maxpool) layers to decrease the volume size. This type of model is known for being easy to use and highly effective. Khan et al. (2020) suggest that boosting the depth of neural networks is beneficial for improving the network's ultimate performance. In ILSVRC 2014, VGGNet finished second with an error state of 7.32%. Figure 2.8 provides a graphic depiction of the VGG architecture of different layers depths (Simonyan & Zisserman, 2015).

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 2.8: VGG architectures layers (11,13,16 and 19) (Simonyan & Zisserman, 2015)

iv. ResNet

He, Zhang, Ren & Sun (2016) proposed ResNet, a residual network consisting of 34 layers. It addresses the problem of disappearing gradients in deep networks by using skip connections to learn residuals. ResNet outperformed the GoogLeNet Inception v3 in the 2015 ILSVRC image classification with an error rate of 3.6% (Khan et al., 2020). Figure 2.9 provides a graphic depiction of the ResNet residual block (Airola, Hager & Alfredsson, 2017)

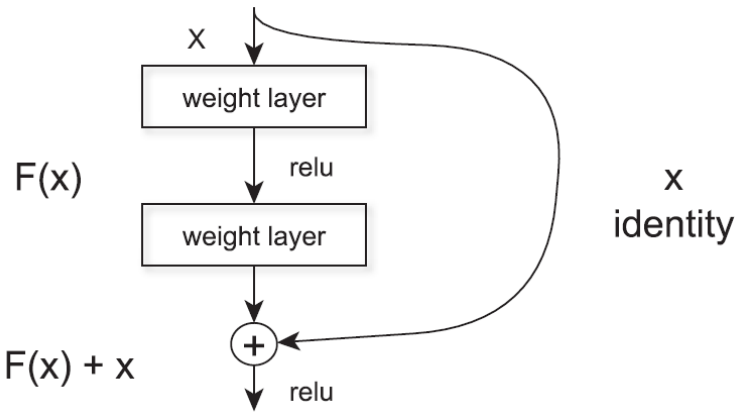


Figure 2.9: ResNet residual block (Airola et al., 2017)

2.3.4 Pre-trained model: Inception-V3

This model was developed by researchers at Google and introduced in the 2015 paper titled "Rethinking the Inception Architecture for Computer Vision" by Szegedy, Liu, Jia, and Sermanet. According to Reed, Anguelov, Vanhoucke & Rabinovich (2015), the InceptionV3 model is a deep CNN that was trained on the ImageNet dataset.

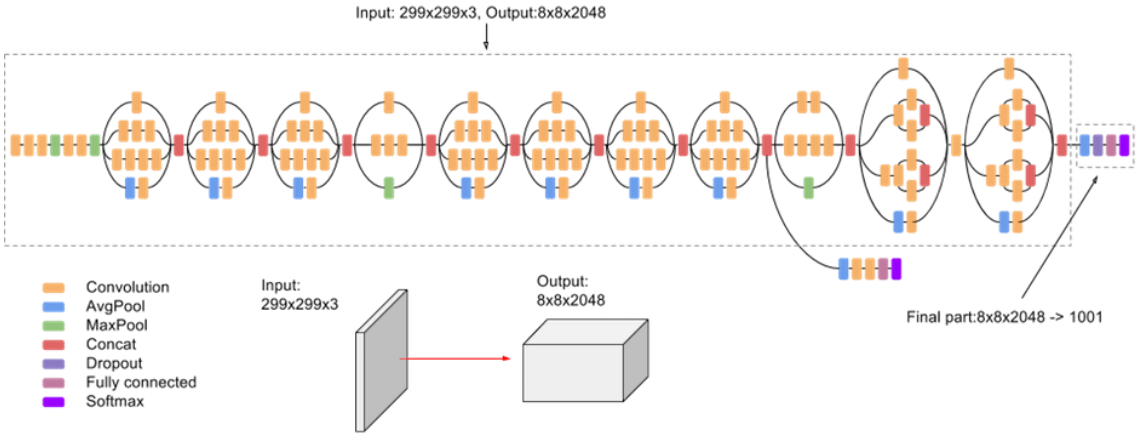


Figure 2.10: Architecture of the inception model (Biswas, Klingensmith & Umbaugh 2021)

The InceptionV3 model is built upon the Inception architecture, which was developed to address the issue of the increasing depth of CNNs leading to higher computational cost and decreased efficiency (Alzubaidi et al., 2021). The Inception architecture introduces the concept of modules, which are collections of convolutional and pooling layers that operate on different scales and at different resolutions (Khan et al., 2020). By enabling the model to acquire characteristics at various levels, it diminishes the number of parameters and computational expenses, while enhancing its efficiency.

Khan et al. (2020) stated that the InceptionV3 model is composed of a set of modules, where each module is comprised of a sequence of convolutional and pooling layers as depicted in Figure 2.10. The first module uses large filters to learn coarse-scale features, while subsequent modules use smaller filters to learn increasingly fine-scale features. The last module employs global average pooling to condense the acquired features and input them into a fully connected layer to classify.

One key innovation of the InceptionV3 model is the use of batch normalisation and rectified linear units (ReLU) for improved training and regularisation. Batch normalisation is utilised to normalise the inputs to each layer, which stabilises the distribution of activations and enhances the convergence of the model, as Rawat and Wang (2017) noted. Meanwhile, ReLU is employed as the activation function, which has been proven to enhance the performance of

deep learning models by providing a computationally efficient non-linearity that can learn complex functions, according to (Khan et al., 2020).

Khan et al. (2020) reported that the InceptionV3 model achieved state-of-the-art performance on the ImageNet dataset, with a top-5 error rate of 3.46% in the 2015 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) image classification task, surpassing the previous state-of-the-art result of 4.94%.

2.3.5 Pre-trained model: VGG-16

Simonyan and Zisserman (2015), from Oxford University, proposed the use of VGGNet. The ILSVRC object identification method won in 2014. The model attained an accuracy of 92.7% in the top-5 test by utilising ImageNet. It is one of the popular CNN architectures because it uses small-sized filters (Ghosh et al., 2019). The architecture of the vgg-16 model is depicted in Figure 2.11, as presented by Kogilavani, Prabhu, Sandhiya, Kumar, Subramaniam, Karthick, Muhibbullah & Imam (2022).

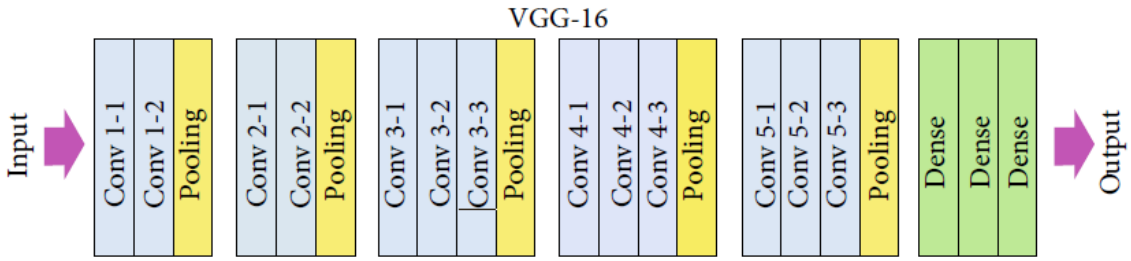


Figure 2.11: VGG-16 architecture (Kogilavani et al., 2022)

This type of model is known for being easy to use and highly effective. The model is composed of 13 conv layers with a filter size of 3 x 3, 3 dense layers with 4096 neurons each, and 5 pooling layers. Khan et al. (2020) suggest that boosting the depth of neural networks can be beneficial in improving the final performance of the network. After each convolution, a rectified linear unit (ReLU) action is applied to reduce the spatial dimension, and a max pooling operation is performed at the end of each block, as stated by Leonardo, Carvalho, Rezende, Zucchi & Faria (2019). Nevertheless, the model comes with a disadvantage in that it requires a greater expense for evaluation and a substantial amount of memory and parameters (Leonardo, Carvalho, Rezende, Zucchi & Faria, 2019).

2.4 Explainable AI

Explainable AI referrers to the methods and techniques applied within the field of AI, which enables human users to comprehend and trust the results created by the ML model (Mohseni et al., 2021). For the AI model to be applied in agriculture, it is essential to comprehend how the final results were arrived at. So that the user may trust in the prediction, the model needs

a minimum of interpretability. The use of AI in agriculture is complicated, for instance, due to the different plant types, and the various climate change and other interfering elements. According to Wei, Chen, Zhang, Fan, Wu, Liu & Chen (2022a), it is crucial to enhance the interpretability of these models.

Traditional assessment measures such as the classification report fail to sufficiently capture the underlying process undertaken by the AI system to reach a result and lack provisions for interpreting the outcome. To assist practitioners in effectively using DL techniques to address classification problems in the domain of agriculture, Wei et al. (2022a) investigated the application of XAI by employing DL models such as VGG, ResNet and GoogleNet for the identification of leaf disease classification.

2.4.1 Machine learning explainability

Over the past 20 years, ML has advanced quickly, which has resulted in the adoption of computerised decision-making systems (Lipton, 2018). According to Adadi (2018), interested researchers and practitioners can gain insight into the young and rapidly growing body of research related to XAI through their work, which serves as an entry point. Arrieta, Díaz-Rodríguez, Del Ser, Bennetot, Tabik, Barbado, Garcia & Molina (2020), emphasise the importance of comprehending how AI approaches are utilised to produce such results, particularly when automated decisions have an impact on individuals. Gerlings (2021) refers to explainable AI as methods and techniques that provide humans with understandable results/output that they can trust. The lack of transparency in the algorithm has emphasised the necessity for AI that can be explained. Gunning (2019), highlights the importance of trust in AI output, through the Defense Advanced Research Projects Agency (DARPA). The following Figure 2.12 explains the concept of XAI.

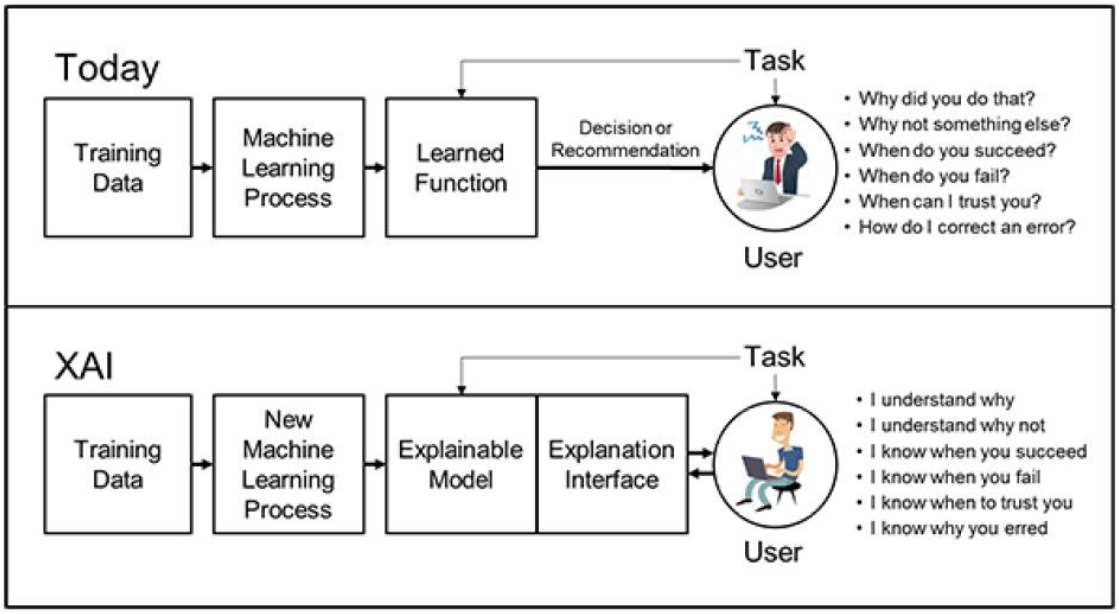


Figure 2.12: Concept of XAI (Rodrigo, 2020)

The misunderstanding and debate surrounding key ideas like interpretability, transparency, and explanation are some of XAI's main limitations (Kim, Atakishiyev, Babiker, Farruque, Goebel, Zaïane, Motallebi & Rabelo, 2021). The terms interpretability and explainability in the field of XAI are being used interchangeably because the field is not yet standardised (Gunning, 2019). Widely varying definitions have been used. Arrieta et al. (2020) define interpretability as the capacity to convey significance in terms that are understandable to humans. Interpretability is defined by Akata (2020) as an abstract concept translated into insights useful for domain knowledge. This section will clarify several taxonomies in XAI.

i. **Transparency:** If a model can understand itself, it can be considered transparent. Transparency can be divided into three distinct categories, overall model (simulatability), individual components (decomposability) and the complexity of the training process (algorithmic transparency) (Mohseni et al., 2021). Lipton (2018) defines a model as simulatable if it can be replicated by human calculation within a reasonable timeframe. Decomposability is concerned with the explainability of each part of the model. The final category, algorithmic transparency focuses on the user's ability to understand the whole model's process and to get the output from its input data (Lipton, 2018).

ii. **Interpretability:** It is the capacity of a model to be understood by humans. Interpretability can be categorised as transparency and post-hoc interpretability (Lipton, 2018). Humans can get explanations or the ability to comprehend terms simultaneously thanks to interpretability. Another factor to consider when designing for interpretability is the user's background knowledge.

iii. **Explainability:** The capacity to communicate the model's logic to a person in a way that they can understand it (Guidotti, Monreale, Ruggieri, Turini, Giannotti & Pedreschi 2018). Explainability can improve a user's trust in critical decision-making systems.

iv. **Understandability:** Denotes a user's understanding of how the model functions, without understanding its internal structure (Guidotti et al., 2018).

The aforementioned phrases are frequently used interchangeably by academics from various fields without considering their varying meanings. Arrieta et al. (2020), summarises the findings of XAI literature as follows: "Given an audience, an explainable Artificial Intelligence produces details or reasons to make its functioning clear or easy to understand."

2.4.2 XAI design goals with user groups

Mohseni, Zarei & Ragan (2021) address the three major user groups that XAI is intended to help and outline its design objectives. As illustrated by Mohseni et al. (2021), in Figure 2.13, there are eight design goals divided into three categories: data experts, AI experts, and AI novices.

i. **AI Novices**

According to Mohseni et al. (2021), social media and e-commerce end-users are examples of people who use AI products in their daily lives but have limited experience with ML systems.

ii. **Data Experts**

According to Mohseni et al. (2021), individuals who use ML for analysis or research and have expertise in the relevant domain are known as data experts.

iii. **AI Experts**

AI experts are ML scientists and engineers who design ML algorithms and interpretability for XAI systems, according to Mohseni et al. (2021).

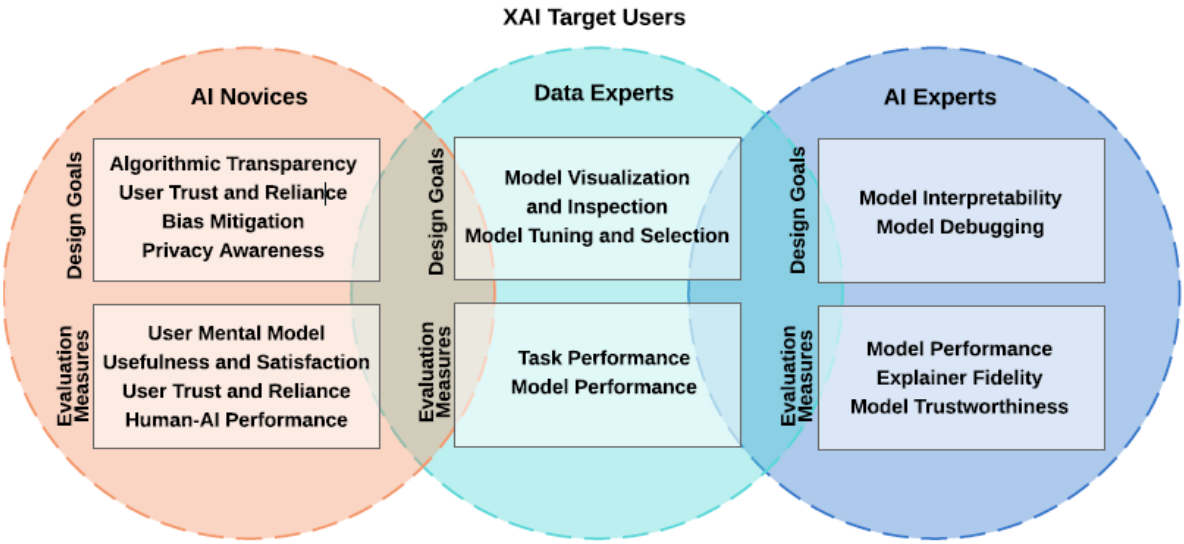


Figure 2.13: Design goals and evaluation measures (Mohseni et al., 2021).

2.4.3 Methods for XAI system

In their 2021 study, Mohseni et al. (2021) explore six prevalent categories of justifications used in XAI system designs, presented in Table 2.2, to gain a better understanding of the demands of designing XAI systems for different scenarios.

Table 2.2: Designing XAI systems for different scenarios (Mohseni et al., 2021).

Type of Explanation	**Description**	**approaches**
How Explanations	This exemplifies how the model functions.	Decision boundaries, model graphs by Lakkaraju, Bach & Leskovec (2016), and visual representations are used.
Why Explanations	Explains the rationale behind a prediction.	According to Ribeiro et al. (2016), Model-independent or model-agnostic are viable solutions.
Why-not Explanations	Mohseni et al. (2021), suggest aiding individuals in	Feature importance

	understanding why certain outputs were excluded from the system's output.	
What-If Explanations	Focuses on how different algorithmic parameters and inputs result in different outputs.	Tune model directly
How-to-Explanations	Mohseni et al. (2021) suggest explaining how altering the model or input could result in a distinct output.	Ad-hoc and model-agnostic
What-else Explanations	According to Mohseni et al. (2021), users should be given input examples that resemble the ones used by the model and produced comparable or indistinguishable outcomes.	Explanation by example

2.4.4 Local Interpretable Model-Agnostic Explanations (LIME)

The distinct types of explanation models include global methods, local methods, and introspective methods. Global methods help in understating how a model makes decisions for the overall structure. Additionally, we can explain the model's entire behaviour. In contrast, the use of local techniques can assist in understanding the decision-making process of the model for a specific instance.

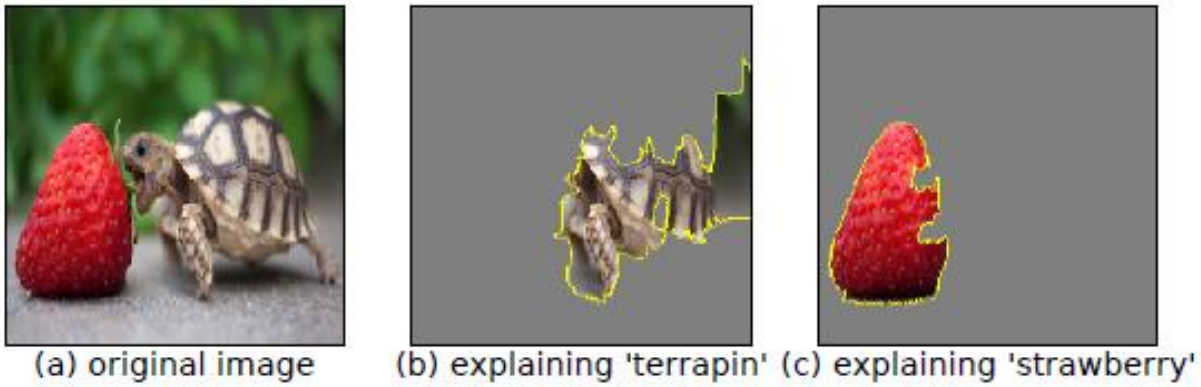


Figure 2.14: LIME explanation for object identification in images (Garreau & Von Luxburg, 2020)

Figure 2.14 shows how the LIME explanation is represented visually. Garreau and von Luxburg (2020) used Inception as the black-box model. The original image (a) shows a turtle and a strawberry. (b) outlines how this prediction was made by LIME. The superpixels with the top

five coefficients in the surrogate linear model are represented by the areas of the image that are highlighted. For (c) the same experiment was conducted for a strawberry.

Ribeiro et al. (2016) introduced LIME, a model-agnostic explanation technique that can be used post-hoc. LIME treats every model like a black box. It is independent of the original classifier which means it can be used for explaining any classifier. Misheva, Osterrieder, Hirsa, Kulkarni & Lin (2021) claim that Lime uses test data points to try to fit a local model. Mathematically, local surrogate models with interpretability constraints can be expressed as follows:

$$explanation(x) = \arg\min_{g \epsilon G} L\,(f, g, \pi_x) + \Omega\,(g); \qquad (2.4)$$

- Where g is the group of models that may be understood, including decision trees (Misheva et al., 2021).
- $g \epsilon G$: An explanation considered as a model (Misheva et al., 2021)
- f : Explaining the primary classifier.
- $\pi_x\,(z)$: Distance between instance z and instance x (Misheva et al., 2021)
- $\Omega\,(g)$: A metric for evaluation of how complicated an explanation is $g \epsilon G$.

Minimising the locality-aware loss L is the objective (e.g., mean squared error) without making any assumption about the original model (e.g., Radom Forest model).

2.4.5 Shapley Additive exPlanations (SHAP)

Lundberg and Lee (2017) developed a strategy called Shapley Additive exPlanations (SHAP). SHAP uses a feature-relevant explanation approach. Any ML model may be understood and explained using SHAP. DL architectures, among other complex ML approaches, exhibit a black-box nature. The foundation of SHAP values is Shapley values, a notion borrowed from the work of Linardatos, Papastefanopoulos & Kotsiantis (2021) in game theory. Singh et al. (2021), state that SHAP utilises Shapley values to provide an interpretation and explanation of any machine learning model, as shown in Figure 2.15.

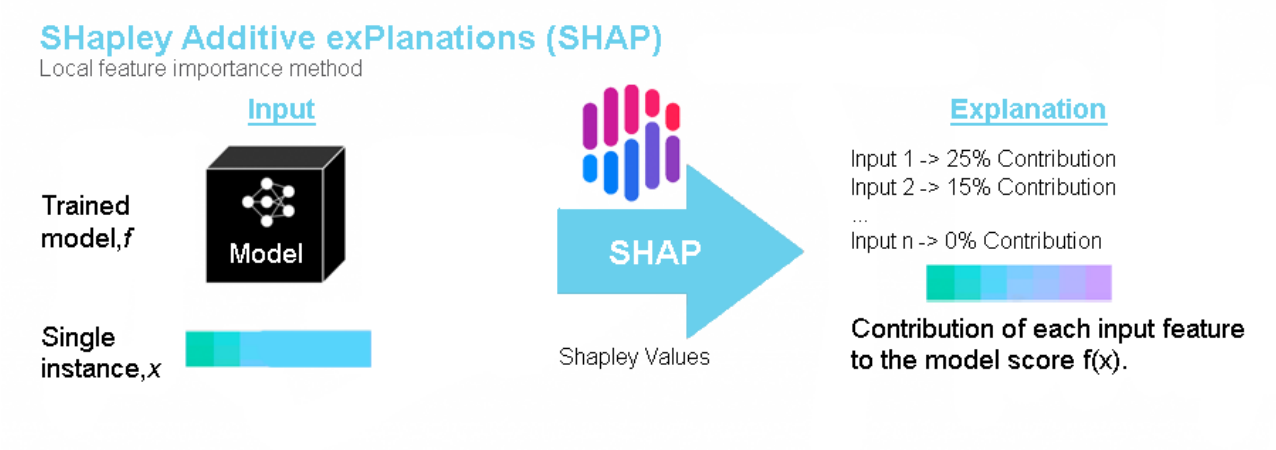


Figure 2.15: SHAP values (Knapič, Malhi, Saluja & Främling 2021)

SHAP is not concerned with the model we use. Today, SHAP is regarded as a fundamental advancement in the field of XAI. Shapley values have been demonstrated as the only approach that preserves two crucial characteristics, local accuracy, and consistency, for determining the relevance of a feature, as shown by Lundberg and Lee (2017).

i. **Local accuracy –** Antwarg, Miller, Shapira & Rokach (2020) state that the explanation model should, at the very least, replicate the outcomes of the original model.

$$f(x) = g(z') = \emptyset_0 + \sum_{j=1}^{M} \emptyset_j Z'_j \qquad (2.5)$$

- The model $g(Z')$ matches the original model $f(x)$ when $x = h_x(x')$ (Lundberg & Lee, 2017)

ii. **Consistency –** Changing a model to depend more on a feature should not decrease the significance of that feature, regardless of other features (Antwarg et al., 2020).

Let $f_x(z')$ = f $(h_x(z'))$ and $z'\backslash i$ denotes setting $z'_i = 0$. For any two models f and f' , if $f'_x(z') - f'_x(z'\backslash i) \geq f_x(z') - f_x(z'\backslash i)$

For all inputs $z' \in \{0,1\}^M$, then $\emptyset_i(f', x) \geq \emptyset_i(f, x)$ (Lundberg & Lee, 2017).

2.4.6 Gradient-weighted Class Activation Mapping (Grad-CAM)

To enhance the transparency of decisions made by CNN-based models, Selvaraju, Cogswell, Das, Vedantam, Parikh & Batra (2017) introduced a method called Gradient-weighted Class Activation Mapping (Grad-CAM) for presenting "visual explanations"

heatmap to understand the decisions made by these models. The fundamental motivation for the creation of Grad-CAM is to leverage the information present in the convolutional layer to identify the significant parts of image classification. According to Panesar (2022), Grad-CAM combines feature maps using gradients such that the original network does not need to be changed.

In Figure 2.16 the Grad-CAM technique is employed to classify a dog and cat. (a) The original image of a dog and cat, (b,c) an illustration of a dog image produced using the Grad-CAM method. This method highlights the areas of the input image that the model focuses on during the classification process. The heatmap's important regions are highlighted in red, while unimportant regions are highlighted in blue (Setiawan & Rulaningtyas, 2023). Moreover, according to Panesar (2022), this suggests that the feature maps created in the final convolutional layer contain the spatial information necessary for identifying the visual pattern.

Grad-CAM heatmaps are becoming a popular visualisation technique in the agriculture sector. For instance, Xu, Tan, Zhang, Zha, Yang & Yang (2022) used a Grad-CAM-based approach to improve maize seed classification, while Morshed, Ahmed, Ahmed, Islam & Rahman (2022) employed the same approach to gain insights into the fruit quality assessment. Additionally, Chopra and Whig (2021) identified tomato leaf disease using the same method. These studies demonstrate the versatility of Grad-CAM in agriculture and their potential to improve crop yield and quality assessment.

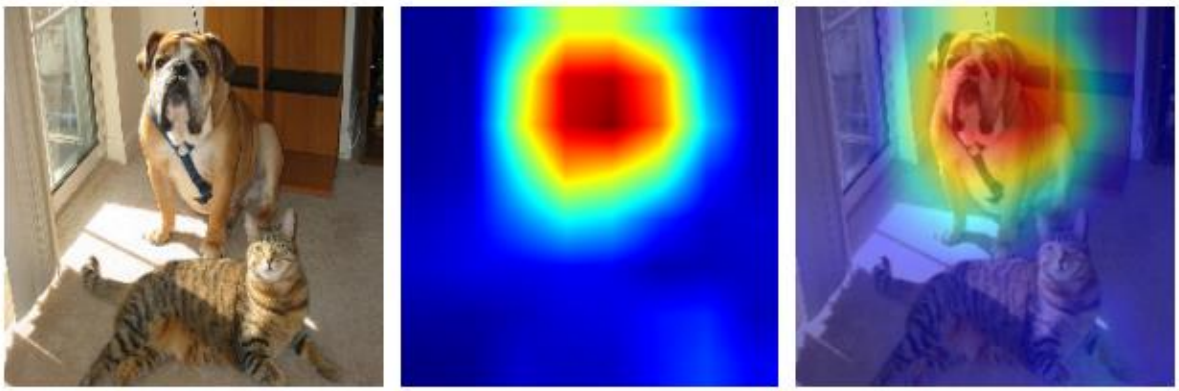


Figure 2.16: An illustration of a dog image produced using the Grad-CAM method

According to Panesar (2022), the importance weights for each position in the feature map are obtained by globally averaging the gradients that flow into the last convolutional layer across the width and height dimensions represented by i and j , using the formula (Panesar,2022):

$$\propto \frac{c}{k} = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}} \quad (2.6)$$

$\propto \frac{c}{k}$ = important weight of class c gradient k

Z = Dimension (i x j)

y^c = Score of class c

A^k_{ij} = vector (i, j) of feature map k

$\frac{\partial y^c}{\partial A^k_{ij}}$ = Derivative of gradient score class c for feature map k

2.4.7 Explainable Artificial Intelligence (XAI) Evaluation Measures

The understandability of an ML system is believed to be influenced by two main factors, the characteristics of the system itself and the human's ability to comprehend it (Zhou, Gandomi, Chen & Holzinger, 2021). Various methods have been suggested for assessing the quality of an explanation. Figure 2.17 illustrates the three categories in which evaluations of explainable ML can be classified.

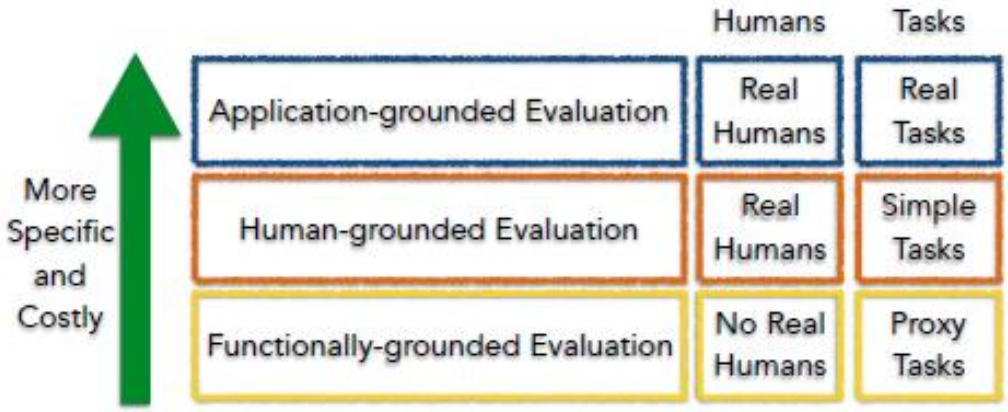


Figure 2.17: Taxonomy of Interpretability Evaluation (Doshi-Velez & Kim, 2017)

i. **Functionally grounded**: Human experiments are not necessary for functionally grounded evaluation. Doshi-Velez and Kim (2017) propose a formal definition of interpretability as a substitute for the quality of an explanation. This type of assessment is appropriate once a class of models or regularising has been validated, for example, through human-grounded experiments (Doshi-Velez & Kim, 2017).

ii. **Application-grounded:** Human experiments within a practical application are conducted as part of the application-grounded evaluation (Bodria, Giannotti, Guidotti,

Naretto, Pedreschi & Rinzivillo, 2021). They are frequently used in particular contexts. For instance, if the model is a tool to help doctors make decisions, the doctors conduct the validation (Doshi-Velez & Kim, 2017).

iii. **Human-grounded:** Conducting more basic human-subject experiments while maintaining the core functionality of the intended application is known as human-grounded evaluation (Doshi-Velez & Kim, 2017). Normal people can conduct evaluations, expanding the pool of potential test subjects. Moreover, this validation is most suitable for general testing ideas about the quality of an explanation (Bodria et al., 2021).

2.5 Evaluation Criteria

There are several measures used to evaluate the performance of DL models, such as Precision, F1-score, and accuracy. Through the use of multiple evaluation metrics, the study achieved an understanding of the model's performance and made informed decisions about its effectiveness. Table 2.3 shows the equations for the evaluation metrics.

Table 2.3: Evaluation Criteria

	Evaluation focus	**Formula**
Precision	It focuses on what proportion out of all predictions was correct.	$Precision = \frac{tp}{tp+fp}$ (2.7)
Accuracy	It is used to measure the percentage of correct classifications (Liu, Zhou, Wen & Tang, 2014).	$Accuracy = \frac{tp+tn}{tp+tn+fp+fn}$ (2.8)
F1 Score	The harmonic mean of precision and recall is represented by it (Liu et al., 2014).	(2.9) $f1 = \frac{2}{recall^{-1} + precision^{-1}} = 2 * \frac{precision * recall}{precision + recall} = \frac{tp}{tp + \frac{1}{2}(fp + fn)}$
Recall	Describes the precision of a model, which is the ratio of true positives to all predicted positives.	$Recall = \frac{tp}{tp+fn}$ (2.10)

Explanation of the terms associated with the evaluation parameters:

- True Positive (TP): The model correctly predicts that the image contains the nutrient.
- True Negatives (TN): The model correctly predicts that the image does not contain the nutrient.
- False Positives (FP): The model incorrectly predicts that the image contains the nutrient when it does not.
- False Negatives (FN): The model incorrectly predicts that the image does not contain the nutrient when it does.

There are different ways to interpret the F1-Score and AUC Score. This study followed the general rule by Allwright (2022) :

Table 2.4 F1-Score interpretation

F1-Score	Interpretation
>0.9	Very good
0.8 – 0.9	Good
0.5 – 0.8	Ok
<0.5	Not good

Table 2.5 AUC Score Interpretation

AUC Score	Interpretation
> 0.8	Very good performance
0.7 – 0.8	Good performance
0.5 – 0.7	Ok performance
0.5	As good as a random choice

2.6 Cross-validation (CV)

Stone (1974) introduced cross-validation, which has since become a widely used statistical tool for estimating the generalisation risk of learning algorithms. Accurate evaluation of a model's performance is vital in ML, enabling informed decisions about its effectiveness. It serves not only as an estimator of generalisation performance but also as a means to uncover issues related to selection bias and overfitting (Soper, 2021).

F-fold cv is a commonly employed technique for model selection, where the final model is selected based on achieving the minimum cross-validation error, as demonstrated by Zhong, He & Chalise (2020). According to Nti, Nyarko & Aning (2021), in k-fold CV, the dataset is partitioned into k subsets called folds. During each iteration, one-fold serves as the testing

data, while the remaining folds are used for training, as depicted in Figure 2.18, providing an overview of the k-fold CV process.

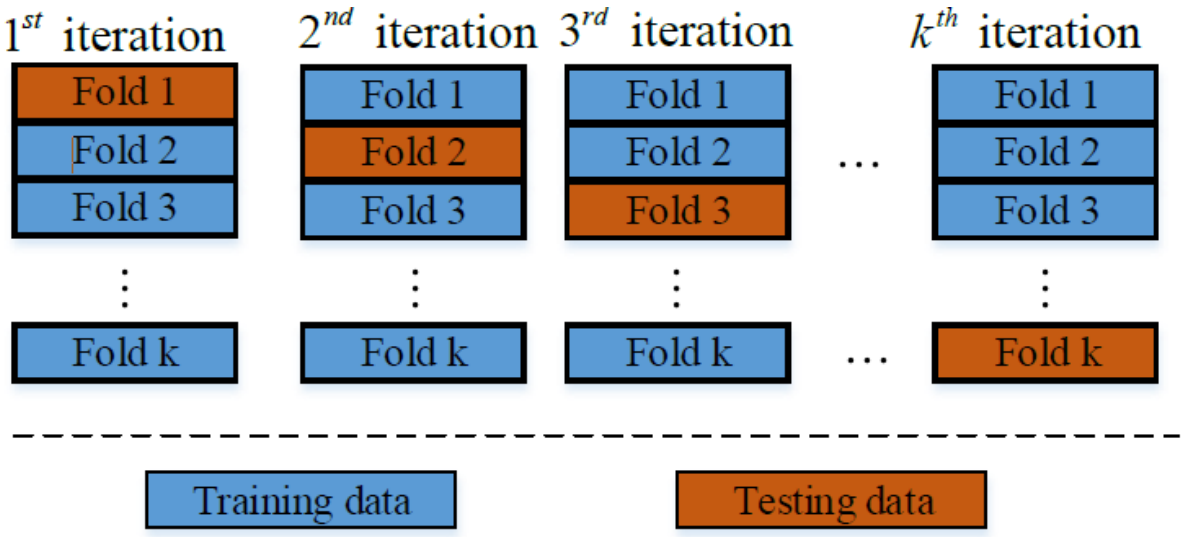


Figure 2.18: K-Fold Cross-Validation (Nti et al., 2021)

2.7 Related work

The following section provides an overview of the use of deep learning for plant nutrient deficiency identification.

Watchareeruetai, Noinongyao, Wattanapaiboonsuk, Khantiviriya & Duangsrisai (2018) proposed a method that uses CNNs to identify plant nutrient deficiencies, including Mg, Ca, and N, among others. The proposed experiment was conducted on a set of black gram plants. To assess how well the experiment performed, the researcher compared the results with two trained humans. Compared to the two people's accuracy performance (27.37%), the proposed model performed with 43.02% more accuracy. The experiment revealed that if there are diverse types of nutrients to identify, it becomes difficult to accurately identify the nutrients.

Tran, Choi, Le & Kim (2019) accurately predicted nutrient deficiencies to increase crop production and prevent tomato pathology caused by a lack of nutrients such as Calcium, Nitrogen, and Potassium. Noon, Amjad, Qureshi & Mannan (2020) conducted a review of 45 DL techniques proposed for 33 different crops using 14 well-known convolutional neural architectures. Meanwhile, Xu, Guo, Zhu, He, Zhao, Han & Subedi (2020) explored the accuracy of various DCNNs for diagnosing nutrient deficiencies in rice. Abade, Ferreira & Vidal (2020) conducted a systematic review of the literature to determine the current state of the art in using CNNs for identifying and classifying plant diseases. Their findings demonstrate the potential of DL, particularly CNNs, in the field of agriculture and plant disease detection, while also highlighting trends and gaps in the research.

The performance of palm leaf nutrient deficiency detection using CNN was evaluated by Ibrahim, Hasan, Sabri, Abu & Rahimi (2022) through a confusion matrix. The results showed an overall mean accuracy of 94.29%, a sensitivity of 80%, and a specification of 96.67%.

Another vision-based monitoring system was proposed by Claudio (2020). The experiment aimed to identify tomato plants that were deficient in the main nutrients such as nitrogen. Furthermore, the implementation of such systems can increase the quantity and quality of plants. The proposed CNN model had an accuracy rate of 86.57%. Zeng, Song, Li, Chusap & Liu (2021) introduced a human-in-the-loop method that utilises verbatim neighbourhood manifestation to explain ML models. Similarly, Veldhuis, Ariëns, Ypma, Abeel & Benschop (2022) draw on XAI techniques in their 2022 study to aid users in comprehending the reasoning behind particular predictions.

Rodrigo (2020) addressed the issue of CNN-based solutions' lack of explainability by utilising the CNN model and YOLOv3 to generate easily understandable outcomes for nutrient deficiencies. During the study, the model was trained to detect nutrient deficiencies in lettuce, specifically nitrogen, magnesium, and calcium. The training data resulted in a mean average precision of 94.38%. In a similar vein, another author Humaion, Salman, Shafi, Samanta, Rabeya & Mehedi (2022) used three pre-trained CNN models to detect plant leaf disease. Among the three models, EfficientNetV2L demonstrated a remarkable accuracy of 99.63%. To explain the predictions made by EfficientNetV2L, the author used the classification XAI framework LIME.

In their 2022 study, Wei et al. (2022) delved into the interpretability of DL models across various agricultural classification tasks. Their findings indicate that the ResNet model outperformed all other models with an accuracy of above 99.11% in all experiments conducted. Additionally, the study compared three visualisation methods and concluded that the GradCAM method is the most appropriate for agricultural classification tasks.

Rakesh and Indiramma (2022), focus on the use of the DL model with XAI for crop disease detection. The experimental results show that both Inception-V3 and ResNet-9 achieved high accuracy, 98.18% and 99.2% respectively. The study showcases the potential of XAI in enhancing the transparency of DL models. Additionally, the author employs XAI tools such as LIME and Grad-CAM to offer visual interpretations for the model's predictions.

Ghosal, Blystone, Singh, Ganapathysubramanian, Singh & Sarkar (2018) show the capability of DCNN to recognise and categorise a variety of foliar stressors in soybean. By studying more than 25,000 photos, the researcher was able to accurately identify and categorise several diseases and nutritional deficiencies. The availability of a reliable and fast explainable model would significantly impact both scientific research and agricultural production. To evaluate the

nitrogen content of wheat, the researchers Singh, Roy, Setia & Pateriya (2021) suggested using machine learning and XAI. Six machine-learning regression models were used by the author to forecast the nitrogen of wheat. The gradient-boosting regression and random forest combination outperformed (R^2> 0.85) previous combinations. To provide local and global explanations, SHAP values were used. According to Singh et al. (2021), based on the findings, it appears that using XAI as a tool to explain machine learning models is a promising idea.

Summarily, this study distinguishes itself from the previous related work in various aspects. Table 2.4 presents a comparison of the classification performance between the study and other ML models that used the same two datasets as this study. The related work mostly focuses on one plant type such as tomato and maize. This study expands on the limitations of previous studies by looking at a more diverse range of plants. By doing so the study provides a more comprehensive analysis.

Additionally, most of the previous research assesses the accuracy of the DL models, but they do not consider the explainability of the models. While some related work considers XAI techniques, they do not compare the different XAI techniques to understand how these models arrive at their predictions. Therefore, an integrated approach that considers both performance and explainability is lacking in the existing body of work. This study seeks to fill the gap by comparing the explainability of two prominent XAI techniques, GRAD-CAM, and Shapley Additive exPlanations.

Table 2.6 Compares the accuracy of the two datasets.

Previous Studies	**Dataset**	**Accuracy %**	**Model**	**XAI**
Han, Maneerat, Sepsirisuk & Hamamoto (2023)	Banana	87.89%	ConvNextTiny	No
Sathyavani, Jaganmohan & Kalaavathi (2023)	Rice	96.13%	Xception	No
Simul, Talukder & Sarkar (2022)	Rice	86.66%	InceptionV3	No

2.8 Chapter Summary

This chapter presented literature related to explainable DL models for the identification of plant nutrient deficiencies. The research looks at the importance of understanding the relationship between plant nutrient deficiency and plant diseases. It introduces the significance of DL and how they are applied in the agriculture sector. Furthermore, the research highlights the importance of XAI, and it delves into different XAI tools such as LIME, SHAP, and GRAD-CAM

Table 2.7: Summary of research literature

Title	**Plant**	**Nutrient(s)/ Disease(s)**	**Architecture**	**Best Accuracy**	**Research Gap**
Watchareeruetai,Noinongyao, Wattanapaiboonsuk, Khantiviriya & Duangsrisai (2018)	Black gram	Nutrients: Calcium, phosphorus, iron, potassium, magnesium, and nitrogen	AlexNet, ResNet50, MobileNet, MobileNet-v2, Xception, Inception-v3, and VGG16	65.44%	-The study used only one plant, which could affect the generalisation of the results. -Lack of model explainability
Tran, Choi, Le & Kim (2019)	Tomato	Nutrients: Calcium, nitrogen, and potassium	CNN	87.27%	-Limited data set -Lack of model explainability
Xu, Guo, Zhu, He, Zhao, Han & Subedi (2020)	Rice	Nutrients: Nitrogen, manganese, calcium, magnesium, potassium,	DCNN: Dense Net, ResNet, Inception-v3, and NasNet-large	97.44% - DenseNet121	-The study used only one plant, which could affect the generalisation of the results.

		phosphorus, zinc, iron, and Sulfur.			-Limited data set; the study used a total of 1818 images -Lack of model explainability
Claudio (2020)	Tomato	Nutrients: Nitrogen, Phosphorus, and potassium	CNN	86.57%	-There is no comparison with other models. - The study used only one plant (tomato) -Lack of model explainability
Rodrigo (2020)	Lettuce	Nutrients: Calcium, Nitrogen, and Magnesium	CNN	94.38%	-There is no comparison with other models. - The study used only one plant (tomato)

Singh, Roy, Setia & Pateriya (2021)	Wheat	Nutrients: Nitrogen	Six regression models (i.e., Random Forest) XAI tool: SHAP	(R^2 = 0.89) Random Forest	-The study used only one plant and one nutrient, which could affect the generalisation of the results. -The study does not compare different XAI approaches
Sabri, Kassim, Ibrahim, Roslan & Abu (2020)	Maize	Nutrients: Nitrogen, Magnesium, and Potassium	Random Forest	78.35%	-There is no comparison with other models. -Limited data set. -Lack of model explainability.
Ghosal, Blystone, Singh, Ganapathysubramanian, Singh & Sarkar (2018)	Soybean	Nutrients: Potassium. Diseases: Frog-eye leaf spot, bacterial blight,	DCNN	94.13%	-There is no comparison with other models.

		Septoria brown spot, and herbicide injury			
Wei, Chen, Zhang, Fan, Wu, Liu & Chen (2022)	Mixed	Diseases: 34 categories of fruit leaf diseases	VGG, GoogLeNet and ResNet XAI tool: Grad-CAM, SmoothGrad and LIME	99.89% Resnet	-Limited data set
Rakesh and Indiramma (2022)	Mixed	Diseases: Seventeen diseases caused by fungi	Inception-V3 and ResNet-9 XAI tool: LIME and Grad-CAM	99.2% ResNet-9	-Limited data set
Humaion, Salman, Shafi, Samanta, Rabeya & Mehedi (2022)	Mixed	Diseases: Thirty-eight types of leaf diseases	EffecientNetV2L, MobileNetV2, and ResNet152V2 XAI tool: LIME	99.63% EfficientNetV2L	-The study does not compare different XAI approaches

CHAPTER THREE

RESEARCH METHODOLOGY

This chapter presents the research methodology employed for this study. This chapter covers the research approach, methodological choice, research strategy, and research design. The study's ethical considerations are also elaborated on in this chapter.

3.1 Research Approach

According to Creswell (2014), research approaches encompass a range of strategies and methods used to conduct studies, which can vary from general hypotheses to specific techniques for collecting, analysing, and interpreting data. Two variables must be considered while choosing a research approach: the topic's features and the available time for the study (Pandey & Pandey, 2016). According to Jonathan (2010), there are two widely used research approaches in research, inductive and deductive. However, a hybrid method known as abductive is also available. Different methods of approach can be used to acquire knowledge that does not yet exist.

3.1.1 Deductive approach

According to Jonathan (2010), a deductive approach is the process of creating a hypothesis based on an established theory. Saunders, Lewis & Thornhill (2007) identified several significant qualities that are encompassed by the deductive approach. The effort to understand the causes of the correlations between variables comes first. To fulfil the objective of the research, a deductive research approach was chosen. To achieve the aim of this study, the use of a structured and systematic approach to test pre-defined hypotheses based on existing theories and literature is necessary. The deductive approach works from general to specific. The researcher first examines existing theories related to the phenomenon under investigation and then evaluates the hypotheses that arise from those theories. As a result, this study employed the deductive approach.

3.2 Quantitative research methodology

According to Aramay (2014), methodology refers to the way the researcher deploys the whole research design. The construction of the research methodology is explained by Saunders, Lewis & Thornhill (2019), through their theoretical concept of the research onion. The research methodology outlines the procedures and techniques utilised to determine the data to be

collected, and to conduct data processing and evaluation. Qualitative, quantitative, or mixed-method research methodologies are the three categories that research methods fall under (Daniel, 2016).

Saunders et al. (2007) assert that quantitative research is centred on measuring quantity or amount. It employs analytic tools like graphs or diagrams to provide accurate results. Quantitative research, in contrast to qualitative research, is frequently linked to a deductive approach (Jonathan, 2010). The primary focus of quantitative research is on the “when”, “where”, “what”, and how frequently a particular phenomenon occurs. It is vital to remember that this method is crucial since it enables generalisation from a sampling of populations. Based on these facts Håkansson (2013) states that experiments and testing are supported by quantitative research methods.

The study's objectives will be achieved by using a quantitative research methodology. The study involves the use of collecting and analysing numerical data. The quantitative research methodology will make it easier to measure and compare the performance of the three DL models using standardised metrics. Moreover, to validate the study’s findings and conclusions, it is crucial to use a quantitative methodology that enables other researchers to reproduce the findings.

3.3 Research Strategy

McNabb (2010) describes the research strategy as a way of collecting data and how we produce answers to the research question. According to Creswell (2014), the strategy offers a precise orientation for the methods used in a research plan. Saunders et al (2007) state that there are several research strategies such as surveys and experiments. Therefore, Johannesson and Perjons (2014) suggest that a research strategy needs to be appropriate for the task at hand, otherwise, it will not be able to assist the researcher in answering the issue they are trying to answer. Surveys and experiments are the key tools used in quantitative methodology. According to Pandey (2016), experiments are typically employed to address “how” and “why” questions. As a result, an experimental research strategy has been selected for this study.

Kapur (2018) states that experimental research uses two sets of variables and a scientific approach to conduct the study. Two essential characteristics define experiments. The first involves the systematic manipulation of the independent variable’s level by the researchers. The researcher controls the experiment, which is its second component. The objective is to change the independent variable and then observe how it changes the dependent variable (Jonathan, 2010). According to Saunders et al. (2007), it is necessary for an experiment to

clearly define its variables to enable readers to comprehend which groups are undergoing the experimental process and which outcomes are being monitored.

3.4 Experimental Design of the Study

Research design, according to Robinson (2015), is the road map for data collection, measurement, and analysis. Robinson (2015) states that there is not a single design that works in every situation. Each researcher must develop his or her design. Naturally, a variety of factors will influence the design(s) you choose. These include your research goal, its timetable, its methodology, and the data's accessibility (Håkansson, 2013). Figure 3.1 illustrates the experimental research design of this study. According to McNabb (2010), Experimental design (ED) is a process of designing, conducting, and evaluating the results of all types of experiments.

Research design is necessary because it makes it easier for various research activities to go smoothly, which maximises the amount of information that can be obtained with the least amount of time, effort, and money (Kothari, 2004). Additionally, according to Robinson (2015), a significant factor affecting the validity of the obtained result is the research design. However, how any of these can be accomplished largely relies on the research purpose.

The experimental design was selected for this study because it is essential for the development and implementation of ML algorithms and models. By carefully planning, implementing, and analysing the results of experiments, researchers can ensure the validity and reliability of their findings and can make confident conclusions about the research question or hypothesis under investigation.

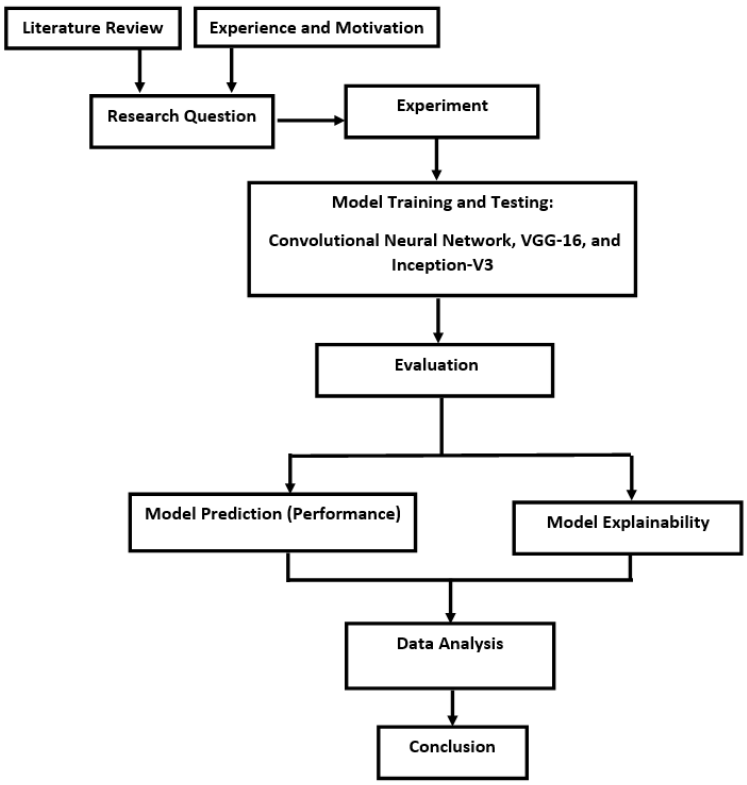


Figure 3.1: Overview of the experimental design of this study (Source Researcher)

3.5 Ethical Considerations

McNabb (2010), refers to ethics as a part of philosophy, which focuses on the moral behaviour of humans in society. The research ethics guidelines of the Cape Peninsula University of Technology, South Africa were followed during the study. The study was conducted using publicly available datasets, adhering to their respective terms and conditions.

The study did not interfere with the procedures for gathering and analysing data. Another consideration in the study was to accept the terms and conditions for using Python software and all its libraries. The experiment did not make use of external participants; therefore it did not need to do informed consent enrolment.

3.6 Chapter Summary

This chapter provided an overview of the research methodology used for this study. It included the research approach, methodological choice, research strategy and research design. Additionally, this chapter covers the ethical considerations taken into account to ensure that the research process was conducted appropriately.

CHAPTER FOUR

EXPERIMENTATION

The procedures used for the experiments that produced the findings of this study are described in this chapter. This chapter presents details of the models' configurations and the dataset that was used.

4.1 System infrastructure

Table 4.1 shows the hardware setup that was used to perform experiments on the selected machine-learning models:

Table 4.1 Workstation setup

Operating System	**Windows 10 Enterprise**
CPU	Intel Core i7-1185G7
System Type	64-bit OS
GPU	Intel(R) Iris(R) Xe Graphics (7.8GB)
RAM	16.0 GB
Environment	Google Colab

4.1.1 Software and Frameworks

Google Collaboratory (Google Colab), served as the integrated development environment (IDE) for the experimental process. Google Colab is a free cloud-based, setup-free Jupyter Notebook environment. It is a web-based platform that enables users to create and execute code, save, and distribute their findings, and utilise robust computing resources. The Google Cloud Platform hosts the Google Colab infrastructure (Carneiro, Nobrega, Nepomuceno, Bian, Albuquerque & Filho, 2018). The study used the normal Colab version, which provides access to a GPU, to satisfy the intensive computing needs of neural networks. Python (version 3.9.19) was used for the code development and implementation. The following is a discussion of the software environment, which comprises frameworks and the main Python libraries that were employed in the experimental process, Table 4.2 highlights the library version used.

TensorFlow: The Google Brain team, a group of researchers and engineers within Google, developed an open-source library called TensorFlow to research ML and DNN. As per Pang, Nijkamp & Wu (2020), TensorFlow is the most extensively used deep-learning library among

several others. It only accepts tensors, which are multi-dimensional arrays with more dimensions, as the type of data.

NumPy: Numerical Python is known as NumPy. It is a fundamental package for working with arrays (Hao & Ho, 2019). Due to its efficiency and speed, it is ideal for calculation in science or mathematics.

Pandas: It is a NumPy-based open-source library that is widely utilised for data analysis and machine learning purposes.

Matplotlib: Matplotlib, the Python visualisation library, is not only fantastic but also simple to use. It consists of a variety of plots and is built on NumPy arrays, making it compatible with other SciPy stack libraries.

PyTorch: It is an open-source machine learning framework that utilises the Torch library and is implemented in Python. It was created in 2017 by the Facebook AI research team. PyTorch has strong GPU computation. As a result, people frequently choose it for fast experimentation.

Scikit-learn (Sklearn): Sklearn is the most comprehensive machine learning package. It supports Python numerical libraries such as NumPy and includes a variety of algorithms like SVM and RF.

Table 4.2: Libraries used

Library	**Version**
Pandas	1.3.5
NumPy	1.21.6
Matplotlib	3.2.2
TensorFlow	2.9.2
Cv2	4.6.0

4.2 Dataset

According to Kairuz, Crump & Brien (2007), the researcher chooses a method of data collection and considers the strengths and limitations when evaluating the results. For this study, two open-access datasets were used. Sample images from the different datasets are shown in Figure 4.1

Figure 4.1: Images from the two datasets (a) Banana, (b) Rice

4.2.1 Dataset Details

i. The rice dataset

The dataset is available on the Kaggle data science community (https://www.kaggle.com/datasets/guy007/nutrientdeficiencysymptomsinrice). There are 1156 images of rice in this dataset each varying in size. Showing three distinct types of nutrient deficiency: Potassium (K), Nitrogen (N), and Phosphorus (P). The distribution of the dataset classes is depicted in Figure 4.2

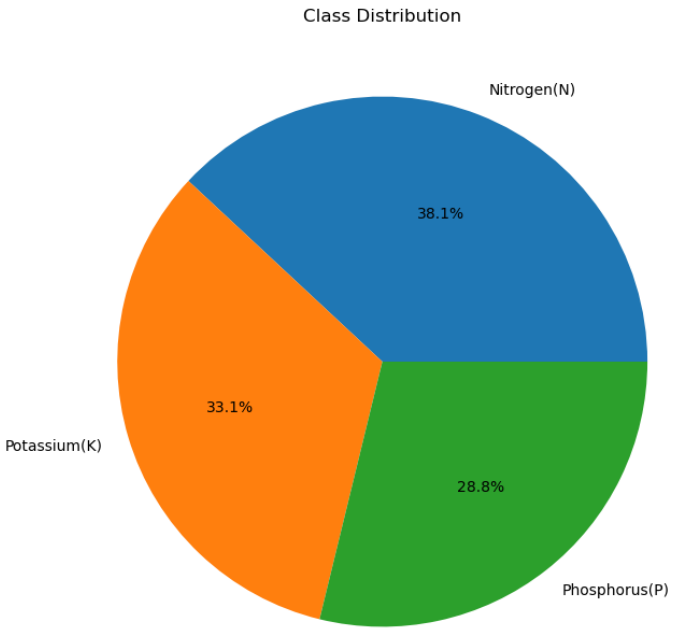


Figure 4.2: Rice dataset class distribution

ii. Banana dataset

The banana dataset is available on Mendeley data (https://data.mendeley.com/datasets/7vpdrbdkd4/1). The dataset comprises banana

leaves from different banana types such as Robusta, Poovan, and Monthan. The dataset consists of three classes: calcium, iron, and healthy leaves. Figure 4.3 shows how the three classes are distributed.

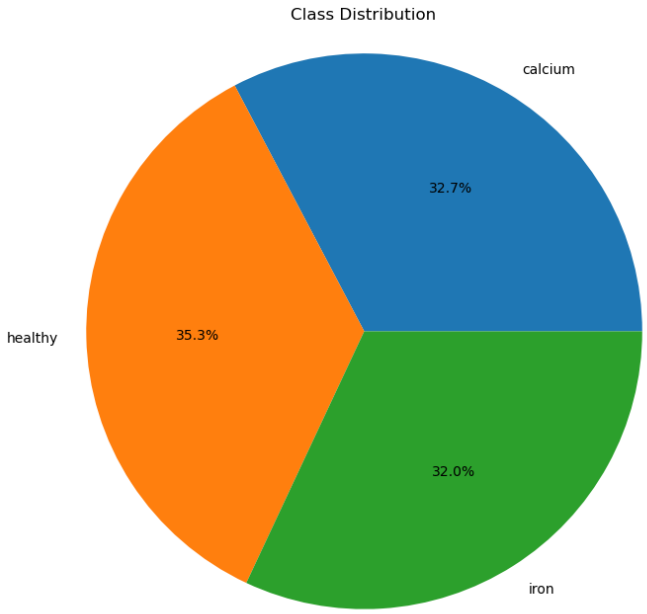


Figure 4.3: Banana dataset class distribution

4.3 Image Pre-processing

Images need to be pre-processed before being used for model training and testing (Kumar & Debika, 2018). Images may be pre-processed using a variety of techniques for, example, scaling images, making them grayscale, and enhancing images. This study's dataset comprises images of rice and banana in the JPEG — Joint Photographic Experts Group — format (.jpg), categorised into various classes

4.3.1 Data Augmentation

As per the research done by Kumar and Debika (2018), data augmentation is an ML technique that involves applying various transformations to the original images to increase the size of the training dataset. This technique is utilised for pre-processing purposes. The use of dataset augmentation is a common approach to expand datasets and enhance the diversity of data, thereby improving the performance of DL models under various conditions. It serves as a regulariser and reduces overfitting when training a model. In this study, a Keras image data generator class *'ImageDataGenerator'* was used for data augmentation. The data augmentation parameters are shown in Table 4.3.

Table 4.3: Data Augmentation Parameters

Augmentation Technique	**Parameter Setting**	**Outcome**
Rescaling	*rescale*=1. / 255	It normalises the pixel values of the images to the range [0,1].
Rotation	*rotation_range*=15	The images are randomly rotated by a maximum of 15 degrees clockwise or counterclockwise.
Width Shift	*width_shift_range*=0.1	This allows the images to be randomly shifted horizontally by a maximum of 10% of the image width.
Height Shift	*height_shift_range* =0.1	This allows the images to be randomly shifted vertically by a maximum of 10% of the image height.
Zoom	*zoom_range*=0.2	Images can be randomly zoomed in or out by a maximum of 20%.
Shear Transformation	*shear_range*=0.2	Applies the random shear transformation to the images.
Brightness Adjustment	*brightness_range*=[0.8, 1.2]	Randomly adjusts the brightness of the images.
Horizontal Flip	*horizontal_flip*=True	Random horizontal flipping of the images.
Vertical Flip	*vertical_flip*=True	Random vertical flipping of the images.

After applying the transformations to the training data, the augmented data is combined with the original training data and labels, and the resulting dataset is returned. This augmented dataset can be used to train a machine learning model, providing diverse and additional data for the model to learn from.

4.4 Model Training and Testing

To train and test the machine learning models the study used a built-in function called “train-test-split” from the Sklearn model selection module. The function divides the dataset into training and test sets with an 80:20 ratio. The split choice is based on the proposal by Mohanty et al. (2016). The study used a standard CNN and two pre-trained models (Inception-V3 and VGG-16). Using pre-trained CNN models primarily revolves around the rapid and effortless training of a CNN by using randomly initialised weights (Deniz, Şengür, Kadiroğlu, Guo, Bajaj & Budak, 2018). Moreover, the study used one-hot encoding for our categorical labels.

Two callbacks were used in the model's training and validation processes. Monitoring the validation loss and learning rate decrease by a factor of 0.5 was done using the first callback. To stop early, a second callback was used. Both callbacks were applied within 30 epochs as well to prevent the model from over-fitting. In the following sections, the conventional CNN, along with two pre-trained models, namely Inception-V3 and VGG-16 are outlined.

4.4.1 Convolutional Neural Networks

CNN is one of the most important algorithms in the computer vision field used for image classification (Alzubaidi et al., 2021). Figure 4.4 shows the plot of the CNN model architecture used for this study.

The first layer is a *Conv2D* layer with 32 filters, a kernel size of *(5,5)*, and a ReLu activation function. The input shape is the height and width of the input image, and 3 represents the number of colour channels (RGB). The second *Conv2D* layer has the same parameters as the first, followed by a *MaxPool2D* layer with a pool size of *(2,2)* and a Dropout layer with a rate of *0.25.*

The third and fourth layers are similar to the second layer but with 64 filters instead of 32. The fifth layer is a Flatten layer that converts the output of the previous layers into a 1D array. The sixth layer is a Dense layer with 256 units and a ReLu activation function, followed by a Dropout layer with a rate of 0.5. The final layer is a Dense layer with the number of units equal to the number of classes, and a softmax activation function. In total, all the layers of this CNN

model are trainable, and it comprises *22,514,211* parameters. Table 4.4 shows the parameters used to train the CNN model.

Table 4.4: CNN Parameters

Parameter	**CNN**
Cross-Validation	k-fold (k=10)
Epochs	30
Batch Size	32
Early Stopping	Patience =10
Initial learning rate	0.001
Optimizer	Adam
Loss	Categorical Crossentropy

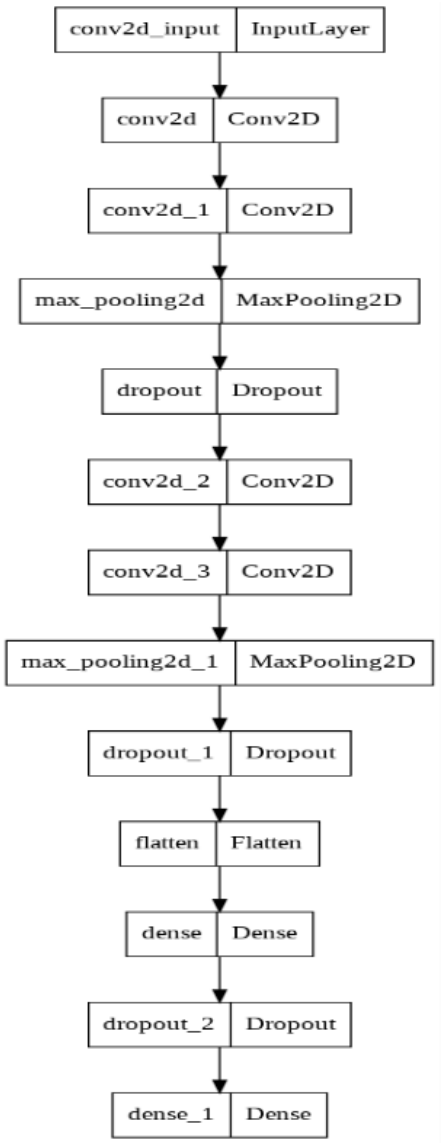


Figure 4.4: CNN Model plot

4.4.2 Pre-trained Inception-V3

The Inception v3 model's configuration comprises multiple layers of different types, such as convolutional, input, batch normalisation, pooling, and activation layers.

In this configuration of the Inception v3 model, there are several layers of diverse types. These include input layers, convolutional layers, pooling layers, batch normalisation layers, and activation layers.

The input layer: This layer accepts the input images and passes them on to the next layer (Szegedy et al. 2015). The input images are *224x224* pixels in size and have three colour channels (RGB), as indicated by their output shape (*None, 224, 224, 3*).

The convolutional layer: This layer is where the convolution operation is performed (Szegedy et al. 2015). The input data undergoes a series of filters, each of which is tailored to identify a particular pattern within the data. In this configuration, the convolutional layer applies thirty-two filters of size *3x3* to the input data.

The batch normalisation layer: According to Szegedy et al. (2015), the normalisation layer is used to standardise the output of the convolutional layer, resulting in a mean of zero and a standard deviation of one. This process can enhance the model's performance.

The activation layer: The output of the batch normalisation layer is passed through an activation function in this layer (Szegedy et al. 2015). ReLU is commonly used as the activation function in CNNs.

The pooling layer: According to Szegedy et al. (2015), this layer is utilised to decrease the data's size. It does this by applying a "pooling" operation, which typically involves taking the maximum or average value of a group of pixels in the input data. Reducing the amount of data that the next layer needs to process can accelerate the overall processing time.

The fully connected layer: The "fully connected" layer, where all neurons are connected to those in the previous layer, is responsible for the final classification or prediction based on the input data. This layer uses the output of the previous layers to make the ultimate prediction (Szegedy et al. 2015). For training the Inception-v3, the utilised parameters are presented in Table 4.5.

Table 4.5: Inception-v3 Parameters

Parameter	**Inception-V3**
Epochs	30
Cross-Validation	k-fold (k=10)
Batch Size	32
Early Stopping	Patience =10
Initial learning rate	0.001
Optimizer	Adam
Loss	Categorical Crossentropy

4.4.3 Pre-trained VGG-16

This study used the pre-trained VGG-16 CNN as the backbone architecture for the image classification task. First, the VGG-16 architecture was imported from the Keras applications library. Also, the necessary layers and the model object were imported.

This study defined the input tensor with the desired image size of 150x150 and 3 channels (RGB) and then passed this input tensor to the VGG-16 architecture, setting the weights argument to 'imagenet' to use the pre-trained weights. All the layers in the VGG-16 model were made to be non-trainable to avoid retraining the existing weights of the architecture.

Utilising the VGG-16 architecture with pre-trained weights, defining input and output layers, creating a model object, and compiling it using the categorical crossentropy loss function, Adam optimizer, and accuracy metrics. For training the VGG-16 model, the parameters used are presented in Table 4.6.

Table 4.6: VGG-16 Parameters

Parameter	**VGG-16**
Epochs	30
Cross-Validation	k-fold (k=10)
Batch Size	32
Early Stopping	Patience =10
Initial learning rate	0.001
Optimizer	Adam
Loss	Categorical Crossentropy

4.5 Chapter Summary

This chapter describes the procedures used in the experiments that produced the findings of the study. Further outlining the system infrastructure used, and model parameters, including the software and libraries used. Additionally, the chapter delves into the architecture of the standard CNN model and pre-trained Inception-V3 and VGG-16

CHAPTER FIVE

EVALUATION

This chapter examines the effectiveness of the three models used in the study and their explainability. To gauge the effectiveness of the models, the study used standard metrics like accuracy, precision, and F1-score. Subsequently, the study explores the use of SHAP and GRAD-CAM for model explainability.

5.1 Evaluation of Models' Performance

The study assessed the performance of three well-known CNN architectures: standard CNN, VGG-16, and Inception-v3. These architectures were tested on two different datasets, rice and banana. The performance of each model was measured using four standard classification metrics which are accuracy, precision, recall, and the F1-Score.

5.1.1 Classification Performance on Rice Dataset

Table 5.1 shows the performance of each model when applied to the rice dataset. The three models performed well across all metrics, with Inception-V3 and Inception-V3 achieving the highest scores of 93% for all metrics Accuracy, precision, recall, and F1-score.

Table 5.1 Rice dataset evaluation

	Rice Dataset			
	Mean			
Classifiers	**Accuracy**	**Precision**	**Recall**	**F1-Score**
CNN	84%	84%	84%	84%
VGG-16	93%	93%	93%	93%
Inception-V3	93%	93%	93%	93%

The graphical representation of the different models' accuracy on each fold using a 10-fold cross-validation approach is shown in Figure 5.1 – 5.3

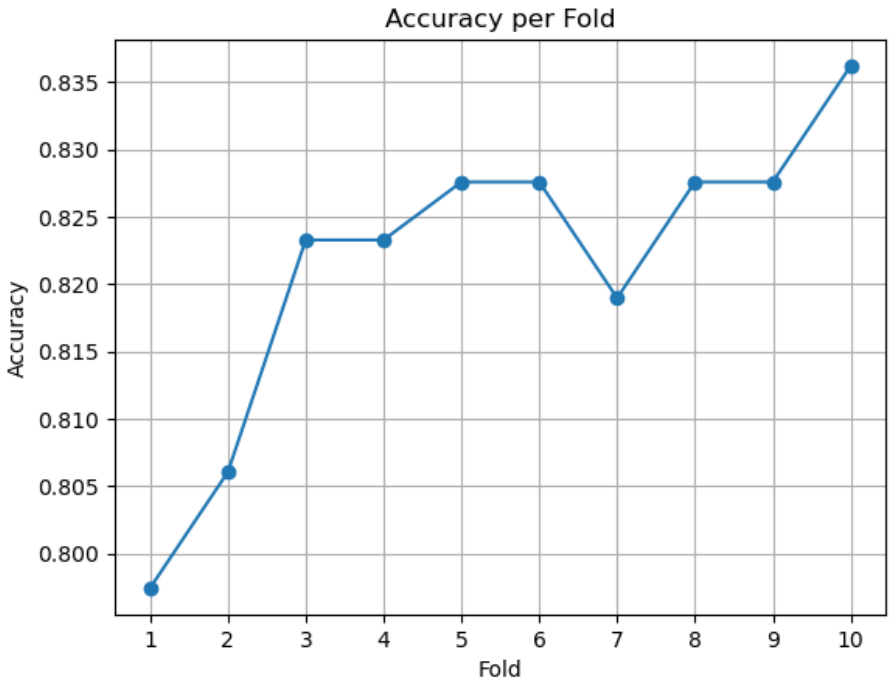


Figure 5.1: Performance accuracy of CNN on each fold

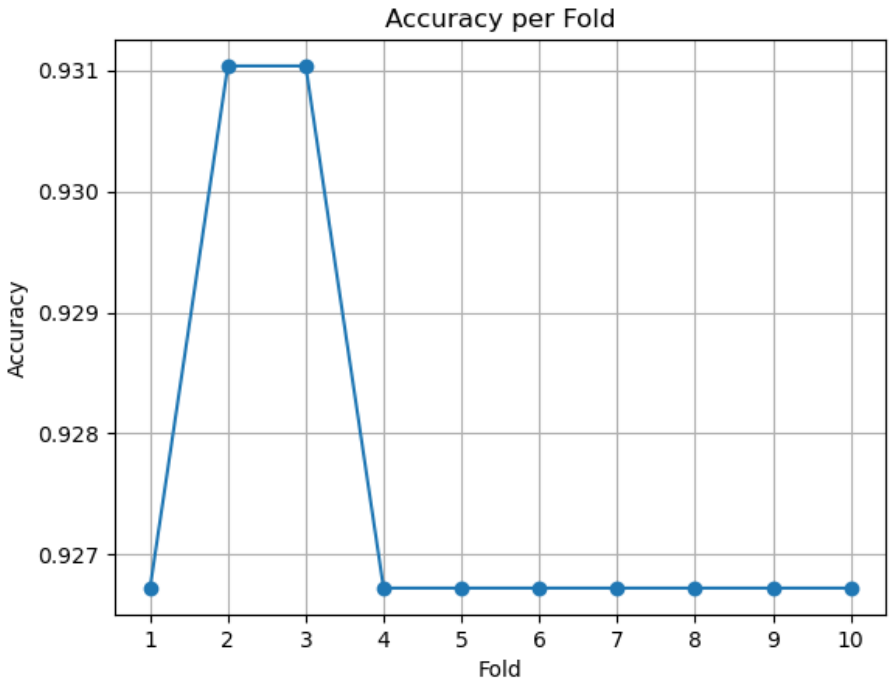


Figure 5.2: Performance accuracy of VGG-16 on each fold

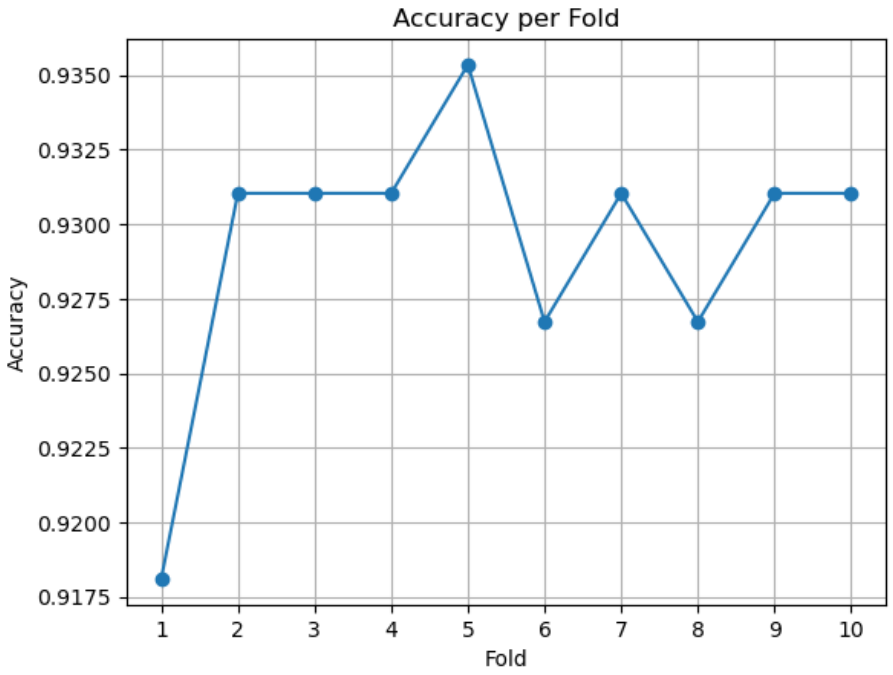


Figure 5.3: Performance accuracy of Inception-V3 on each fold

5.1.2 Classification Performance on the Banana Dataset

As shown in Table 5.2, Inception-V3 achieved the best performance, achieving 92% across all evaluation metrics. With an accuracy of 82%, VGG-16 performed better than the CNN model with a score of 68%. According to these results, Inception-V3 outperforms the other models, making it the preferred choice for classifying the banana dataset. Figure 5.4 to 5.6 depicts the graphical representation of the models used using a 10-fold cross-validation approach for each fold.

Table 5.2 Banana dataset evaluation

	Banana Dataset			
	Mean			
Classifiers	**Accuracy**	**Precision**	**Recall**	**F1-Score**
CNN	68%	68%	68%	68%
VGG-16	82%	81%	82%	81%
Inception-V3	92%	92%	92%	92%

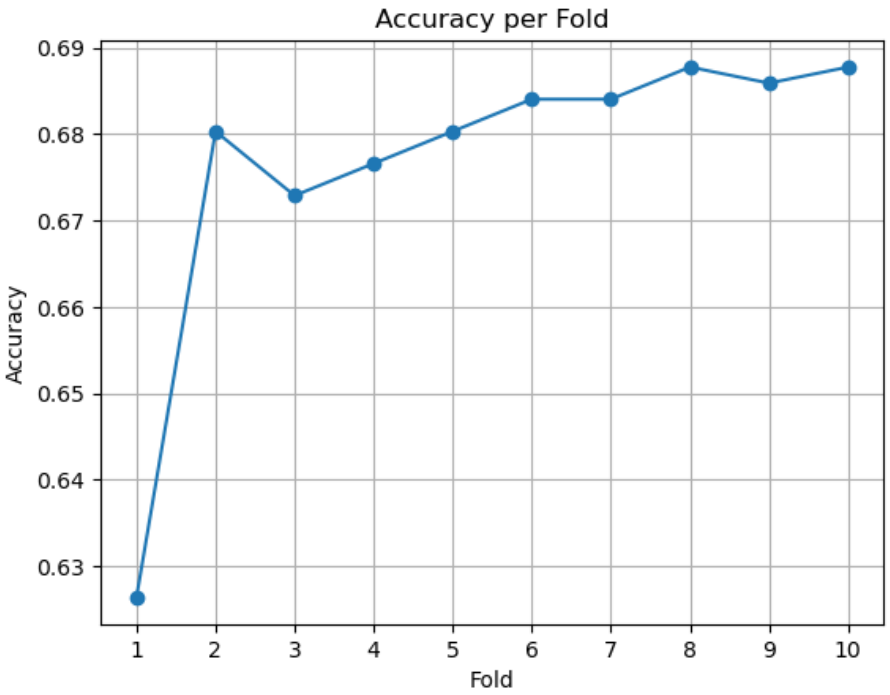


Figure 5.4: Performance accuracy of CNN on each fold

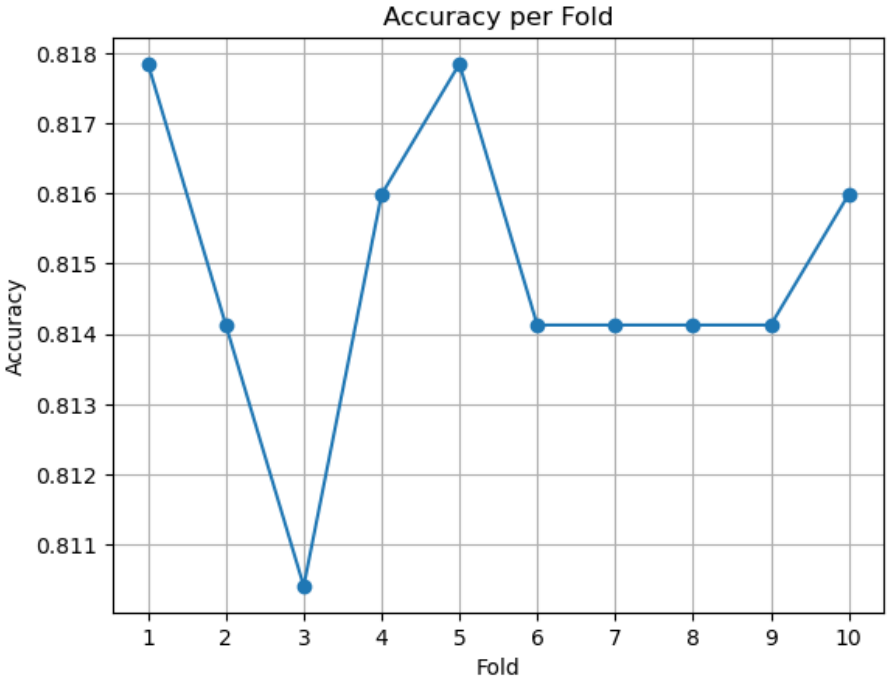


Figure 5.5: Performance accuracy of VGG-16 on each fold

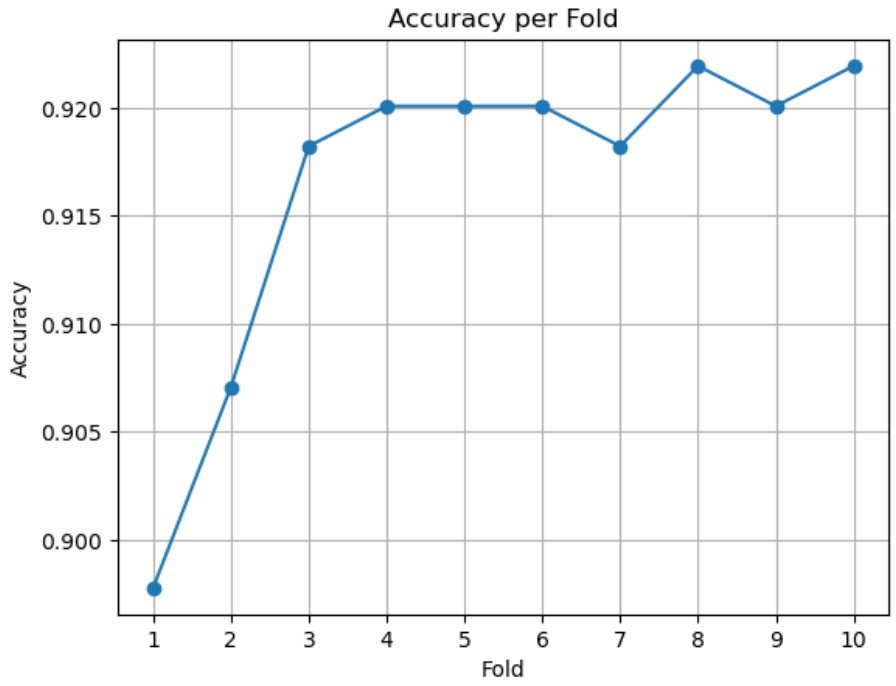


Figure 5.6: Performance accuracy of Inception-V3 on each fold

Table 5.1 to Table 5.2 highlight that Inception-V3 appears to be the most suitable model for the given datasets, as it consistently achieves the highest performance. VGG-16 also performs well and is competitive, but it falls short of Inception-V3. Therefore, Inception-V3 stands out as the preferred model for the given datasets.

5.1.3 Confusion Matrix

A helpful instrument for assessing the effectiveness of ML models is the confusion matrix. It provides a detailed breakdown of how many instances were classified correctly or incorrectly for each class. To assess the efficiency of standard CNN, VGG-16, and Inception-V3 architectures, they underwent testing on two distinct datasets.

i. Rice dataset

The confusion matrix for the three models when applied to the rice dataset is shown in Tables 5.3 to 5.5 The confusion matrix for the CNN model shows that it predicted Nitrogen correctly *75 times*, while incorrectly predicting Phosphorus *7 times* and Potassium *6 times*. Similarly, the model predicted Phosphorus correctly *59 times* but incorrectly predicted Nitrogen *6 times* and Potassium *13 times*. Finally, for the Potassium class, the model correctly predicted it *60 times* but incorrectly predicted Nitrogen 2 *times* and Phosphorus *4 times*. The model's accuracy rate is 84% in total.

The VGG-16 model's confusion matrix shows a higher accuracy compared to the CNN model. It correctly predicted Nitrogen *84 times*, Phosphorus *57 times*, and Potassium *74 times*. The model's accuracy rate is 93% in total.

The overall accuracy of the Inception-V3 model is 93%, as indicated by its confusion matrix. The model correctly predicted Nitrogen *94 times*, Phosphorus *56 times*, and Potassium *66 times*.

Table 5.3 CNN rice dataset confusion matrix

True Labels		Nitrogen	Phosphorus	Potassium
	Nitrogen	75	6	2
	Phosphorus	7	59	4
	Potassium	6	13	60

Table 5.4 VGG-16 rice dataset confusion matrix

True Labels		Nitrogen	Phosphorus	Potassium
	Nitrogen	84	0	2
	Phosphorus	2	57	5
	Potassium	4	4	74

Table 5.5 Inception-V3 rice dataset confusion matrix

True Labels		Nitrogen	Phosphorus	Potassium
	Nitrogen	94	3	1
	Phosphorus	5	56	2
	Potassium	1	4	66

ii. Banana Dataset

The banana dataset underwent evaluation using three ML models – CNN, VGG-16, and Inception-V3. The corresponding confusion matrix outputs for each model are depicted in Tables 5.6 to 5.8, respectively. Overall, all three models performed well on the banana dataset, as indicated by the relatively high number of correct predictions compared to misclassifications.

Table 5.6 CNN banana dataset confusion matrix

True Labels		Calcium	healthy	iron
	Calcium	109	50	28
	healthy	48	132	14
	iron	20	8	129

Table 5.7 VGG-16 banana dataset confusion matrix

True Labels		Calcium	healthy	iron
	Calcium	126	28	20
	healthy	22	156	8
	iron	12	9	157

Table 5.8 Inception-v3 banana dataset confusion matrix

True Labels		Calcium	healthy	iron
	Calcium	154	16	13
	healthy	5	182	3
	iron	5	0	160

5.1.4 Receiver Operating Characteristic (ROC) Curve

According to Chicco and Jurman (2023), the receiver operating characteristic curve is a performance measure used in ML for assessing the efficacy of binary classification models. When there are issues with class imbalance, such as when one class is significantly more

prominent than the other, the AUC (Area Under the Curve) will be used. On the *y-axis* of the ROC curve is the true positive rate and, on the *x-axis* is the false positive rate. The ROC AUC is a widely used statistical measure in scientific research for evaluating binary classifications (Chicco & Jurman, 2023). It quantifies the performance of a classifier ranging from 0 (indicating the worst result) to 1 (representing the perfect result).

The ROC curve of each of the three models is shown in Figure 5.7 – 5.9. All three models demonstrated *very good performance* in image classification, achieving high AUC values for the three different nutrient classes: 1- Nitrogen, 2- Phosphorus, and 3- Potassium. The rice dataset results suggest that DL models like VGG-16 and Inception-V3 are well-suited for multi-class classification tasks.

The banana dataset as depicted in Figure 5.10 – 5.12, shows that the AUC values are constantly above 0.9 when classifying the “healthy” and “Iron” classes for all three models, indicating excellent discrimination between positive and negative samples. Based on the AUC-Scores, the scores fall within the *very good performance* range according to the AUC- Score interpretation scale. While all three models performed well, Inception-V3 stands out as the most robust model, exhibiting high AUC values across the two datasets, as depicted in Table 5.9.

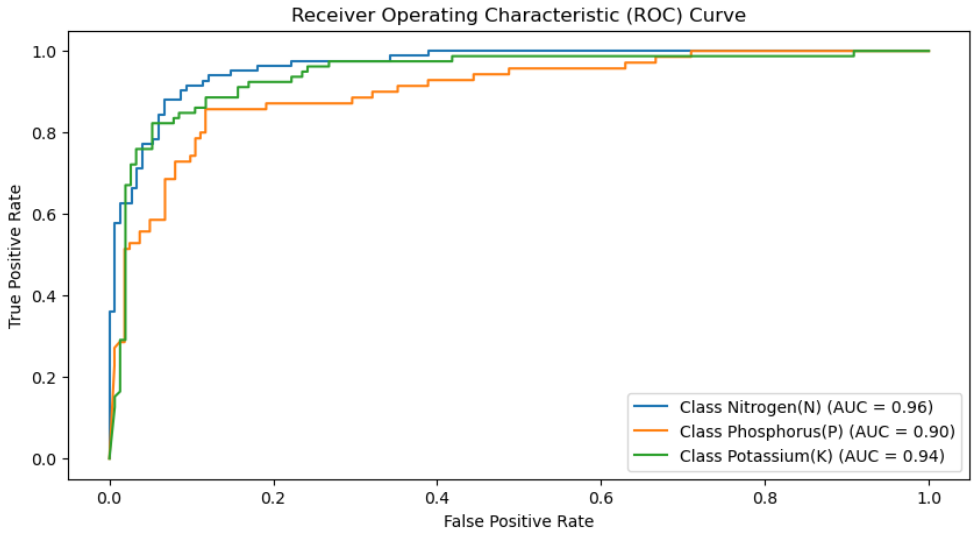


Figure 5.7: Rice dataset - CNN ROC Curve

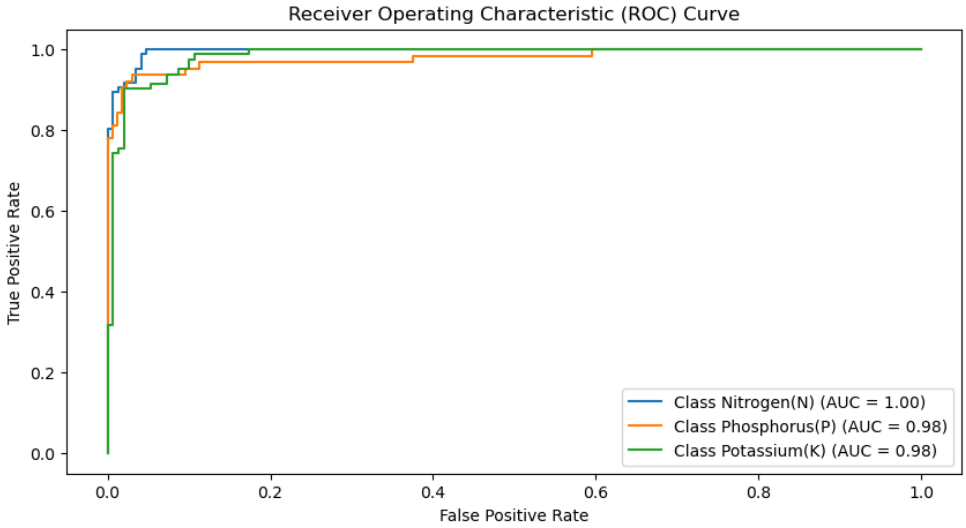


Figure 5.8: Rice dataset - VGG-16 ROC Curve

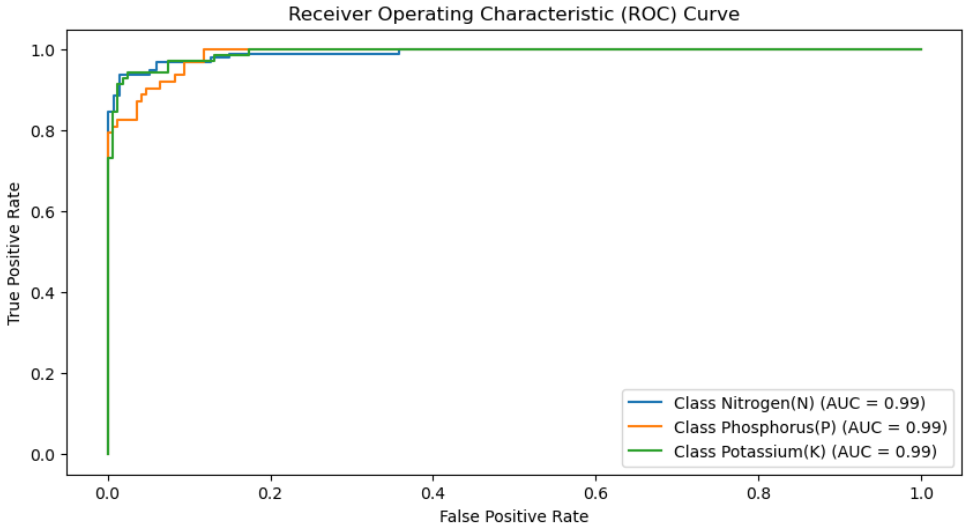


Figure 5.9: Rice dataset - Inception-V3 ROC Curve

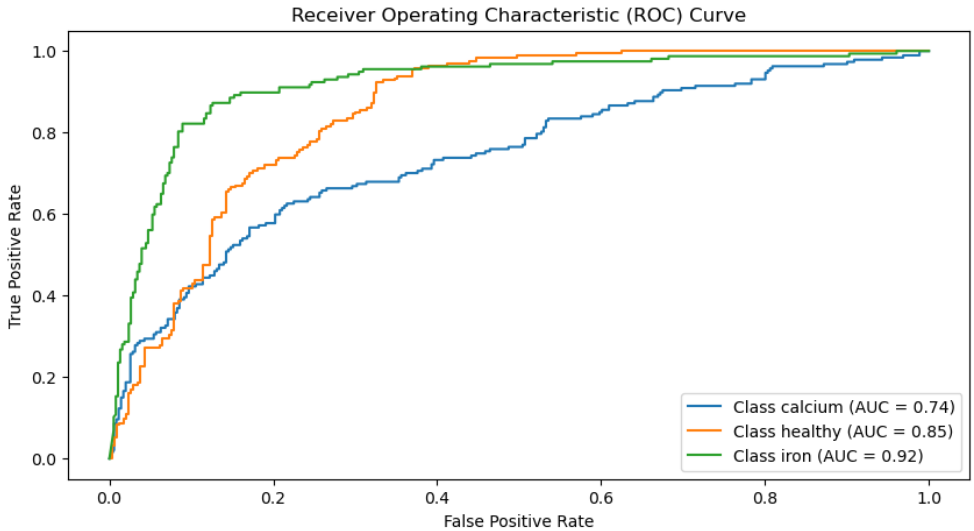


Figure 5.10: Banana dataset - CNN ROC Curve

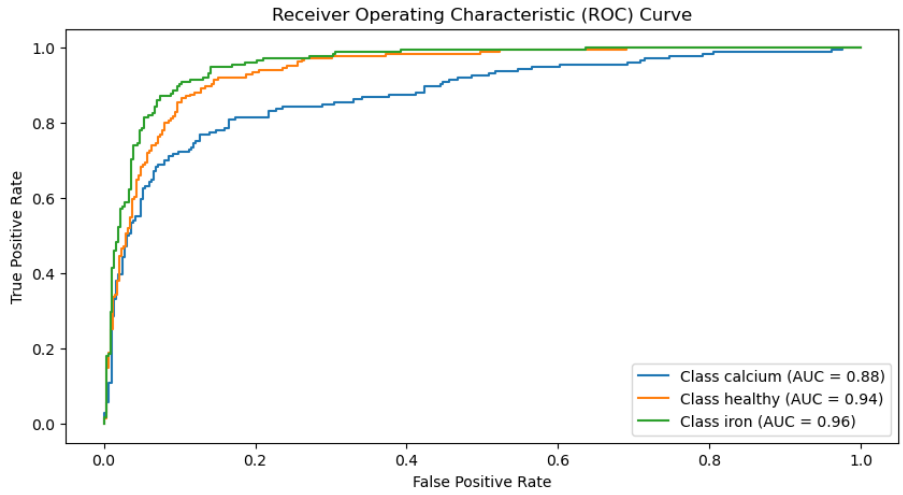


Figure 5.11: Banana dataset - VGG-16 ROC Curve

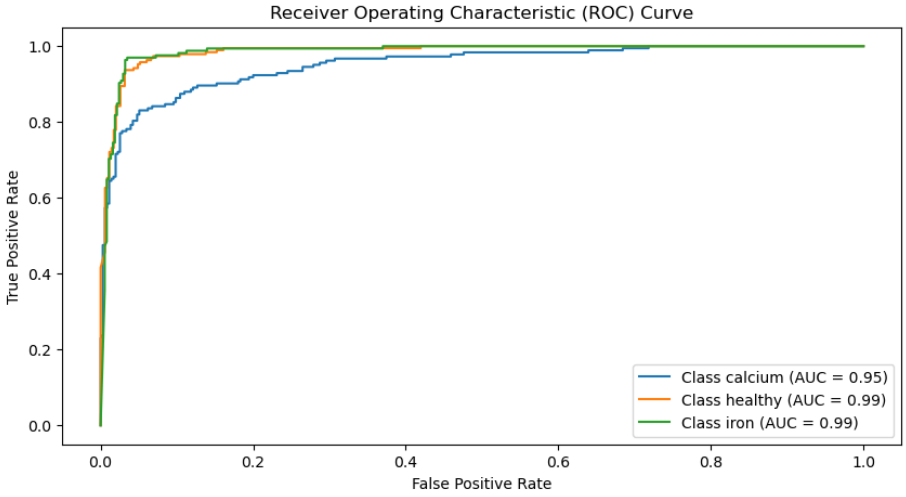


Figure 5.12: Banana dataset - Inception-V3 ROC Curve

Table 5.9 AUC scores of the DL models on the two datasets

RICE Dataset			
	CNN	**VGG-16**	**Inception- V3**
Nitrogen	0.96	1.00	0.99
Phosphorus	0.90	0.98	0.99
Potassium	0.94	0.98	0.99
Banana Dataset			
	CNN	**VGG-16**	**Inception- V3**
Calcium	0.74	0.88	0.95
healthy	0.85	0.94	0.99
Iron	0.92	0.96	0.99

5.2 Analysis of explainability using SHAP

Notwithstanding the remarkable accuracy of the three models, there exists a lack of transparency in these models. To explain any ML model, there are a few generalised XAI methods. For this study, two XAI algorithms, SHAP and Grad-CAM, were used. In this section, the SHAP output of the three models when applied to two different datasets were examined.

5.2.1 Explainability of ML models using SHAP - Rice Dataset

The input image is displayed on the left in Figure 5.13, while the SHAP output of the CNN model is highlighted in red and blue pixels. Red tones show positive contributions and blue tones show negative contributions to the prediction of this category. The model displays the three categories that it deems to be associated with the image.

The probability of phosphorus prediction is shown to increase in the first explanation image through the display of a red tone. The explanations for nitrogen and potassium contain fewer red colours. Overall, the figure suggests that the CNN model relies heavily on a few prominent features to make its predictions. Figure 5.14 shows the SHAP output of the VGG-16 model. This model also appears to rely on a limited set of notable features, similar to the CNN model. However, the contribution of each feature is more evenly distributed, with fewer features having an extremely positive or negative contribution

The SHAP output of the Inception-v3 model is displayed in Figure 5.15. Inception-V3 appears to be more reliant on a broader set of features than CNN, with many features making a significant positive or negative contribution to the final prediction.

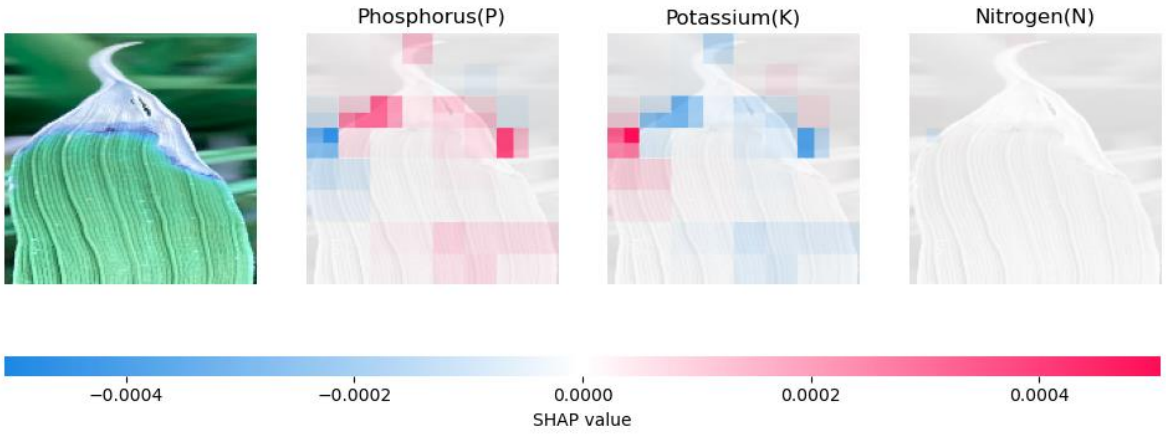


Figure 5.13: CNN (rice) SHAP output

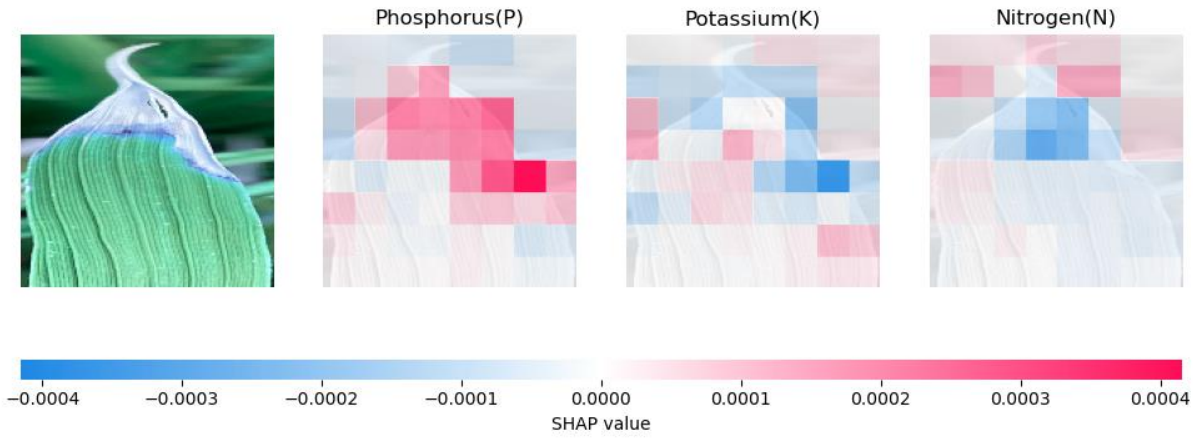


Figure 5.14: VGG-16 (rice) SHAP output

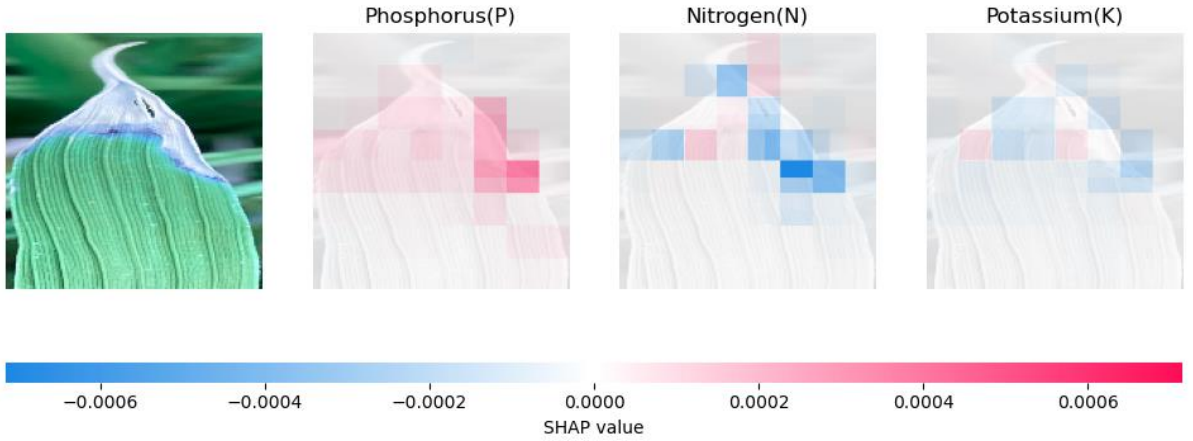


Figure 5.15: Inception-V3 (rice) SHAP output

5.2.2 Explainability of ML models using SHAP - Banana Dataset

By analysing the SHAP output from the CNN model as shown in Figure 5.16, it is obvious that the most influential regions on the banana leaf are on the sides of the leaf. Compared to the other models, the VGG-16 in Figure 5.17 distribute the SHAP values across the whole leaf. The Inception-V3 models in Figure 5.18 rely on broader context clues, resulting in different patterns of SHAP values across the banana leaf.

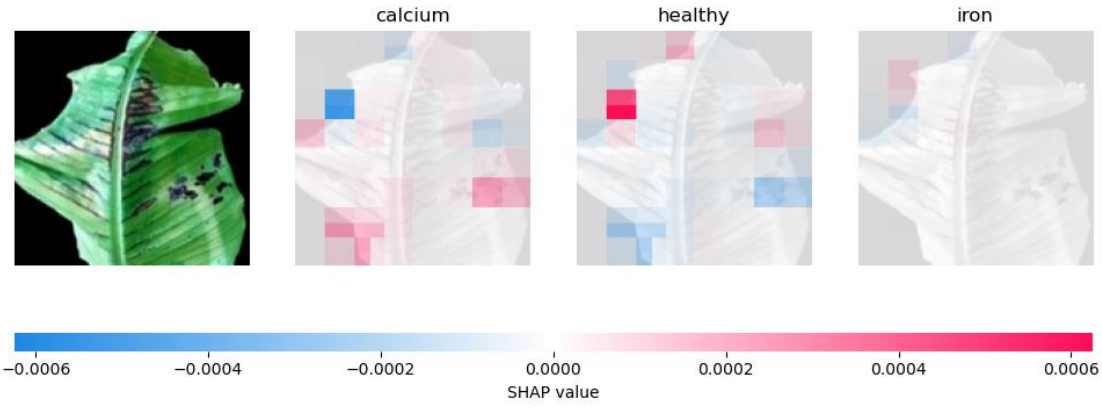


Figure 5.16: CNN (banana) SHAP output

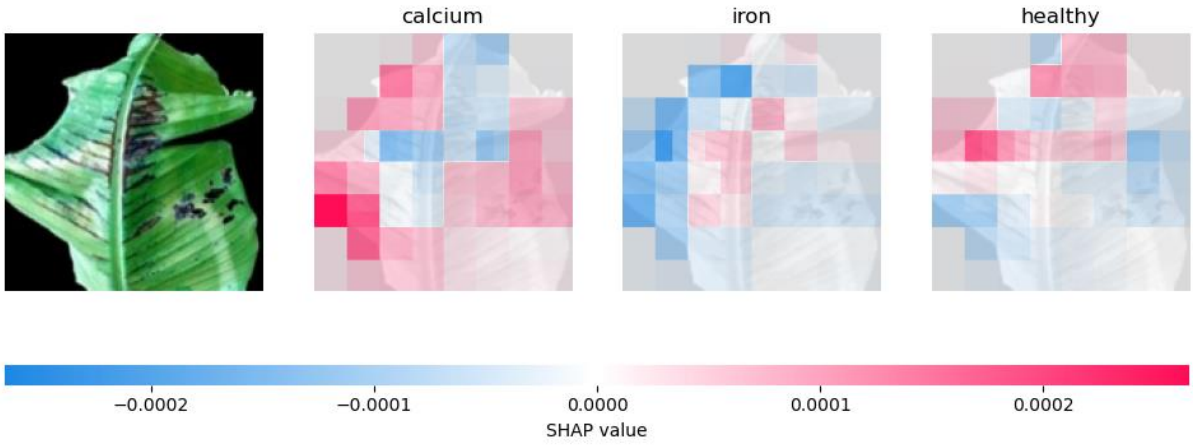


Figure 5.17: VGG-16 (banana) SHAP output

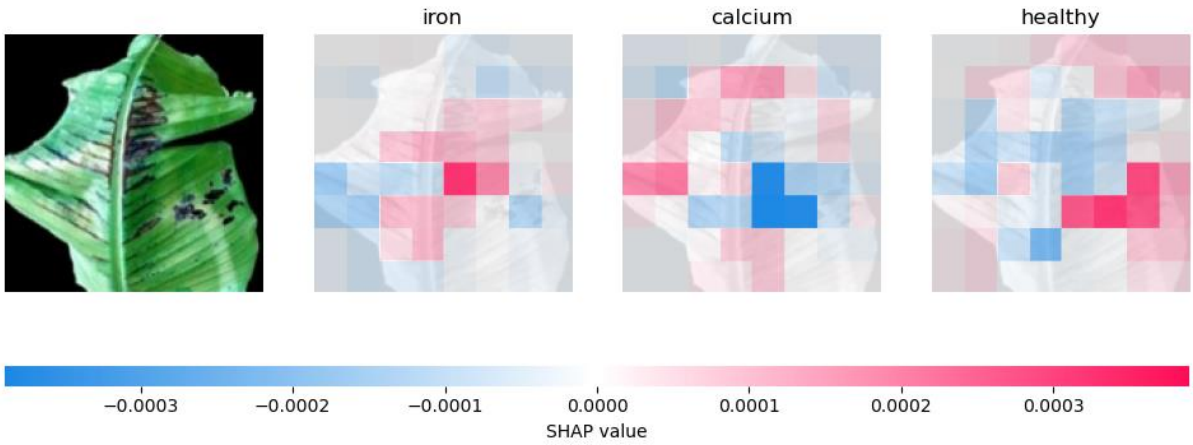


Figure 5.18: Inception-V3 (banana) SHAP output

Overall, the SHAP output of these three deep learning models suggests that each model relies on a distinct set of key features to make predictions as summarised in Table 5.10. The CNN model appears to be the most reliant on a small number of features, while Inception-V3 and VGG-16 rely on a broader set of features. The implication is that the selection of the model can influence which features are deemed crucial for the prediction.

Table 5.10 Summary of Explainability of ML Models based on SHAP

Models	**Rice Dataset**	**Banana Dataset**
CNN	Relies heavily on prominent features, especially for phosphorus prediction	Influential regions on the sides of the leaf.

Inception-V3	Relies on a broader set of features	Relies on broader context clues, resulting in a different pattern of SHAP values
VGG-16	Similar to CNN, but with more evenly distributed feature contributions	Distributes SHAP values across the whole leaf.

5.3 Analysis of explainability using Grad-CAM

For each of the three models, a GRAD-CAM visualisation technique is used to debug the prediction process and emphasise the intriguing areas of the plant picture that determine the final decision.

CNN possess a complex architecture often referred to as a “black box,” resulting in a lack of transparency regarding its internal operations. However, Grad-CAM is a technique proposed by Selvaraju et al. (2017) that can be used to debug almost any CNN model by locating and processing the gradient of the target for the last convolutional layer.

According to Setiawan and Rulaningtyas (2023), visual explanation provides an overview by generating a heatmap where pixels with high to low intensity are coloured from red, yellow, green, and blue. This technique can be used to determine whether the model accurately predicts the absence of potassium based on the infected region of the plant.

5.3.1 Explainability of ML models using Grad-CAM - Rice Dataset

The CNN GRAD-CAM heatmap reveals that the model mainly focuses on the tip of the leaf, where the nutrient deficiency is prominent. The VGG-16 considers a similar part of the leaf. However, the heatmap shows a more accurate part of the plant compared to the CNN model. The way that VGG-16 localised the defected region shows that the model can be trusted. People can better understand how VGG-16 uses image classification to identify the lack of nutrients in plants by examining the highlighted area.

Inception-V3 GRAD-CAM heatmap shows a different pattern compared to the first two models. The model highlights various broader areas rather than a single dominant area. The approach used in the study for nutrient deficiency identification is consistent with the findings of Chopra and Whig (2021). The results presented in Figure 5.19 demonstrate the effectiveness of Grad-CAM in providing these models' transparency and interpretability.

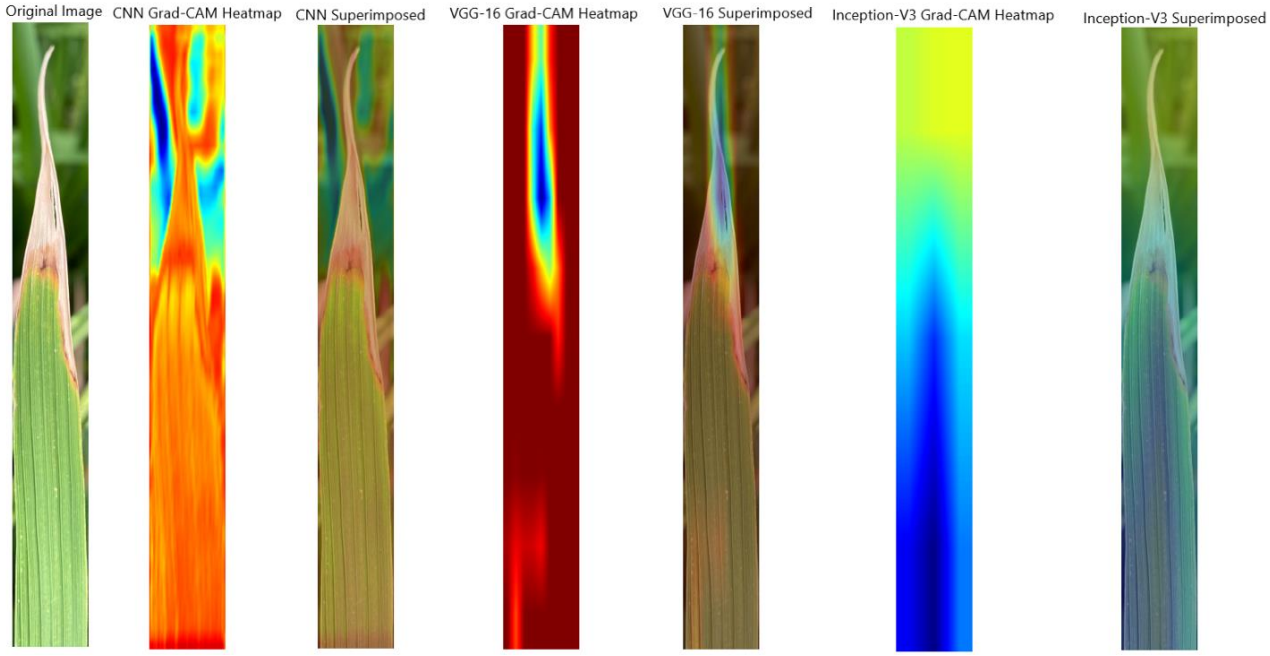


Figure 5.19: Rice Grad-CAM output

5.3.2 Explainability of ML models using Grad-CAM - Banana Dataset

Figure 5.20 illustrates the use of the Grad-CAM approach when applied to the three models to construct the class activation mapping. According to Selvaraju et al. (2017), this mapping helps to localise the specific region of the plant that predominantly influenced the decision. In this instance, the CNN model tends to highlight the contours of the banana leaf. However, the other two models focus on the leaf itself. The Inception-V3 model heatmap lacks a precise overlap with the object of interest.

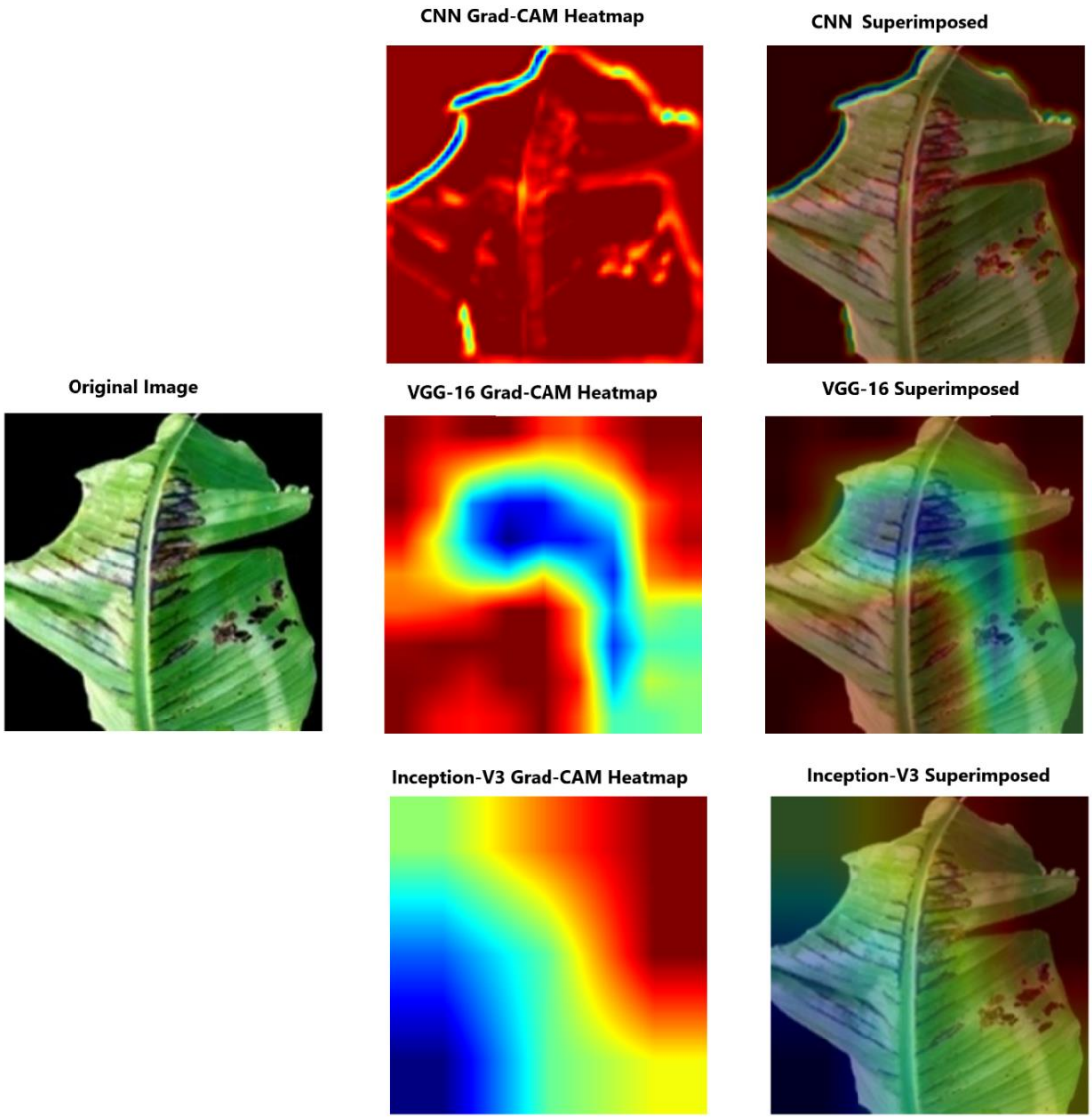


Figure 5.20: Banana Grad-CAM output

In conclusion, Grad-CAM is a powerful tool that helps understand the decision-making process of CNN models such as standard CNN, VGG-16 and Inception-V3. The results presented in Table 5.11 demonstrate the effectiveness of Grad-CAM in providing transparency and interpretability of these models. This visual explanation enables a better understanding of how the model performs image classification. Moreover, Grad-CAM is a potential approach for improving the interpretability and transparency of CNN models, offering practitioners and researchers insightful information about their internal working and decision-making processes.

Table 5.11 Summary of Explainability of ML Models based on Grad-CAM

Models	**Rice Dataset**	**Banana Dataset**
CNN	Grad-Cam heatmap focuses on the tip of the leaf.	Highlights banana leaf contours.
Inception-V3	Highlights various broader areas.	Lacks precise overlap with the object of interest
VGG-16	Localises defected region.	Focuses on the leaf itself

5.4 Discussion of Results

This study compared the performance and explanation of three CNN-based architectures for the identification of plant nutritional deficiencies. The study used two datasets for conducting experiments on the models. The rice dataset was used to classify images based on three classes. Additionally, the banana dataset was used to assess the model's robustness based on three classes.

Based on the results obtained for the rice dataset, the Inception-V3 and VGG-16 models outperformed the CNN model in terms of performance scores both achieving 93% accuracy, which is consistent with the findings of Ibrahim et al. (2022). The study employed CNN to address palm leaf nutrient deficiencies and achieved an overall accuracy of 94.29%.

Precision and recall values provide an overview of the effectiveness of each model in identifying different nutrient deficiencies. As far as precision is concerned, the VGG-16 outperformed the other two models for nitrogen and phosphorus deficiencies, 93% and 93%, respectively. While Inception-V3 has the highest accuracy (96%) for potassium deficiencies.

Based on the F1-Scores, both the VGG-16 and Inception-V3 achieved a *good* score, both getting 93% F1-score for the rice dataset, while the CNN model demonstrated an *ok* score, achieving an 84% F1-score.

For the banana dataset, Inception-V3 excels once more, achieving a *very good* F1-Score of 92%. VGG-16 follows with a *good* F1-Score of 81%. And the CNN, while not as strong as the other models, achieves an *ok* F1-Score of 68%. Based on these findings, it appears that

Inception-V3 is effective in detecting nutrient deficiency in banana plants. The high performance in terms of accuracy of the Inception-V3 model on both datasets can be attributed to its deeper architecture and better feature extraction capabilities.

Although the studies in Table 2.6 focus on the performance of the models, the explainability factor added an extra layer of understanding of the results. The explainability of the models is also a crucial factor to consider. This study used two XAI algorithms, SHAP and Grad-CAM. Regarding SHAP, the CNN and VGG-16 models often rely on a limited set of prominent features. However, Inception-V3 appears to rely on a broader range of features, with many features making significant contributions to the final prediction.

For the banana and rice datasets, it was noted that the CNN model Grad-CAM heatmap highlights the contours of the plant leaf, while the other two models focus on the leaf itself. VGG-16 accurate localisation on the affected region proved to be more reliable due to its heatmap's quality. Overall, the visual explanations allowed for a better understanding of how these models identify the absence of nutrients by examining the highlighted areas.

Inception-V3 is the most accurate model, but it may not be the most understandable model due to its complex architecture. On the other hand, the VGG-16 has a simpler architecture that can explain it more. Therefore, it is important to balance accuracy and explication when selecting a model for a particular task.

5.5 Chapter Summary

This chapter provided a thorough analysis of the model's performance and explainability, which added insightful information to the comparative research of explainable deep learning models for the detection of nutrient deficiencies in plants. The results of the chapter provide a foundation for improving the model's accuracy, interpretability, and practical applicability in the agriculture sector.

CHAPTER SIX

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

Based on the observations made in Chapter 5, this chapter provides a comprehensive overview of the study's findings, conclusions, contributions, and recommendations. The first to the fourth objectives are covered in the summary section. Furthermore, this chapter serves as the conclusion of the study, offering specific insights into the research implications. It elaborates on the study's input and suggests recommendations based on the findings and conclusions, which the researcher deems potentially valuable for future research.

6.1 Summary

The research aimed to compare the performance of deep learning algorithms when applied to plant nutrient deficiency and to ascertain the level of their explainability. The study comprised six chapters, each making a distinct contribution to the overall research.

Chapter 1 discussed the research problem, along with the study's aim, research questions and objectives. Additionally, the chapter introduced the topic and offered a comprehensive background of the study, highlighting the delineation and significance of the study. In Chapter 2, the study presented literature related to explainable DL models for the identification of plant nutrient deficiencies. The research looked at the importance of DL and how they are applied in the agricultural sector. Additionally, the study highlighted the importance of XAI, and different XAI tools such as LIME, SHAP, and Grad-CAM. Chapter 3 provided an overview of the research methodology, including the research approach, research strategy and research design. Chapter 4 described the process used in the experiments that produced the findings in the study. Chapter 5 provided a thorough analysis of the model's performance and explainability. The final chapter, Chapter 6, presents the summary, recommendations, and conclusion of the study.

The following describes how the study's objectives were met:

i. **Objective 1: To identify the features that are essential to determine plant nutrient deficiencies.**

   Through a thorough analysis of the literature (Chapter 2), pertinent characteristics for identifying plant nutrient deficiencies were identified to achieve objective 1. There are several methods of analysing deficiencies in plants, such as visual observations, plant

analysis, and soil analysis. Plants exhibit signs of nutrient inadequacy in their leaves as a result of nutrient shortage.

Nutrient deficiencies in plants often exhibit initial visual symptoms on their leaves, which include yellowing, and reduced leaf size. For instance, a general look of light green to yellow is indicative of a Nitrogen (N) deficiency and is most obvious on older leaves. Lack of Phosphorous (P) causes brown spots and necrosis to appear on leaves, which is followed by reduced plant growth. Similar to this, Potassium (K) deficiency causes brown spots. Lack of Calcium (Ca) results in irregular or distorted development of new leaves. The last sign of Magnesium (Mg) shortage is yellowing of the veins of young leaves and in extreme cases, leaf drop.

ii. **Objective 2: To determine the requirements of explainable deep learning for nutrient deficiency identification.**

This objective was achieved by doing extensive research on machine learning and deep learning to understand the prerequisites and criteria for implementing explainable deep learning in the agriculture sector. In addition, the significance of XAI was highlighted, spotlighting the necessity for interpretable models in the agriculture sector. To improve model explainability, various XAI techniques have been explored by different authors in the literature. These include Local Interpretable Model-Agnostic Explanations (LIME), Shapley Additive exPlanations (SHAP), and Gradient-weighted Class Activation Mapping (Grad-Cam).

iii. **Objective 3: To explore how explainable deep learning could be applied to a plant image dataset to identify plant nutrient deficiencies.**

To achieve this objective of applying explainable deep learning to identify plant nutrient deficiencies, the study used two datasets. The datasets contained two different plant types, namely rice and banana. Each plant exhibits distinct nutrient deficiency symptoms. With the use of different datasets, the study was able to assess how well the chosen DL models worked with various plant types and nutrients.

For this study, three different DL models were used, including Convolutional Neural Networks (CNNs) and two pre-trained CNN models: Inception-V3 and VGG-16. CCNs were a good option for our objective because they have proven to be very successful in image-related tasks.

Two XAI techniques, namely Grad-CAM and SHAP were then applied to the Dl models. Through this, the visual explanations that helped in understanding how these models make predictions and gain insight into their decision-making processes were obtained.

iv. Objective 4: To determine the performance of selected deep learning algorithms when used for plant nutrient deficiencies and their explainability.

This objective was achieved in two aspects, by evaluating the model's performance in identifying plant nutrient deficiencies and their level of explainability through the two XAI techniques that were applied (Grad-CAM and SHAP).

The study used standard evaluation metrics like accuracy, precision, recall, F1-Score, and AUC-Score to assess the model's effectiveness in identifying nutrient deficiencies in plants. These metrics provided a quantifiable assessment of the model's effectiveness, allowing the researcher to evaluate how well each model identified certain nutrient deficiencies across different datasets.

In addition to the performance evaluation, the study also focused on the models' explainability. The model predictions were visually explained using XAI techniques, Grad-Cam and SHAP. By using heatmaps and feature attributions, these visual explanations highlighted the areas of the input images that affected the model's predictions.

By evaluating both performance and explainability, the researcher gained a comprehensive understanding of the DL models' strengths and limitations in identifying nutrient deficiencies in plants.

6.2 Contributions of the Study

This study made valid theoretical and practical contributions which are outlined below.

6.2.1 Theoretical Contribution

The study contributes to the literature by incorporating explainable deep learning in the context of plant nutrient deficiency identification. Moreover, unlike prior research, as highlighted in Table 2.6, that primarily evaluated accuracy without considering explainability, the study addressed this gap by comparing the explainability of GRAD-CAM and SHAP techniques, shedding light on how these models arrive at their predictions. Bringing attention to the

importance of a holistic approach that considers both performance and explainability in the agriculture sector will enhance trust and usability in the models' predictions.

6.2.2 Practical Contribution

The practical contribution of this thesis is in the practical implementation of explainable deep learning in identifying plant nutrient deficiencies. The study demonstrates their effectiveness in real-world scenarios by applying the models and XAI techniques to a diverse range of plants. This study has important applications for the agriculture sector, as it offers trustworthy and understandable solutions for quick and accurate detection of nutrient deficiencies in plants. This will contribute to the improvement of crop management, productivity, and sustainability in agricultural practices.

6.3 Limitations of the study

The study focuses on Grad-CAM and SHAP, two well-known XAI techniques. Even though these techniques have been explored and implemented, the research does not consider any new or alternative approaches for explainable deep learning. Additionally, the AI models are used differently by various users, hence the explanations they offer are susceptible to different interpretations. The study does not include a usability evaluation that focuses on the experiences of farmers, or agronomists in their attempt to use the machine learning models in practice.

The study is also limited by the unavailability of plant nutrient deficiency image datasets, particularly in Africa. Despite searching extensively, even well-known open-source repositories lacked such datasets. Consequently, the study had to rely on only two available datasets.

6.4 Conclusion

In evaluating the performance of three deep learning models (CNN, Inception-V3, and VGG-16), the study found that these models exhibited varying degrees of effectiveness in identifying nutrient deficiencies in rice and banana plants. This variation in model performance emphasizes the importance of selecting the right model for a particular task in agriculture. Crucially, the study also addressed the often-neglected aspect of model explainability, offering visual insights that enhance transparency and trust in model decision making.

The research successfully addressed its objectives, providing valuable insights into both the theoretical and practical aspects of this domain. The study's holistic approach and valuable

findings pave the way for the integration of XAI techniques in agriculture, adding value to the field and opening avenues for future research and innovation.

6.5 Recommendations and Future Work

For future work, it will be crucial to conduct a user study to assess the effectiveness of implemented DL models and explainability for the lack of nutrients in plant diagnosis among farmers. Moreover, it will help to determine whether the models are perceived as trustworthy and comprehensible by farmers and whether they are deemed useful tools for improving their diagnostic process. To assist farmers in identifying nutrient deficiencies on the spot in the field, future work should investigate the viability of deploying the models using edge devices such as smartphones or tablets. Finally, future studies can explore the possibility of implementing explainable DL models in various agricultural domains, including but not limited to yield prediction and disease identification.

REFERENCES

Abade, A.S., Ferreira, P.A. & Vidal, F. de B. 2020. Plant Diseases recognition on images using Convolutional Neural Networks: A Systematic Review. http://arxiv.org/abs/2009.04365.

Adadi, A. & Berrada, M. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6: 52138–52160.

Airola, R., Hager, K. & Alfredsson, S. 2017. *Image Classification, Deep Learning and Convolutional Neural Networks*.

Akata, Z., Balliet, D., De Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., Van Der Gaag, L., Van Harmelen, F., Van Hoof, H., Van Riemsdijk, B., Van Wynsberghe, A., Verbrugge, R., Verheij, B., Vossen, P. & Welling, M. 2020. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect with Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer*, 53(8): 18–28.

Allwright, S. 2022. How to interpret F1 score. https://stephenallwright.com/interpret-f1-score/ 2 September 2023.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. & Farhan, L. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1).

Antwarg, L., Miller, R.M., Shapira, B. & Rokach, L. 2020. *Explaining Anomalies Detected by Autoencoders Using SHAP*.

Aramay, R. 2014. Chapter 3 Research and Methodology. , (January): 38–46. https://www.academia.edu/7729576/Chapter_3_RESEARCH_AND_METHODOLOGY.

Aye, H.N. & Masih, S. 2023. Role of Nutrients in Plants, Its Deficiency and Management. *International Journal of Plant & Soil Science*, 35(10): 129–136. https://journalijpss.com/index.php/IJPSS/article/view/2932.

Barredo, A.A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. & Herrera, F. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58: 82–115.

Bengio, Y. & Haffner, P. 1998. Gradient-Based Learning Applied to Document Recognition. , 86(11).

Bhaduri, D., Rakshit, R. & Chakraborty, K. 2014. Primary and Secondary Nutrients-a Boon to Defense System against Plant Diseases. *International Journal of Bio-resource and Stress Management*, 5(3): 461.

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K. & Ghayvat, H. 2021. Cnn variants for computer vision: History, architecture, application, challenges and future scope. *Electronics (Switzerland)*, 10(20).

Billericay Fertiliser Services. 2015. *Nutrient Deficiency Identification Guide Diagnosing symptoms in major arable crops*.

Bodria, F., Giannotti, F., Guidotti, R., Naretto, F., Pedreschi, D. & Rinzivillo, S. 2021. Benchmarking and Survey of Explanation Methods for Black Box Models. http://arxiv.org/abs/2102.13076.

Canziani, A., Paszke, A. & Culurciello, E. 2016. An Analysis of Deep Neural Network Models for Practical Applications. : 1–7. http://arxiv.org/abs/1605.07678.

Carneiro, T., Da Nobrega, R.V.M., Nepomuceno, T., Bian, G. Bin, De Albuquerque, V.H.C. & Filho, P.P.R. 2018. Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access*, 6: 61677–61685.

Chandra, A.L., Desai, S.V., Guo, W. & Balasubramanian, V.N. 2020. Computer Vision with Deep Learning for Plant Phenotyping in Agriculture: A Survey. http://arxiv.org/abs/2006.11391.

Chicco, D. & Jurman, G. 2023. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Mining*, 16(1).

Chiu, C.-C., Sainath, T.N., Wu, Y., Prabhavalkar, R., Nguyen, P., Chen, Z., Kannan, A., Weiss, R.J., Rao, K., Gonina, E., Jaitly, N., Li, B., Chorowski, J. & Google, M.B. 2018. *State-of-the-art speech recognition with sequence-to-sequence models*.

Chopra, G. & Whig, P. 2021. Analysis of Tomato Leaf Disease Identification Techniques. *Journal of Computer Science and Engineering (JCSE)*, 2(2): 98–103.

Claudio, C. 2020. *Vision-Based Analysis on Leaves of Tomato Crops for Classifying Nutrient Deficiency using Convolutional Neural Networks*.

Creswell. 2014. Research Design: Qualitative, Quantitative and Mixed Methods Approaches.

Creswell, J.W. 2014. The Selection of a Research Approach. *Research Design*: 3–23. http://www.sagepub.com/upm-data/55588_Chapter_1_Sample_Creswell_Research_Design_4e.pdf.

Daniel, E. 2016. *The Usefulness of Qualitative and Quantitative Approaches and Methods in Researching Problem-Solving Ability in Science Education Curriculum*. Online. www.iiste.org.

Das, A. & Rad, P. 2020. Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey. : 1–24. http://arxiv.org/abs/2006.11371.

Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V. & Budak, Ü. 2018. Transfer learning based histopathologic image classification for breast cancer detection. *Health Information Science and Systems*, 6(1).

Doshi-Velez, F. & Kim, B. 2017. Towards A Rigorous Science of Interpretable Machine Learning. http://arxiv.org/abs/1702.08608.

Garreau, D. & Von Luxburg, U. 2020. *Explaining the Explainer: A First Theoretical Analysis of LIME*.

Gerlings, J., Shollo, A. & Constantiou, L. 2021. Reviewing the Need for Explainable Artificial Intelligence (xAI). , (January).

Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A. & Sarkar, S. 2018. An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences of the United States of America*, 115(18): 4613–4618.

Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A. & De, D. 2019. *Fundamental concepts of convolutional neural network*.

Goyal, D., Amit, ·, Gupta, K., Piuri, V., Ganzha, M. & Paprzycki, M. 2021. *Lecture Notes in Networks and Systems 166 Proceedings of the Second International Conference on Information Management and Machine Intelligence*. http://www.springer.com/series/15179.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. & Pedreschi, D. 2018. A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5).

Gunning, D. & Aha, D.W. 2019. DARPA’s Explainable Artificial Intelligence Program. : 44–58.

Håkansson, A. 2013. *Portal of Research Methods and Methodologies for Research Projects and Degree Projects*. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-136960.

Han, K.A.M., Maneerat, N., Sepsirisuk, K. & Hamamoto, K. 2023. Banana Plant Nutrient Deficiencies Identification using Deep Learning. In Institute of Electrical and Electronics Engineers (IEEE): 5–9.

Hao, J. & Ho, T.K. 2019. Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. *Journal of Educational and Behavioral Statistics*, 44(3): 348–361.

He, K., Zhang, X., Ren, S. & Sun, J. 2016. *Deep Residual Learning for Image Recognition*. http://image-net.org/challenges/LSVRC/2015/.

Ibrahim, S., Hasan, N., Sabri, N., Abu Samah, K.A.F. & Rahimi Rusland, M. 2022. Palm leaf nutrient deficiency detection using convolutional neural network (CNN). *International Journal of Nonlinear Analysis and Applications*, 13(1): 1949–1956.

Inácio, D. & Rieder, R. 2018. Computer vision and arti fi cial intelligence in precision agriculture for grain crops : A systematic review. , 153(August): 69–81.

Jeyalakshmi, S. & Radha, R. 2017. A review on diagnosis of nutrient deficiency symptoms in plant leaf image using digital image processing. *ICTACT Journal on Image and Video Processing*,7(4):1515–1524.
http://ictactjournals.in/paper/IJIVP_Vol_7_Iss_4_Paper_7_1515_1524.pdf.

Johannesson, P. & Perjons, E. 2014. Research Strategies and Methods. In *An Introduction to Design Science*. Springer International Publishing: 39–73.

Jonathan, W. 2010. *Essentials of business research: a guide to doing your research project*. www.methodspace.com.

Jose, A., Nandagopalan, S., Ubalanka, V. & Viswanath, D. 2021. Detection and classification of nutrient deficiencies in plants using machine learning. In *Journal of Physics: Conference Series*. IOP Publishing Ltd.

Kairuz, T., Crump, K., District, W., Board, H. & Brien, A.J.O. 2007. Perspectives on qualitative research. Part 2: Useful tools for data collection and analysis. , (March).

Kali, Y. & Linn, M. 2006. Science. *International Encyclopedia of Education*, 313(July): 468–474.

Kamelia, L., Rahman, T.K.B.A., Saragih, H. & Haerani, R. 2020. The comprehensive review on detection of macro nutrients deficiency in plants based on the image processing technique. In *Proceedings - 2020 6th International Conference on Wireless and Telematics, ICWT 2020*. Institute of Electrical and Electronics Engineers Inc.

Kamilaris, A. & Prenafeta-Boldú, F.X. 2018. A review of the use of convolutional neural networks in agriculture. *Journal of Agricultural Science*, 156(3): 312–322.

Kapur, R. 2018. Research Methodology: Methods and Strategies. https://www.researchgate.net/publication/324588113.

Khan, A., Anabia, S., Umme, Z. & Aqsa, S.Q. 2020. A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*.

Kihara, J., Bolo, P., Kinyua, M., Rurinda, J. & Piikki, K. 2020. Micronutrient deficiencies in African soils and the human nutritional nexus: opportunities with staple crops. *Environmental Geochemistry and Health*, 42(9): 3015–3033. https://doi.org/10.1007/s10653-019-00499-w.

Kim, M.-Y., Atakishiyev, S., Babiker, H.K.B., Farruque, N., Goebel, R., Zaïane, O.R., Motallebi, M.-H., Rabelo, J., Syed, T., Yao, H. & Chun, P. 2021. A Multi-Component Framework for the Analysis and Design of Explainable Artificial Intelligence. *Machine Learning and Knowledge Extraction*, 3(4): 900–921.

Kogilavani, S. V., Prabhu, J., Sandhiya, R., Kumar, M.S., Subramaniam, U.S., Karthick, A., Muhibbullah, M. & Imam, S.B.S. 2022. COVID-19 Detection Based on Lung Ct Scan Using Deep Learning Techniques. *Computational and Mathematical Methods in Medicine*, 2022.

Kothari. 2004. Research Methodology Methods and Techniques.

Krizhevsky, A., Sutskever, I. & Hinton, G.E. 2012. *ImageNet Classification with Deep Convolutional Neural Networks*. http://code.google.com/p/cuda-convnet/.

Kumar, J.M. & Debika, B. 2018. *Emerging Technology in Modelling and Graphics*. http://www.springer.com/series/11156.

Lakkaraju, H., Bach, S.H. & Leskovec, J. 2016. Interpretable decision sets: A joint framework for description and prediction. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery: 1675–1684.

Lecun, Y., Bengio, Y. & Hinton, G. 2015. Deep learning. *Nature*, 521(7553): 436–444.

Leena, N. & Saju, K.K. 2018. Vision based nutrient deficiency classification in maize plants using multi class support vector machines. *AIP Conference Proceedings*, 1952.

Leonardo, M.M., Carvalho, T.J., Rezende, E., Zucchi, R. & Faria, F.A. 2019. Deep Feature-Based Classifiers for Fruit Fly Identification (Diptera: Tephritidae). In *Proceedings - 31st Conference on Graphics, Patterns and Images, SIBGRAPI 2018*. Institute of Electrical and Electronics Engineers Inc.: 41–47.

Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. 2021. Explainable ai: A review of machine learning interpretability methods. *Entropy*, 23(1): 1–45.

Lipton, Z. 2018. The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. , Volume 16(3).

Liu, W., Wang, Z., Liu, X., Liu, Y. & Alsaadi, F.E. 2016. A Survey of Deep Neural Network Architectures and Their Applications. *Neurocomputing*. http://dx.doi.org/10.1016/j.neucom.2016.12.038.

Liu, X., Zhang, F., Hou, Z., Wang, Z., Mian, L., Zhang, J. & Tang, J. 2020. Self-supervised Learning: Generative or Contrastive. http://arxiv.org/abs/2006.08218.

Liu, Y., Zhou, Y., Wen, S. & Tang, C. 2014. A Strategy on Selecting Performance Metrics for Classifier Evaluation. *International Journal of Mobile Computing and Multimedia Communications*, 6(4): 20–35.

Lu, T., Han, B., Chen, L., Yu, F. & Xue, C. 2021. A generic intelligent tomato classification system for practical applications using DenseNet-201 with transfer learning. *Scientific Reports*, 11(1).

Lundberg, S.M. & Lee, S.I. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 2017-Decem(Section 2): 4766–4775.

Mccann, B., Bradbury, J., Xiong, C. & Socher, R. 2017. *Learned in Translation: Contextualized Word Vectors*.

Mccauley, A. 2013. *Plant Nutrient Functions and Deficiency and Toxicity Symptoms*.

McNabb, D.E. 2010. *Research Methods for Political Science: quantitative and qualitative approaches*.

Md Humaion Kabir, M., Salman, H., Shafi, A., Samanta Tabassum, P., Rabeya Khatun, M. & Mehedi, H. 2022. *Plant Leaf Disease Detection using Transfer Learning and Explainable AI*. IEEE.

Md Zahangir, A., Tarek M., T., Christopher, Y., Stefan, W. & Paheding, S. 2018. The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches.

Mehdipour Ghazi, M., Yanikoglu, B. & Aptoula, E. 2017. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*, 235: 228–235.

Minni, N., Rehna, N., Nadu, T. & Symptoms, B.D. 2016. Detection of Nutrient Deficiencies in Plant Leaves using Image Processing. , (84): 84–87.

Misheva, B.H., Osterrieder, J., Hirsa, A., Kulkarni, O. & Lin, S.F. 2021. Explainable AI in Credit Risk Management. http://arxiv.org/abs/2103.00949.

Mohanty, S.P., Hughes, D.P. & Salathé, M. 2016. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7(September): 1–10.

Mohseni, S., Zarei, N. & Ragan, E.D. 2021. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Transactions on Interactive Intelligent Systems*, 11(3–4): 1–45.

Morshed, Md.S., Ahmed, S., Ahmed, T., Islam, M.U. & Rahman, A.B.M.A. 2022. Fruit Quality Assessment with Densely Connected Convolutional Neural Network. http://arxiv.org/abs/2212.04255.

Moujahid, H., Cherradi, B., Al-Sarem, M., Bahatti, L., Eljialy, A.B.A.M.Y., Alsaeedi, A. & Saeed, F. 2022. Combining cnn and grad-cam for covid-19 disease prediction and visual explanation. *Intelligent Automation and Soft Computing*, 32(2): 723–745.

Noon, S.K., Amjad, M., Qureshi, M.A. & Mannan, A. 2020. Use of deep learning techniques for identification of plant leaf stresses: A review. *Sustainable Computing: Informatics and Systems*, 28.

Nti, I.K., Nyarko-Boateng, O. & Aning, J. 2021. Performance of Machine Learning Algorithms with Different K Values in K-fold CrossValidation. *International Journal of Information Technology and Computer Science*, 13(6): 61–71. https://www.mecs-press.org/ijitcs/ijitcs-v13-n6/v13n6-5.html.

Nutifafa, A. 2020. *Visual Symptoms: A Handy Tool in Identifying Nutrient Deficiency in Row Crops*.

O'Shea, K. & Nash, R. 2015. An Introduction to Convolutional Neural Networks. http://arxiv.org/abs/1511.08458.

Páez, A. 2019. The Pragmatic Turn in Explainable Artificial Intelligence (XAI). *Minds and Machines*, 29(3): 441–459.

Pandey, P., Pandey, N.S., Shamim, M., Srivastava, D., Dwivedi, D.K., Awasthi, L.P. & Singh, K.N. 2015. Molecular tools and techniques for detection and diagnosis of plant pathogens. In *Recent Advances in the Diagnosis and Management of Plant Diseases*. Springer India: 253–271.

Pandey, Prabhat. & Pandey, M.Mishra. 2016. *Research methodology : tools & techniques*.

Panesar, A. 2022. *Improving Visual Question Answering by Leveraging Depth and Adapting Explainability*.

Pang, B., Nijkamp, E. & Wu, Y.N. 2020. Deep Learning With TensorFlow: A Review. *Journal of Educational and Behavioral Statistics*, 45(2): 227–248.

Rakesh, S. & Indiramma, M. 2022. Explainable AI for Crop disease detection. In *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. IEEE: 1601–1608. https://ieeexplore.ieee.org/document/10074303/.

Rawat, W. & Wang, Z. 2017. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9): 2352–2449.

Ribeiro, M.T., Singh, S. & Guestrin, C. 2016. 'Why Should I Trust You?' Explaining the Predictions of Any Classifier. *NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session*: 97–101.

Robinson, N.C. 2015. *Research methodology: tools and techniques*. www.sagepublications.com.

Rodrigo, A.R.S.P. 2020. *Deficiency Identification of Greenhouse Lettuce using Explainable AI*.

Rudin, C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5): 206–215.

Sabri, N., Kassim, N.S., Ibrahim, S., Roslan, R., Mangshor, N.N.A. & Ibrahim, Z. 2020. Nutrient deficiency detection in maize (Zea mays L.) leaves using image processing. *IAES International Journal of Artificial Intelligence*, 9(2): 304–309.

Sathyavani, R., Jaganmohan, K. & Kalaavathi, B. 2023. Sailfish optimization algorithm with deep convolutional neural network for nutrient deficiency detection in rice plants. *Journal of Pharmaceutical Negative Results* ¦, 14.

Saunders, M.N.K., Lewis, P. & Thornhill, A. 2019. *'Research Methods for Business Students' Chapter 4: Understanding research philosophy and approaches to theory development*. www.pearson.com/uk.

Saunders, M.N.K., Lewis, P. & Thornhill, Adrian. 2007. *Research methods for business students*. Financial Times/Prentice Hall.

Sawyer, J. 2004. *Integrated Pest Management Nutrient Deficiencies and Application Injuries in Field Crops Nitrogen deficiency in corn*.

Schmidt, S.B., Jensen, P.E. & Husted, S. 2016. Manganese Deficiency in Plants: The Impact on Photosystem II. *Trends in Plant Science*, 21(7): 622–632.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. & Batra, D. 2017. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. http://gradcam.cloudcv.org.

Setiawan, W. & Rulaningtyas, R. 2023. Visual Explanation of Maize Leaf Diaseases Classification using Squeezenet and Gradient-Weighted Class Activation Map. In *AIP Conference Proceedings*. American Institute of Physics Inc.

Sharma, M., Nath, K., Sharma, R.K., Kumar, C.J. & Chaudhary, A. 2022. Ensemble Averaging of Transfer Learning Models for Identification of Nutritional Deficiency in Rice Plant. *Electronics (Switzerland)*, 11(1).

Shete, S., Gosavi, N., Bulbule, A., Patil, D. & Pawar, R. 2020. Foliar application of macro and micronutrients and productivity of oilseed crops: A review. *International Journal of Chemical Studies*, 8(2): 2851–2855.

Simonyan, K. & Zisserman, A. 2015. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*: 1–14.

Simul, M., Talukder, H. & Sarkar, A.K. 2022. *Nutrients Deficiency Diagnosis of Rice Crop with High Accuracy via Weighted Average Ensemble of Different Augmented Transfer Learning Architectures*. https://ssrn.com/abstract=4217286.

Singh, H., Roy, A., Setia, R.K. & Pateriya, B. 2021. Estimation of nitrogen content in wheat from proximal hyperspectral data using machine learning and explainable artificial intelligence (XAI) approach. *Modeling Earth Systems and Environment*, (0123456789). https://doi.org/10.1007/s40808-021-01243-z.

Singh, V. & Misra, A.K. 2017. Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, 4(1): 41–49. http://dx.doi.org/10.1016/j.inpa.2016.10.005.

Soper, D.S. 2021. Greed is good: Rapid hyperparameter optimization and model selection using greedy k-fold cross validation. *Electronics (Switzerland)*, 10(16).

Spann, T.M. & Schumann, A.W. 2010. *Mineral Nutrition Contributes to Plant Disease and Pest Resistance 1*. http://edis.ifas.ufl.edu.

Sparrow, R., Howard, M. & Degeling, C. 2021. Managing the risks of artificial intelligence in agriculture. *NJAS: Impact in Agricultural and Life Sciences*, 93(1): 172–196. https://doi.org/10.1080/27685241.2021.2008777.

Stone, M. 1974. *Cross-Validatory Choice and Assessment of Statistical Predictions*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. 2015. Going Deeper with Convolutions. : 1–9.

Timothy, spann & Arnold, schumann. 2009. *The Role of Plant Nutrients in Disease Development with Emphasis on Citrus and Huanglongbing*.

Tran, T.T., Choi, J.W., Le, T.T.H. & Kim, J.W. 2019. A comparative study of deep CNN in forecasting and classifying the macronutrient deficiencies on development of tomato plant. *Applied Sciences (Switzerland)*, 9(8).

Veldhuis, M.S., Ariëns, S., Ypma, R.J.F., Abeel, T. & Benschop, C.C.G. 2022. Explainable artificial intelligence in forensics: Realistic explanations for number of contributor predictions of DNA profiles. *Forensic Science International: Genetics*, 56.

Wang, H. 2017. On the Origin of Deep Learning. http://arxiv.org/abs/1702.07800.

Watchareeruetai, U., Noinongyao, P., Wattanapaiboonsuk, C., Khantiviriya, P. & Duangsrisai, S. 2018. Identification of Plant Nutrient Deficiencies Using Convolutional Neural Networks. *iEECON 2018 - 6th International Electrical Engineering Congress*: 1–4.

Wei, K., Chen, B., Zhang, J., Fan, S., Wu, K., Liu, G. & Chen, D. 2022a. Explainable Deep Learning Study for Leaf Disease Classification. *Agronomy*, 12(5).

Wei, K., Chen, B., Zhang, J., Fan, S., Wu, K., Liu, G. & Chen, D. 2022b. Explainable Deep Learning Study for Leaf Disease Classification. *Agronomy*, 12(5).

Wolanin, A., Mateo-Garciá, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liangzhi, Y. & Guanter, L. 2020. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environmental Research Letters*, 15(2).

Xu, P., Tan, Q., Zhang, Y., Zha, X., Yang, S. & Yang, R. 2022. Research on Maize Seed Classification and Recognition Based on Machine Vision and Deep Learning. *Agriculture (Switzerland)*, 12(2).

Xu, Z., Guo, X., Zhu, A., He, X., Zhao, X., Han, Y. & Subedi, R. 2020. Using deep convolutional neural networks for image-based diagnosis of nutrient deficiencies in rice. *Computational Intelligence and Neuroscience*, 2020.

Yi, J., Krusenbaum, L., Unger, P., Hüging, H., Seidel, S.J., Schaaf, G. & Gall, J. 2020. Deep learning for non-invasive diagnosis of nutrient deficiencies in sugar beet using RGB images. *Sensors (Switzerland)*, 20(20): 1–19.

Zeng, X., Song, F., Li, Z., Chusap, K. & Liu, C. 2021. Human-in-the-loop model explanation via verbatim boundary identification in generated neighborhoods. http://arxiv.org/abs/2106.13093.

Zhong, Y., He, J. & Chalise, P. 2020. Nested and repeated cross validation for classification model with high-dimensional data. *Revista Colombiana de Estadistica*, 43(1): 103–125.

Zhou, J., Gandomi, A.H., Chen, F. & Holzinger, A. 2021. Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics (Switzerland)*, 10(5): 1–19.

APPENDICES

APPENDICES A: ETHICS CERTIFICATE

Appendix A serverd the objective of adhering to the ethical guidelines delineated in the Cape Peninsula University of Technology research code of ethics.

Cape Peninsula
University of Technology
creating futures

PO Box 1906, Bellville,7535 Symphony Way, Bellville, Cape Town, South Africa
+ 27 (0)21 959 6767 www.facebook.com/cput.ac.za info@cput.ac.za www.cput.ac.za

Office of the Research Ethics Committee
Faculty of Informatics and Design
Room 2.09
80 Roeland Street
Cape Town
Tel: 021-469 1012
Email: ndedem@cput.ac.za
Secretary: Mziyanda Ndede

20 September 2022

Mr Junior Mkhatshwa
c/o Department of Information Technology
CPUT

Reference no: 214011097/2022/21

Project title: Comparative analysis of explainable deep learning models for identification of plant nutrient deficiencies

Approval period: 20 September 2022 – 31 December 2023

This is to certify that the Faculty of Informatics and Design Research Ethics Committee of the Cape Peninsula University of Technology conditionally approves the methodology and ethics of Mr Junior Mkhatshwa (214011097) for Master of ICT.

Any amendments, extension or other modifications to the protocol must be submitted to the Research Ethics Committee for approval.

The Committee must be informed of any serious adverse event and/or termination of the study.

Dr Blessing Makwambeni
Acting Chair: Research Ethics Committee
Faculty of Informatics and Design
Cape Peninsula University of Technology

APPENDICES B: PROFESSIONAL EDITOR'S CERTIFICATE

Appendix B presents the certificate document issued by a professional and accredited editior of the Cape Peninsula University of Technology, who was responsible for editing this thesis.

NewsPlus Communications (Pty) Ltd
(2016/523137/07)
Unit 15 Silver Willows, 106 Furrow Road, Equestria Pretoria 0184
Cell Phone: +27 62 240 1084
Email: editor@newsplus.ng
odianosent@gmail.com
(Editing, Proofreading, News Writing and Reporting, Speech Writing, Rapporteuring, Web Content Management)

MEMBER,

Professional EDITORS Guild

TO WHOM IT MAY CONCERN

This document certifies that I, Anthony A. Ekata, a member of the Professional Editors' Guild South Africa, Nigerian Guild of Editors, and an accredited editor of the Cape Peninsula University of Technology, Cape Town, South Africa, edited the Master of Technology: Information Technology thesis with the title below for proper Language and Style.

Title:

Comparative analysis of explainable deep learning models for identification of plant nutrient deficiencies

By

Junior Mkhatshwa

Faculty of Informatics and Design, Cape Peninsula University of Technology

Date September 15, 2023

SIGNED

Anthony A. Ekata

+27 62 240 1084