UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL

Título
Artificial intelligence and non-invasive sensing technologies for early detection of downy mildew in grapevine
Autor/es
Inés Hernández Casado
Director/es
Javier Tardáguila Laso, Salvador Gutiérrez Salcedo y Juan Félix San Juan Díaz
Facultad
Facultad de Ciencia y Tecnología
Titulación
Departamento
Matemáticas y Computación
Curso Académico
2024-2025

Artificial intelligence and non-invasive sensing technologies for early detection of downy mildew in grapevine

PhD Thesis

Inés Hernández Casado

Written under the supervision of

Prof. Dr. Manuel Javier Tardáguila Laso
Dr. Salvador Gutiérrez Salcedo
Prof. Dr. Juan Félix San Juan Díaz

UNIVERSIDAD DE LA RIOJA

Programa de Doctorado en Matemáticas y Computación

Logroño, January 2025

The work leading to these results was funded by the European H2020 FetOpen (Future and Emerging Technologies) project entitled "NoPest: Novel Pesticides for a Sustainable Agriculture" from the European Commission under grant agreement ID 828940.

The PhD thesis was also funded by a pre-doctoral contract for the training of research personnel (FPI) granted by the University of La Rioja in resolution 1150/2020 of 9 December 2020. It was also supported by the grants for conducting doctoral theses of the University of La Rioja (ATUR), in the resolutions of 2022, 2023 and 2024.

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisors, Professor Javier Tardáguila Laso, Doctor Salvador Gutiérrez Salcedo and Professor Juan Félix San Juan Díaz, for their invaluable guidance, patience and support throughout this process. Their commitment and trust were essential for the development of this work.

I am also grateful to Professor Pedro Melo Teixeira Pinto and Doctor Rui Manuel Machado Silva, of the Centre for the Research and Technology of Agro-Environmental and Biological Sciences (CITAB) at the University of Trás-os-Montes e Alto Douro in Vila Real, Portugal. Their expertise and support during my doctoral stay in 2023 and 2024 significantly enriched my knowledge of artificial intelligence.

I extend my gratitude to my colleagues in the Televitis research group, Professor María Paz Diago Santamaría, Doctor Juan Fernández Novales, Doctor Fernando Palacios, Doctor Sara Ceballos Marcaida, and especially to Ignacio Barrio Fernández and Rubén Íñiguez Mangado for the ideas shared and for making every day a space of mutual learning and for their support on a personal level, without whom this thesis would not have been possible. I could not forget about the rest of the collaborators of the research group during the period of my PhD thesis. I am fortunate to have had the opportunity to work with all of them over the years.

I would also like to thank my lifelong friends for their support and the good times we have shared, and the friends I have met along the way, who, despite cultural differences, have become like family to me, especially Esra for her support throughout these last months.

Finally, I would like to express my deepest gratitude to my family for their unconditional support, and to my brother-in-law Ignacio for his valuable advice. In particular, I am especially grateful to my sister Elena, my best friend, who has been my main pillar, constant support and source of strength at every stage of this journey.

Agradecimientos

Quisiera expresar mi más sincero agradecimiento a mis directores de tesis, el Catedrático Javier Tardáguila Laso, el Doctor Salvador Gutiérrez Salcedo y el Catedrático Juan Félix San Juan Díaz, por su inestimable orientación, paciencia y apoyo a lo largo de todo este proceso. Su compromiso y confianza fueron esenciales para el desarrollo de este trabajo.

Agradezco también al Profesor Pedro Melo Teixeira Pinto y al Doctor Rui Manuel Machado Silva, del Centro de Investigación y Tecnología Agroambiental y Biológica (CITAB) de la Universidad de Trás-os-Montes e Alto Douro en Vila Real, Portugal. Su experiencia y apoyo durante mi estancia doctoral en 2023 y 2024 enriquecieron significativamente mis conocimientos sobre inteligencia artificial.

Extiendo mi agradecimiento a mis compañeros del grupo de investigación Televitis, la Profesora María Paz Diago Santamaría, el Doctor Juan Fernández Novales, el Doctor Fernando Palacios, la Doctora Sara Ceballos Marcaida, y especialmente a Ignacio Barrio Fernández y Rubén Íñiguez Mangado por las ideas compartidas y por hacer de cada día un espacio de aprendizaje mutuo y por su apoyo a nivel personal, sin los cuales esta tesis no hubiera sido posible. No podía olvidarme del resto de colaboradores del grupo de investigación durante el periodo de realización de mi tesis doctoral. Soy afortunada de haber tenido la oportunidad de trabajar con todos ellos a lo largo de estos años.

También me gustaría dar las gracias a mis amigas de toda la vida por su apoyo y los buenos momentos que hemos compartido, y a los amigos que he conocido por el camino, que, a pesar de las diferencias culturales, se han convertido en una familia para mí, especialmente a Esra por su apoyo a lo largo de estos últimos meses.

Por último, quiero expresar mi más profunda gratitud a mi familia por su apoyo incondicional, y a mi cuñado Ignacio por sus valiosos consejos. En particular, quisiera agradecer muy especialmente a mi hermana Elena, mi mejor amiga, que ha sido mi pilar principal, apoyo constante y fuente de fortaleza en cada etapa de este viaje.

Abstract

Diseases and pests in agriculture represent a major problem worldwide, severely impacting crop quality and yield. Among them, downy mildew is a particularly devastating disease affecting grapevine. Early detection is crucial for timely intervention, preventing disease spread and reducing chemical treatments. Traditional evaluation relies on experts, which can be laborious, subjective and time-consuming. The integration of artificial intelligence into agricultural practices presents a promising solution for disease management, facilitating the automation of qualitative and quantitative disease assessment.

The main objective of the PhD thesis was to develop new artificial intelligence and computer vision-based methods for early assessment of grapevine downy mildew using non-invasive sensing technologies under laboratory and field conditions. In particular, the following objectives were proposed: i) the exploration of artificial intelligence and non-invasive technologies to evaluate downy mildew under laboratory conditions; ii) the development and validation of a method to estimate downy mildew severity under laboratory conditions combining fuzzy logic and computer vision; iii) the use of convolutional neural networks and explainable artificial intelligence to early detect downy mildew under laboratory conditions; iv); the in-field detection and localisation of downy mildew applying explainable deep learning; and v) the use of deep semantic segmentation to assess downy mildew severity in images taken in commercial vineyards.

For the first objective, artificial intelligence was used for analysing RGB and hyperspectral images of grapevine leaf discs. Spectral pre-processing, computer vision and machine learning were used to identify downy mildew infection in hyperspectral images. At the same time, classic computer vision was used to locate the symptoms in RGB images. The results demonstrated the potential of artificial intelligence and non-invasive technologies to early detect downy mildew and to estimate its severity accurately and objectively.

For the second objective, classic computer vision was used to localise downy mildew symptoms on RGB images of grapevine leaf discs. Then, fuzzy logic was used to evaluate the pixels detected as symptoms with a degree of infection according to their intensity. The results demonstrated that computer vision and fuzzy logic can automatically and accurately estimate

the severity of grapevine downy mildew under laboratory conditions.

For the third objective, convolutional neural networks were applied to early detect downy mildew and classify disease stages in RGB images of grapevine leaf discs. In addition, Grad-CAM was used to interpret model predictions. The results highlighted the accurate early detection of grapevine downy mildew under laboratory conditions using low-cost techniques.

For the fourth objective, a sliding window was used for analysing the grapevine canopy in images captured considering the variability of field conditions. Convolutional neural networks and vision transformers used transfer-learning for detecting regions with downy mildew in the canopy. Predictions were interpreted with explainable artificial intelligence methods. The results highlighted the use of convolutional neural networks for the automatic and explainable detection and localisation of grapevine downy mildew under field conditions.

Finally, different semantic segmentation architectures were compared to detect downy mildew symptoms in grapevine canopy images. Imbalance problems due to small symptom size were reduced with data augmentation, MixUp, oversampling and undersampling techniques. Neural networks trained with light-weight encoders and using the Dice loss function allowed accurate and fast assessment of downy mildew severity in grapevine under field conditions.

The results of the research presented in this PhD thesis demonstrated the capability of artificial intelligence and computer vision for objective, rapid and accurate early assessment of grapevine downy mildew under both laboratory and field conditions. The potential adaptability of these methods to other crops, diseases and pests offers important advances in precision agriculture. Furthermore, the integration of these methods on mobile platforms, such as tractors, would allow for enhanced disease management over large crop areas, optimising monitoring and intervention directly in the field.

Resumen

Las enfermedades y plagas en cultivos agrícolas representan un grave problema en todo el mundo, repercutiendo gravemente en la calidad y rendimiento de los cultivos. Entre ellas, el mildiu es una enfermedad especialmente devastadora de la vid. La detección precoz es crucial para intervenir a tiempo, evitando la propagación de la enfermedad y reduciendo los tratamientos químicos. La evaluación tradicional depende de expertos, lo que puede resultar laborioso, subjetivo y lento. La integración de la inteligencia artificial en las prácticas agrícolas presenta una solución prometedora para la gestión de enfermedades, facilitando la automatización de la evaluación cualitativa y cuantitativa de las enfermedades.

El objetivo principal de la tesis doctoral era desarrollar nuevos métodos basados en inteligencia artificial y visión artificial para la evaluación temprana del mildiu de la vid utilizando sensores no invasivos en condiciones de laboratorio y de campo. En concreto, se propusieron los siguientes objetivos: i) la exploración de la inteligencia artificial y tecnologías no invasivas para evaluar el mildiu en condiciones de laboratorio; ii) el desarrollo y validación de un método para estimar la severidad de mildiu en condiciones de laboratorio combinando lógica difusa y visión artificial; iii) el uso de redes neuronales convolucionales e inteligencia artificial explicable para la detección temprana de mildiu en condiciones de laboratorio; iv); la detección y localización de mildiu en campo aplicando aprendizaje profundo explicable; y v) el uso de segmentación semántica profunda para evaluar la severidad de mildiu en imágenes tomadas en viñedos comerciales.

Para el primer objetivo, se utilizó inteligencia artificial para analizar imágenes RGB e hiperespectrales de discos foliares de vid. Se utilizó preprocesamiento espectral, visión artificial y aprendizaje automático para identificar la infección de mildiu en imágenes hiperespectrales. Al mismo tiempo, se utilizó visión artificial clásica para localizar los síntomas en imágenes RGB. Los resultados demostraron el potencial de la inteligencia artificial y las tecnologías no invasivas para detectar precozmente el mildiu y estimar su severidad de forma precisa y objetiva.

Para el segundo objetivo, se utilizó la visión artificial clásica para localizar los síntomas de mildiu en imágenes RGB de discos foliares de vid. A continuación, se utilizó la lógica difusa para evaluar los píxeles detectados

como síntomas con un grado de infección en función de su intensidad. Los resultados demostraron que la visión artificial y la lógica difusa pueden estimar automáticamente y con precisión la severidad de mildiu de la vid en condiciones de laboratorio.

Para el tercer objetivo, se aplicaron redes neuronales convolucionales para detectar precozmente el mildiu y clasificar las fases de la enfermedad en imágenes RGB de discos foliares de vid. Además, se utilizó Grad-CAM para interpretar las predicciones del modelo. Los resultados resaltaron la precisión de la detección precoz del mildiu de la vid en condiciones de laboratorio utilizando técnicas de bajo coste.

Para el cuarto objetivo, se utilizó una ventana deslizante para analizar el dosel de la vid en imágenes tomadas considerando la variabilidad de las condiciones de campo. Las redes neuronales convolucionales y los transformadores de visión utilizaron el aprendizaje por transferencia para detectar regiones con mildiu en el dosel. Las predicciones se interpretaron con métodos de inteligencia artificial explicable. Los resultados remarcaron el uso de redes neuronales convolucionales para la detección y localización automática y explicable del mildiu de la vid en condiciones de campo.

Por último, se compararon diferentes arquitecturas de segmentación semántica para detectar síntomas de mildiu en imágenes del dosel de la vid. Los problemas de desequilibrio debidos al pequeño tamaño de los síntomas se redujeron con técnicas de aumento de datos, MixUp, sobremuestreo y submuestreo. Las redes neuronales entrenadas con codificadores ligeros y utilizando la función de pérdida Dice permitieron una evaluación precisa y rápida de la severidad de mildiu en la vid en condiciones de campo.

Los resultados de la investigación presentada en esta tesis doctoral demostraron la capacidad de la inteligencia artificial y la visión artificial para la evaluación temprana objetiva, rápida y precisa del mildiu de la vid tanto en condiciones de laboratorio como de campo. La potencial adaptabilidad de estos métodos a otros cultivos, enfermedades y plagas ofrece importantes avances en la agricultura de precisión. Además, la integración de estos métodos en plataformas móviles, como tractores, permitiría mejorar la gestión de enfermedades en grandes extensiones de cultivo, optimizando el seguimiento y la intervención directamente en el campo.

Articles

This PhD thesis has been made as a compendium of the following published works:

1. HERNANDEZ, I., GUTIERREZ, S., CEBALLOS, S., IÑIGUEZ, R., BARRIO, I., TARDAGUILA, J. (2021). Artificial intelligence and novel sensing technologies for assessing downy mildew in grapevine. Horticulturae, 7(5), 103. https://doi.org/10.3390/horticulturae7050103. Impact Factor: 3.1 (Q1)
2. HERNANDEZ, I., GUTIERREZ, S., CEBALLOS, S., PALACIOS, F., TOFFOLATTI, S. L., MADDALENA, G., DIAGO, M. P., TARDAGUILA, J. (2022). Assessment of downy mildew in grapevine using computer vision and fuzzy logic. Development and validation of a new method. OENO One, 56(3), 41–53. https://doi.org/10.20870/oeno-one.2022.56.3.5359. Impact Factor: 2.2 (Q3)
3. HERNÁNDEZ, I., GUTIÉRREZ, S., TARDAGUILA, J. (2024). Image analysis with deep learning for early detection of downy mildew in grapevine. Scientia Horticulturae, 331, 113155. https://doi.org/10.1016/j.scienta.2024.113155. Impact Factor: 3.9 (Q1)
4. HERNÁNDEZ, I., GUTIÉRREZ, S., BARRIO, I., ÍÑIGUEZ, R., TARDÁGUILA, J (2024). In-field disease symptom detection and localisation using explainable deep learning: Use case for downy mildew in grapevine. Computers and Electronics in Agriculture, 226(September), 109478. https://doi.org/10.1016/j.compag.2024.109478. Impact Factor: 7.7 (Q1)

Another publication, already submitted and currently under review, is also included in the PhD thesis:

1. HERNÁNDEZ, I., SILVA, R., MELO-PINTO, P., GUTIÉRREZ, S., TARDÁGUILA, J. Early detection of downy mildew in vineyards using deep neural networks for semantic segmentation. Biosystems Engineering (under review). Impact Factor: 4.4 (Q1)

Contents

Chapter 1

Introduction

1.1 Disease assessment in agriculture

Crop diseases result in considerable economic losses in agricultural production worldwide. A diverse range of pathogens, including fungi, bacteria, mycoplasmas and viruses can cause significant crop diseases that pose a considerable threat to global agriculture. These pathogens may affect plant growth, yield, and quality, and under severe infestations, can even lead to total crop failure (Lee and Tardaguila, 2023). Infected plants frequently exhibit diverse visual and characteristic symptoms on different organs, including stems, leaves, and fruits, while some infections may initially be asymptomatic. Effective monitoring and timely interventions are of paramount importance for the mitigation of these threats, ensuring food quality, and improving agricultural productivity. Such measures assist in the reduction of crop damage, the reduction of reliance on chemical treatments, the analysis of plant breeding, or the understanding of biological processes related to diseases (Bock et al., 2010).

Disease assessment, whether in the laboratory or the field, is essential for understanding plant-pathogen interactions and implementing effective management strategies. In addition, key parameters, such as incidence (the proportion of infected plants) and severity (the percentage of affected tissue), are crucial for determining the extent of an outbreak (Madden et al., 2017). Traditional methods depend on trained professionals who perform intricate analyses such as DNA assays to detect diseases (Lee and Tardaguila, 2023), or visual assessments to identify symptoms like discolouration or lesions in the plants. While these approaches are often effective, they are frequently costly, subjective, time-consuming, and susceptible to errors, particularly when symptoms are subtle or cryptic (Bock et al., 2010; Paulus et al., 1997). Advances in digital phenotyping, such as high-throughput imaging, have significantly improved the precision and efficiency of these evaluations (Mahlein et al., 2019). By leveraging these technologies, farmers and researchers can enhance early detection, measure disease severity accurately, and identify pathogens with greater reliability.

1.1.1 Sensing technologies for disease detection

The integration of non-invasive and proximal sensing technologies in agriculture has revolutionised disease detection, enabling rapid and accurate identification of pathogens affecting crops without damaging the plants. Techniques such as thermography, spectroscopy, chlorophyll fluorescence, RGB imaging, multispectral imaging (MSI), and hyperspectral imaging (HSI) offer numerous advantages over conventional diagnostic methods, including objectivity, efficiency, cost-effectiveness, and reliability (Tardaguila et al., 2021). Among these methods, image-based detection using RGB imaging or hyperspectral imaging has been particularly effective in monitoring plant health and detecting diseases (Lee and Tardaguila, 2023; Mahlein, 2016). Both techniques offer distinct advantages and rely on different underlying principles:

- **RGB imaging** captures visual information using the red, green, and blue colour channels, which correspond to the way the human eye perceives colour. This method involves the use of digital cameras equipped with sensors that detect light in these three primary colour bands. The captured images are then processed to produce a composite image that represents the visible spectrum. RGB imaging is particularly useful for visual inspection, allowing for the identification of visible symptoms of diseases, such as discolouration, spots, or lesions in leaves and fruits (Barbedo, 2013).
- **Hyperspectral imaging (HSI)**, captures a wide spectrum of light beyond the visible range, providing detailed spectral information for each pixel in an image. This technique involves the use of hyperspectral sensors that collect data across numerous narrow spectral bands, ranging from the ultraviolet to the near-infrared regions of the electromagnetic spectrum. The resulting hyperspectral images contain rich spectral information that can be used to identify subtle biochemical changes in plants that could precede visible symptoms. This capability allows for the early detection of diseases, enabling timely intervention. Due to the complexity of natural and irregular illumination, HSI data is often collected in controlled laboratory settings. Thus, studies have demonstrated the efficacy of HSI in accurately detecting early-stage diseases such as pear black spot (Pan et al., 2019) or apple rottenness (Zhang et al., 2015).

Sensing technologies play a critical role in agricultural data collection, utilising manual methods as well as mobile platforms, including space-based, air-based, and ground-based systems (Tardaguila et al., 2021). Manual data acquisition, particularly in laboratory settings, remains essential for plant disease research, offering precise, localised observations and facilitating the validation of automated systems. Despite being labour-intensive, these methods are particularly valuable for conducting controlled experiments. However, advancements in sensor technology and automation have made mobile platforms, particularly ground-based systems, increasingly indispensable for precise field-based disease monitoring (Lee and Tardaguila, 2023). In contrast to satellite or drone-based systems, which may exhibit lower spatial resolution and fail to discern intricate plant-level characteristics, ground-based systems offer high-resolution proximal sensing. Such systems can be equipped with sophisticated sensors that permit the collection of georeferenced data from the plant canopy itself, thus facilitating the detection of subtle physiological indicators that may otherwise be overlooked by remote sensing at higher altitudes. Furthermore, ground-based platforms are less susceptible to environmental influences such as cloud cover and atmospheric interference, ensuring consistent data acquisition. Integrated with agricultural machinery, these platforms enable on-the-go monitoring across extensive areas, significantly reducing operational costs and supporting precision agriculture practices like variable rate application (VRA) of fungicides (Román et al., 2020).

1.1.2 Early disease detection

The detection of plant diseases in their early stages is of paramount importance, particularly in the field, in order to facilitate effective disease management and sustainable agricultural practices (Lee and Tardaguila, 2023). Timely identification of infections allows for interventions that can prevent secondary infections or cross-contamination with neighbouring plants, controlling the disease before it progresses to more severe and costly stages. Farmers can adopt sustainable practices, such as treatments only to affected plants or areas within a field (Román et al., 2020). This approach reduces the total volume of chemicals used, leading to lower economic costs and decreases the risk of chemical residues in crops. Additionally, targeted interventions minimise the likelihood of pathogens developing resistance to treatments, a growing challenge in modern agriculture. These measures contribute to more effective disease control while preserving the health of broader crop systems and promoting ecosystem balance.

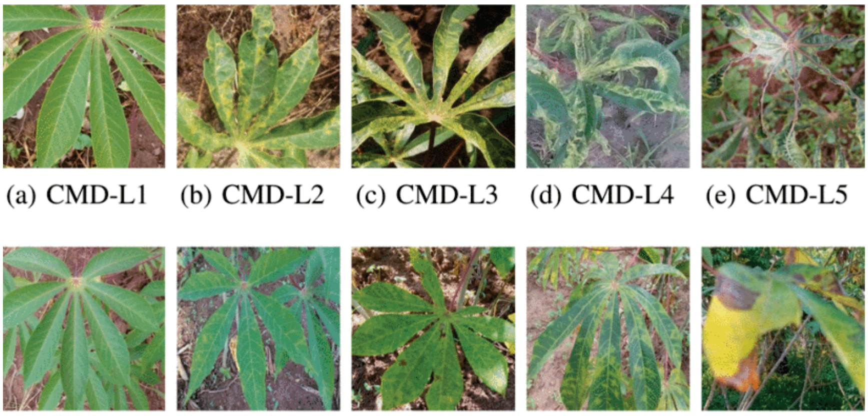
In laboratory settings, early detection plays a pivotal role in the development of new treatments and resistant crop varieties. This detection is usually made with DNA-based or serological techniques. The use of DNA-based methods, including polymerase chain reaction (PCR), enables the precise identification of pathogens through the detection of their genetic material. Serological techniques, such as the enzyme-linked immunosorbent assay (ELISA), are capable of detecting specific proteins that are associated with pathogens. Although these techniques are highly accurate, they necessitate the collection, processing, and analysis of samples, which can require several days and may not be feasible for large-scale applications (Martinelli et al., 2015). As mentioned in Section 1.1.1, the application of non-invasive technologies, such as hyperspectral and RGB imaging in conjunction with artificial intelligence, is transforming the early detection of plant disease. Hyperspectral imaging can capture subtle biochemical changes in plants before visible symptoms appear, as demonstrated by Gao et al. (2020) in the detection of leafroll disease in grapevines. RGB imaging, while simpler, provides high-resolution visual data to identify early structural or colour changes in plants. These changes may help to accurately detect early symptoms of diseases in crops such as apple (Bansal et al., 2021) or sugar beet (Adem et al., 2023).

1.1.3 Disease quantification

Quantification of diseases in agriculture is crucial to effective disease management, complementing early detection by providing essential information for evaluating control strategies (Bock et al., 2010). Disease infection is commonly quantified in terms of incidence, which represents the proportion of affected plant units within a population, and severity, which refers to the percentage of the plant exhibiting visible symptoms of the disease (Madden et al., 2017). Accurate quantification of these metrics is crucial for monitoring crop health, optimising intervention strategies, and improving agricultural productivity. In the field, disease quantification helps to estimate potential yield losses, guide the targeted application of treatments, and adjust crop practices to minimise further damage. Similarly, in laboratory settings, severity assessments are vital for analysing disease progression, evaluating the effectiveness of experimental treatments, and studying plant-pathogen interactions under controlled conditions. They also play a critical role in assessing cultivar resistance to pathogens.

However, disease quantification presents a challenge in both field and laboratory contexts. Traditional methods based on human visual observations are inherently subjective and susceptible to variability, leading to inconsistent results (Bock et al., 2010). The complexity of disease symptoms—varying in intensity, distribution, and appearance—further complicates the standardisation of assessments. In response, advances in image analysis and computer vision have provided transformative tools that offer more accurate, reproducible, and scalable alternatives to manual methods

(Bock et al., 2020). For instance, studies such as that conducted by Mwebaze and Owomugisha (2016) demonstrated its application for the accurate classification of leaves according to disease severity, utilising images captured with a smartphone (Figure 1.1). This research illustrated the potential of integrating low-cost, accessible technologies with artificial intelligence to overcome the limitations of disease assessment, particularly in resource-limited settings.



(a) CMD-L1 (b) CMD-L2 (c) CMD-L3 (d) CMD-L4 (e) CMD-L5

(f) CBSD-L1 (g) CBSD-L2 (h) CBSD-L3 (i) CBSD-L4 (j) CBSD-L5

Figure 1.1: Cassava leaves associated with five severity levels for mosaic disease (CMD) and brown steak disease (CBSD). Source: Mwebaze and Owomugisha (2016)

1.1.4 In-field crop disease assessment

The primary objective of laboratory and field-based assessments is to control the effects of diseases on crops. As highlighted in the previous sections, field-based crop disease assessment plays a significant role in disease management (Lee and Tardaguila, 2023). Early detection of diseases in the field is essential for timely interventions, limiting disease spread, reducing crop losses, and optimising resource utilisation. Accurate identification and quantification of disease symptoms in the plants are essential for monitoring the severity and distribution of infections across extensive crops, evaluating the effectiveness of control measures, and contributing to environmentally sustainable agricultural practices. Traditional methods, carried out by trained personnel, analysing visual symptoms directly in the field, are often subjective, inconsistent, and time-consuming, particularly when applied to large-scale crops. The variability and complexity of the disease symptoms under natural conditions, including the presence of similar symptoms or damage to the plant caused by other diseases, pests, or crop machines, further challenges the accuracy of these assessments. In addition, the heterogeneity of disease distribution within fields, where some areas may be heavily infected while others remain unaffected, requires extensive sampling to obtain a representative assessment. These limitations underscore the necessity for advanced technologies capable of delivering precise and real-time disease detection. The integration of innovative tools into disease management practices offers significant potential to address these challenges and transform agricultural monitoring (Abdullah et al., 2023).

Recent advancements in sensing technologies, artificial intelligence (AI), and decision-support systems have paved the way for precision agriculture, which employs data-driven methods to improve disease and pest management in crops (Tardaguila et al., 2021). These methods typically involve three fundamental steps: data acquisition, information extraction, and actionable management (Figure 1.2). Data acquisition utilises remote and proximal sensing technologies to capture a range of visual, spectral, and spatial information about plants. As commented in Section 1.1.1, image-based sensors, especially RGB images, are a useful tool for disease assessment under field conditions. Nevertheless, the practical deployment of image analysis in the field is confronted with a number of challenges, including variable lighting conditions, background noise, and the limited visibility of early-stage lesions. The analysis of high-resolution images has demonstrated the potential to address these challenges, facilitating the detection of subtle symptoms such as late blight in potatoes (Gao et al., 2021) or downy mildew in grapevine (Abdelghafour et al., 2020). Subsequently, artificial intelligence processes these data to detect disease symptoms, estimate their severity, and localise affected areas. In addition, the extraction of information with mobile sensing platforms such as GPS-equipped tractors or robots utilising non-invasive sensors could provide georeferenced disease assessments without damaging the plants. This is exemplified by Abdelghafour et al. (2020), who detected downy mildew in grapevines through the use of a ground-based imaging platform. These platforms have the potential to facilitate the extraction of comprehensive information across large fields, offering precise and spatially localised insights into crop health. Finally, the analyses derived from these platforms provide actionable insights that contribute to the formulation of targeted management strategies, including the precise application of treatments and the adjustment of cultivation practices, thus supporting the objective of sustainable crop production.

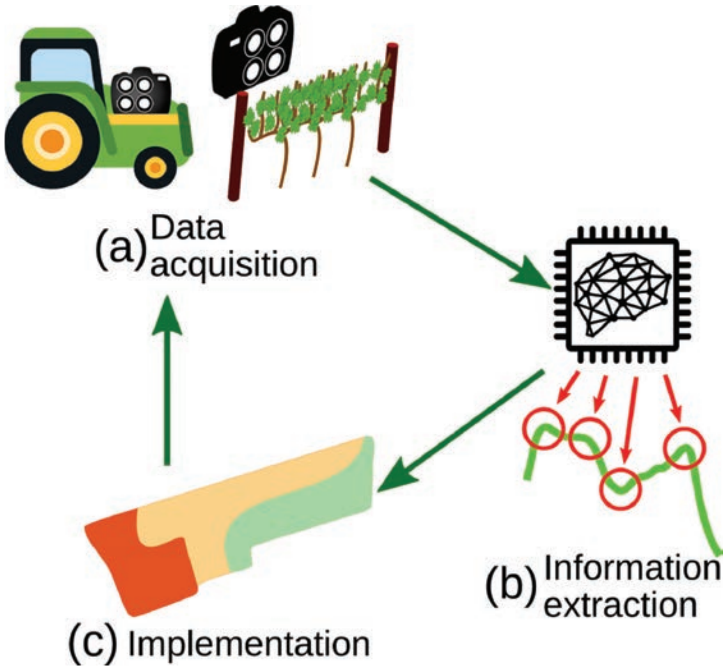
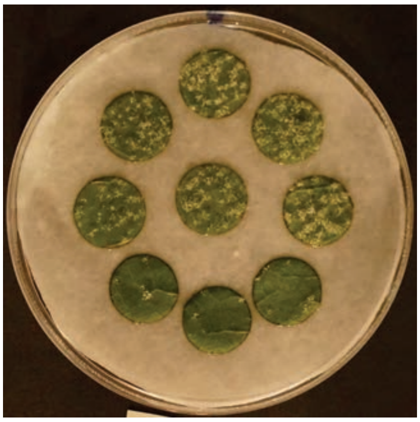


Figure 1.2: Precision agriculture process. Source: Tardaguila et al. (2021)

1.1.5 Downy mildew assessment in viticulture

Viticulture is a crucial agricultural sector that supports the global production of wine, table grapes, and must, contributing significantly to economic activity worldwide (Buonassisi et al., 2017). However, grapevine cultivation is highly vulnerable to diseases caused by fungi, bacteria, and viruses, which severely affect yield and fruit quality. Among these, downy mildew (*Plasmopara viticola*) stands out as one of the most destructive pathogens. Under warm and humid conditions, it has the capacity to infect multiple parts of the plant (Toffolatti et al., 2018). The first symptoms of the disease are characterised by oily spots on the abaxial surface of leaves, often accompanied by white sporulation on the adaxial side (Figure 1.3). If not effectively controlled, these symptoms can lead to rapid disease spread, significant crop losses, and costly protective measures.

Similarly to other diseases in agriculture, the detection of grapevine downy mildew is of critical importance for the advancement of treatment strategies and its effective management. In laboratory settings, symptoms are assessed through controlled inoculation and observation of symptom development by analysing sporulation on leaf discs (Toffolatti et al., 2018). The downy mildew assessment in the laboratory facilitates the precise identification and comprehensive assessment of the pathogen, thereby providing crucial insights for the advancement of novel fungicides, the optimisation of



(a) Sporulation in the laboratory

(b) Oil spots in the field

Figure 1.3: Example of downy mildew symptoms in grapevine leaves

existing treatments and the breeding of grapevine varieties with enhanced resistance. In the vineyard, oil spots usually represent an early indicator of downy mildew (Boso et al., 2005). The detection of these spots is usually made visually by experts. Field-based methods could provide rapid disease monitoring over extensive vineyard areas. The emergence of proximal sensing technologies and artificial intelligence (AI) has led to the development of transformative tools for improving the precision and consistency of downy mildew detection (Bock et al., 2020). For instance, AI-based approaches have been employed to accurately identify downy mildew and differentiate its symptoms from those of spider mite, despite their visual similarity (Gutiérrez et al., 2021). These advances may enable precise monitoring of downy mildew progression in both laboratory and field settings, thereby reducing reliance on subjective and time-consuming visual assessments.

1.2 Artificial intelligence

The term "artificial intelligence" is used to describe the development of computer systems that are capable of performing tasks that are typically associated with human intelligence. Such tasks include decision-making, problem-solving, language comprehension and visual perception. The significance of artificial intelligence (AI) lies in its capacity to process and analyse datasets employing mathematical and statistical techniques, thereby enabling more efficient and precise decision-making processes that can surpass human capabilities in both speed and complexity (LeCun et al., 2015). Since its formal introduction at the Dartmouth Conference in 1956, where John McCarthy defined it as "the science and engineering of making intelligent machines" (McCarthy et al., 1995), AI has evolved from a theoretical concept into a transformative technology. This evolution has been driven by notable advancements in algorithms, computational power, and the exponential growth of data availability.

Initially, AI systems were primarily based on symbolic logic and rule-based methodologies, necessitating explicitly programmed instructions and predefined representations of knowledge (Shortliffe, 1977). Although these systems were effective in constrained and predictable environments, they lacked the necessary adaptability and scalability to cope with the complexities of real-world applications. This limitation gave rise to the advent of machine learning (ML), which introduced data-driven algorithms capable of identifying patterns from examples without the necessity for explicit programming (Mitchell, 1997). ML techniques, including support vector machines and random forests, provided more flexible and robust models for solving diverse and complex problems (Cortes et al., 1995; Breiman, 2001). The subsequent development of deep learning (DL) constituted a paradigm shift within the field of machine learning. The utilisation of artificial neural networks with multiple layers enabled deep learning systems to automatically learn hierarchical features directly from raw data, thereby demonstrating excellence in unstructured domains such as image recognition, speech processing, and natural language understanding (LeCun et al., 2015; Sarker, 2021). These developments have led to significant breakthroughs in fields such as healthcare diagnostics, autonomous vehicles, and intelligent virtual assistants, demonstrating the potential of AI to address both repetitive tasks and intricate problems with remarkable efficiency. Collectively, AI, ML, and DL now represent a leading area of technological innovation, enabling unprecedented progress in automating processes and solving challenges of extraordinary complexity. As these technologies continue to evolve, they are reshaping industries and expanding the boundaries of what intelligent systems can achieve.

In recent years, artificial intelligence (AI) has become a crucial technology, transforming numerous sectors and fields of research, including healthcare, education, finance or manufacturing (Akkem et al., 2023; Jiang et al., 2017; Jordan and Mitchell, 2015). Its integration into agriculture has similarly revolutionised modern farming practices, offering improvements in productivity, sustainability, and food security (Akkem et al., 2023). By combining AI with non-invasive sensing technologies, precision agriculture has advanced significantly, allowing for the optimisation of tasks such as crop monitoring, weed control, irrigation management and disease and pest treatment through automation and data-driven insights (Fuentes et al., 2024; Lee and Tardaguila, 2023). Within precision agriculture, disease detection in crops is a key area of application, where AI systems can process vast amounts of data to identify subtle changes in plant health with both rapidity and accuracy. As an example, machine learning can facilitate the analysis of changes in plant organs generated by structural and biochemical defence mechanisms using non-invasive technologies such as spectral sensors (Mahlein et al., 2019). As AI evolves, its increasing capacity to address specialised agricultural challenges and global issues highlights its pivotal role in shaping a sustainable future.

1.2.1 Machine learning process

Machine learning (ML) is a powerful computational paradigm that enables the automatic extraction of patterns and insights from data, training mathematical models to perform complex tasks based on the relationships inherent in the data. By leveraging data-driven learning mechanisms, ML has become a cornerstone for solving problems in diverse domains, ranging from healthcare and agriculture to finance and autonomous systems. The success of ML, however, depends on several critical factors: the quality of the input data, the architectural design of the model, and its capability to generalise effectively to unseen scenarios, ensuring robust performance in real-world applications (Sarker, 2021).

A reliable ML system should be built on a structured process that transforms raw data into actionable predictions or decisions. As outlined by García et al. (2015), understanding the problem context and defining clear objectives are foundational to initiating the ML pipeline. Then, the ML process involves a series of interdependent steps (Figure 1.4), each playing a vital role in ensuring the accuracy and robustness of the resulting model (García et al., 2015; Paleyes et al., 2023):

1. **Data acquisition** is the first step, where relevant data for solving the problem is collected from various sources, such as sensors or publicly available datasets. The quality and representativeness of the data are critical, as they directly affect model performance. However, challenges such as imbalanced datasets, insufficient coverage, or irrelevant features often arise. For example, in plant pathology applications, like apple leaf disease detection (Li et al., 2021), the imbalance between healthy and diseased samples can hinder accurate model training.
2. **Data preparation**, involves data preparation for model training. This process involves several sub-tasks, including data cleaning, data transformation, and feature engineering. Data cleaning focuses on eliminating errors, inconsistencies, missing values, and noise, thereby improving the quality of the data. Data transformation adjusts raw data into a format that aligns with the requirements of the model, while feature engineering extracts or creates meaningful attributes that enhance the model's learning capability. By addressing these aspects, data preparation significantly boosts the relevance and qual-

ity of the inputs, which is essential for effective pattern recognition. Afterwards, the processed data is typically divided into training and testing datasets, enabling the model to learn from one subset and be evaluated on the other.

3. **Model training** is the phase where the machine learning algorithm learns to recognise patterns and make predictions based on the prepared data. This step involves selecting an appropriate model, tuning its hyperparameters, and iteratively optimising the model to minimise errors and improve accuracy. Depending on the complexity of the data and the task, training can range from straightforward to highly intricate. Often, researchers use pre-trained models and fine-tune them for specific applications to expedite the training process. The primary objective is to develop a model capable of generalising effectively to unseen data, ensuring robust and reliable performance.
4. **Model testing**, the model is evaluated using unseen data to assess its generalisation capabilities and predictive performance. This involves using a separate dataset, known as the test set, to assess how well the model performs on new data. Model testing helps in identifying issues like overfitting and underfitting, and provides insights into the model's accuracy, precision, recall, and other performance metrics. It is crucial to ensure that the test data is representative of real-world scenarios to obtain a reliable evaluation.
5. **Model deployment** is the final step, where the trained model is integrated into a production environment to make predictions on new, real-world data. This involves embedding the model into operational systems, setting up monitoring protocols, and performing regular maintenance to ensure consistent performance over time.

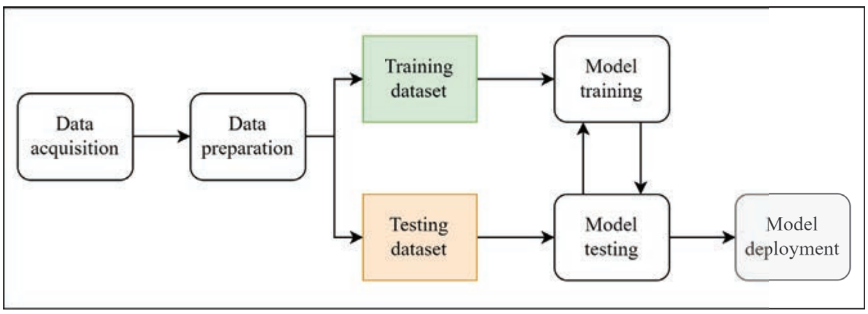


Figure 1.4: Machine learning workflow

1.2.2 Machine learning categories

The ML process can be categorised into four principal types, depending on the data used to evaluate the training process and how the learning process is performed (Sarker, 2021): supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. In supervised learning, models are trained with labelled data, which enables the learning process to concentrate on the anticipated outcome. This approach is particularly suited to classification and regression tasks. In contrast to supervised learning, unsupervised learning operates with unlabelled data, whereby the model identifies hidden patterns or structures. This approach is frequently employed in clustering and dimensional reduction tasks. Semi-supervised learning involves the combination of labelled and unlabelled data, typically utilising a limited amount of labelled data to guide the learning process on a larger set of unlabelled data. This makes it a valuable approach when labelling is costly or time-consuming, such as in fraud detection or text classification. Finally, reinforcement learning focuses on training an agent to make decisions by interacting with an environment, receiving rewards or penalties based on its actions. This makes it particularly well suited to dynamic tasks such as robotics or autonomous driving. Consequently, the complexity of the task to be solved and the characteristics of the data available will determine the most appropriate machine learning category to use.

Supervised machine learning is the most widely used category of ML due to its capability of guiding the learning process with the desired outcomes. This approach is particularly powerful for classification and regression tasks, where the goal is to predict discrete classes or continuous values, respectively, based on input features. Some examples of the models commonly used for these tasks are:

- **Support Vector Machine (SVM)**: it identifies the optimal hyperplane that maximises the margin between two classes, with the support vectors being the data points closest to the hyperplane (Cortes et al., 1995). This capability to maximise the margin makes SVM particularly effective in high-dimensional spaces. For instance, SVM has been effectively used for disease diagnosis, such as cancer classification from gene expression data (Guyon et al., 2002).
- **K-Nearest Neighbors (KNN)**: it classifies data points considering the majority class of their nearest neighbours (Cover and Hart, 1967). The distance between data points is often calculated using Euclidean distance, and the choice of the number of neighbours (k) significantly impacts the performance of the model. Its main advantages are that it is computationally simple and does not require an explicit training

phase. In the context of text classification systems, the algorithm is capable of assigning a category to a given document based on its text similarity with other documents (Bijalwan et al., 2014).

- **Partial Least Squares Discriminant Analysis (PLS-DA)**: it can be thought as a supervised version of Principal Component Analysis (PCA), achieving dimensionality reduction considering the class labels. It models the relationship between predictor variables and class labels by projecting both into a lower-dimensional space, maximising the covariance between them to improve class separation (Barker and Rayens, 2003). PLS-DA is extensively applied in chemometrics, particularly for classifying chemical compounds based on high-dimensional and multicollinear data such as spectral data. For instance, PLS-DA can distinguish between diseased and healthy tomato plants in early stages of the disease by analysing hyperspectral images, even when the number of features far exceeds the number of observations (da Cunha et al., 2023).
- **Multilayer Perceptron (MLP)**: it is an artificial neural network inspired by the structure and function of biological neural networks. It is comprised of multiple layers of interconnected nodes (neurons), linked by weighted edges, and is designed for tasks such as classification and regression (Aggarwal, 2018). An MLP is comprised of three principal components: an input layer that receives data, one or more hidden layers that process and transform input features into higher-level representations, and an output layer that generates predictions or classifications (Figure 1.5). A distinctive feature of MLPs is their utilisation of non-linear activation functions, including ReLU, sigmoid, and tanh, which facilitate the network's capacity to approximate intricate, non-linear mappings between inputs and outputs. This capability renders MLPs suitable for modelling intricate relationships in data across a wide range of applications. Training is conducted using backpropagation (Hinton, 1989), a method wherein the gradient of a loss function, which quantifies the discrepancy between predicted and actual outputs, is propagated backward through the network to iteratively update connection weights. In plant pathology, MLPs have been successfully applied to tasks such as the early detection of bacterial canker in tomatoes by analysing the spectral signatures of plants (Vallejo-Pérez et al., 2021).

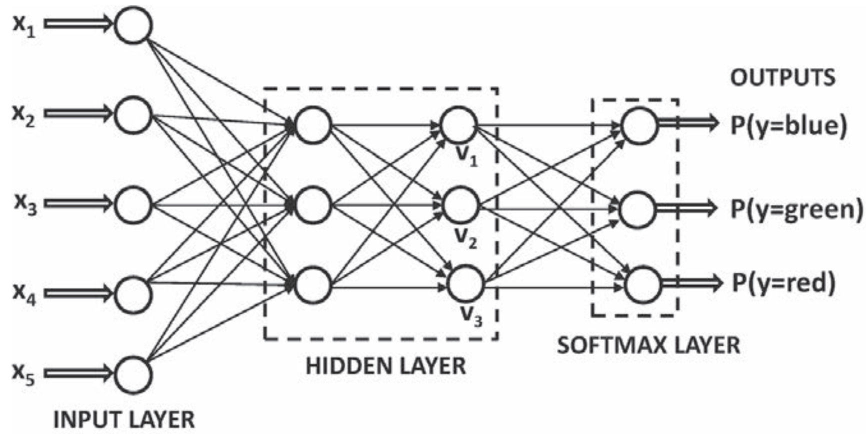


Figure 1.5: Example of an artificial neural network with multiple outputs for categorical classification. Source: Aggarwal (2018)

1.2.3 Deep learning architectures

Deep learning, as a subfield of machine learning, has emerged as a dominant approach for solving complex computational tasks that were previously intractable. Deep learning has demonstrated remarkable advancements in recent years, particularly in tasks that involve high-dimensional data, such as images, video, audio, and text (LeCun et al., 2015). A key strength of DL lies in its capability to automatically learn hierarchical feature representations from raw data, reducing the need for extensive feature engineering, a time-consuming and often labour-intensive pre-processing step in traditional machine learning (ML) pipelines. This reduction in pre-processing efforts optimise the ML workflow and enables more efficient model development.

Several deep learning models have become standard tools in various domains due to their unique architectures and capabilities (Serre, 2019). Convolutional Neural Networks (CNNs) are widely used for tasks involving spatial data, such as image and video processing, by utilising convolutional layers to extract local patterns (Krizhevsky et al., 2012). Recurrent Neural Networks (RNNs) are effective in sequential data tasks such as natural language processing (NLP) and time series prediction, as they can retain information across time steps. Autoencoders are employed for unsupervised learning tasks like dimensionality reduction and anomaly detection, compressing input data into a latent space and reconstructing it. Additionally, Generative Adversarial Networks (GANs), introduced by Goodfellow et al. (2014), consist of two networks (a generator and a discriminator) that are trained simultaneously, producing highly realistic synthetic data and advancing fields like image generation and video synthesis. Finally, Transformer models, particularly those using the attention mechanism, have gained prominence in NLP tasks achieving state-of-the-art results in tasks such as language translation, text summarisation, and question answering (Vaswani et al., 2017).

1.2.4 Challenges of deep learning

Deep learning models often require vast amounts of data to achieve optimal performance. To address these limitations, researchers have adopted techniques such as transfer learning, fine-tuning, and data augmentation, which help improve model generalisation and reduce overfitting:

- **Transfer Learning**: this approach involves utilising neural networks pre-trained on large-scale datasets for a different but related task. Transfer learning allows models to leverage knowledge gained from one domain and apply it to another domain with limited data. A well-known example in image analysis is the utilisation of the ImageNet dataset (Russakovsky et al., 2015) as a basis for training CNNs, due to its general purpose.
- **Fine-Tuning**: it is a specialised technique within transfer learning where a pre-trained model is partially or fully re-trained on a new, smaller dataset to adapt it to a specific task. During fine-tuning, earlier layers of the pre-trained model (which capture general features) are often frozen, while the later layers (which capture task-specific features) are updated. This approach is particularly useful in domains where labelled data is limited, accelerating the training process and reducing overfitting.
- **Data Augmentation**: this method artificially increases the size of the training dataset by applying random transformations to the original data. Shorten and Khoshgoftaar (2019) highlighted how data augmentation improves generalisation in image classification models by exposing them to a broader variety of input data.

These techniques have shown promise in agricultural applications, particularly in field-based disease assessment. Deep learning has proven effective in tasks such as identifying cassava leaf disease (Thai et al., 2021) and differentiating grapevine diseases (Gutiérrez et al., 2021). Deep learning algorithms are capable of adapting to diverse environmental conditions, enabling the identification of complex disease symptoms even in challenging scenarios. However, the practical deployment of these models requires the collection and labelling of substantial datasets, a labour-intensive process. To overcome this, data augmentation, transfer learning and fine-tuning are often employed. For example, Li et al. (2021) demonstrated improved accuracy in detecting apple leaf diseases using these techniques in limited datasets. As these methods evolve, detection systems will become more efficient, providing farmers with actionable insights that enable effective disease management and optimized resource use.

Additionally, while deep learning (DL) models have demonstrated remarkable success across various domains, they often suffer from a lack of interpretability, commonly referred to as the "black box" problem. Unlike traditional machine learning models, which can provide insights into feature importance and decision-making processes, DL models are generally opaque, making it difficult to understand the reasoning behind their predictions. To address this issue, researchers have developed a range of Explainable Artificial Intelligence (XAI) techniques aimed at increasing the transparency and accountability of DL models (Arrieta et al., 2020). One widely used XAI method is Grad-CAM (Gradient-weighted Class Activation Mapping), which generates class-specific heatmaps to highlight the regions of an image that most influence a model's predictions by computing the gradient of the class output with respect to feature maps in the final convolutional layer (Selvaraju et al., 2020). Similarly, attention maps, commonly employed in models handling sequential data, such as in natural language processing and machine translation, provide visual representations of where the model focuses at each time step, offering insights into its decision-making process. In image classification tasks, attention maps can also highlight important regions that the model attends to when making predictions, thereby enhancing interpretability (Dosovitskiy et al., 2020). In conclusion, deep learning has revolutionised machine learning by automating feature extraction and enabling the development of highly accurate models for complex tasks. By addressing challenges such as data scarcity and interpretability, DL has the potential to transform the automation of solving complex real-world problems.

1.3 Computer vision

Computer vision, a subfield of artificial intelligence, focuses on developing techniques that allow machines to process and interpret visual information, such as images and videos, with the goal of emulating human visual comprehension. By leveraging advanced algorithms and models, this technology enhances the automation of tasks that depend on visual analysis, resulting in improvements in efficiency, accuracy, and scalability across various domains (Szeliski, 2022). For instance, in healthcare, computer vision is pivotal in medical image analysis and aiding diagnostics (Litjens et al., 2017), while in the automation of robots, it may enable real-time image processing for navigation and obstacle avoidance (Ball et al., 2016). Similarly, in agriculture, it helps to optimise crop management through weed or disease detection, yield estimation or soil analysis (Rehman et al., 2019). The expanding application of computer vision in these and other sectors underscores its crucial role in driving technological innovation.

Considering the variety of possible applications of computer vision, this technology can cover a range of tasks, each designed to address a specific challenge in image analysis. One of the most common tasks is image classification, which involves assigning predefined labels to entire images based on their content, such as determining whether an image contains a specific object or scene. Regression tasks, similarly, involve the prediction of continuous values associated with images or objects. Object detection extends classification by localising objects within images, identifying their bounding boxes, and assigning class labels. Image segmentation, on the other hand, focuses on partitioning an image into meaningful regions or objects to achieve a more granular understanding of the scene. Segmentation can be categorised into semantic segmentation, where each pixel in an image is classified into a category with consistent labelling across the image, and instance segmentation, which not only classifies pixels but also distinguishes between different instances of the same object type, making it essential for tasks requiring recognition of multiple occurrences of similar objects.

The transition from classical computer vision techniques to those based on deep learning represents a revolutionary change in the field, driven by advances in computational power, algorithmic innovation, and the accessibility of large annotated datasets (O'Mahony et al., 2020). Classic methods depend on manually designed features and algorithms for the analysis of visual data. Although these techniques are effective for specific, well-structured tasks, they are inherently limited in their capability to handle variability caused by factors such as noise, scale, or changes in illumination. Deep learning, particularly through convolutional neural networks (CNNs), addresses these challenges by automating feature extraction and enabling end-to-end learning (Figure 1.6). This allows models to generalise across diverse and complex datasets without the need for task-specific feature engineering (Voulodimos et al., 2018). Rather than supplanting classical methods, deep learning can be integrated with them to create hybrid systems that combine the interpretability and robustness of classical techniques with the adaptability and scalability of neural networks. Such hybrid approaches have been demonstrated to be particularly effective in domains where high reliability and constrained resources are required. For example, Gutiérrez et al. (2021) showed that combining traditional computer vision techniques for the extraction of disease symptoms from grapevine leaves with CNNs for leaf classification resulted in superior outcomes compared to using raw images. As observed by O'Mahony et al. (2020), this complementary relationship offers a promising direction for advancing computer vision applications by leveraging the unique strengths of both paradigms.

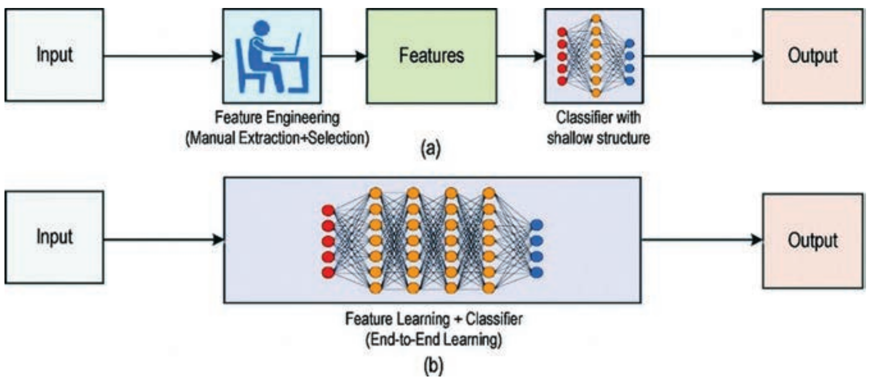


Figure 1.6: Workflows of classic computer vision (a) vs. deep learning (b). Source: O'Mahony et al. (2020)

In agriculture, the different computer vision tasks are pivotal for automating the identification of disease symptoms based on visual patterns such as leaf spots, discolourations, and necrosis. Studies have demonstrated the efficacy in detecting diseases like canker, melanose, sunscald, anthracnose and greening in citrus fruits (Zhang et al., 2022); coffee leaf miner, soybean rust or wheat tan spot (Gonçalves et al., 2021); as well as downy mildew and spider mite in grapevine (Gutiérrez et al., 2021). Due to the specific characteristics of each agricultural problem, deep learning techniques demonstrated their potential to address distinct computer vision tasks. This included image classification using classic computer vision techniques to prepare the data and convolutional neural networks (CNNs) to differentiate grapevine diseases, object detection using a combination of YOLO-v4 and a CNN to localise and classify the fruits, and image segmentation using deep learning models for lesion localisation in leaves affected by diseases or pests (Figure 1.7).

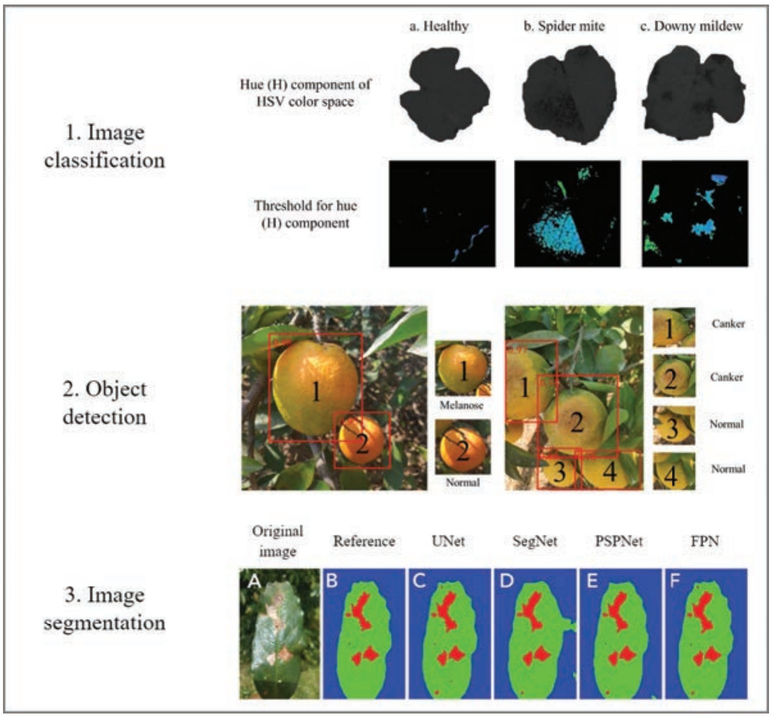


Figure 1.7: Examples of the use of deep learning to solve computer vision tasks in disease detection in agriculture: 1) CNN for disease and pest differentiation in grapevine (Source: Gutiérrez et al. (2021)), 2) YOLO for disease detection in citric fruits (Source: Zhang et al. (2022)), 3) UNet, SegNet, PSPNet and FPN for leaf segmentation (Source: Gonçalves et al. (2021)).

1.3.1 Image preparation with classic methods

Early developments in computer vision relied heavily on manual engineering techniques to pre-process images and extract meaningful features, thereby guiding the subsequent analysis of the data in order to obtain the desired results. A fundamental technique is colour space conversion, which involves transforming the image from RGB (Red, Green, Blue) to colour spaces such as HSL (Hue, Saturation, Lightning), HSV (Hue, Saturation, Value), LAB or grayscale, helping to highlight specific features, such as intensity or chromatic components, facilitating the extraction of relevant information. Contrast Limited Adaptive Histogram Equalization (CLAHE) is frequently applied to improve the local contrast in images with uneven lighting. Kumar and Jindal (2019) remarked the possibility of improving the analysis of foggy images using HSV colour space and CLAHE method. Morphological transformations, such as erosion, dilation, opening and closing, are crucial for refining object boundaries, removing noise, and enhancing specific structures. Noise reduction may also be done using filters such as median filter or blurring, improving the quality of the images, allowing for clearer identification of objects or individuals without compromising the edge details. For instance, Íñiguez et al. (2021) used HSV and RGB colour spaces to segment grapevine components, applying morphological transformations like erosion and dilation to enhance the segmentation quality, particularly in analysing the impact of leaf occlusion on yield assessment. Similarly, Rodríguez et al. (2020) showcased the utility of these techniques in detecting cherry beans on coffee trees. Their method involved converting the colour space, applying a median blur filter to reduce noise, equalizing histograms to enhance contrast, and using morphological transformations to highlight features critical to cherry bean detection. Collectively, these manual engineering techniques underline the importance of pre-processing in traditional computer vision, enabling the effective extraction and enhancement of image features for diverse applications.

1.3.2 Information extraction with classic methods

In addition to image enhancement techniques that facilitate feature extraction for tasks like image classification, classic computer vision methods can also be applied to higher-level tasks such as object detection and image segmentation. The Hough transform is widely used for detecting geometric shapes such as lines, ellipses or circles. Diago et al. (2015) applied this technique to estimate yield components in grapevine, detecting and analysing berries of different varieties represented as circles. Another robust technique is the watershed algorithm, which segments images based on topographical surface analysis by treating pixel intensities as elevations, making it effective in biomedical applications such as cell segmentation in microscopy (Gamarra et al., 2019). While the watershed algorithm excels at separating overlapping objects, it is sensitive to noise and often requires preprocessing steps to enhance accuracy, similar to the Hough transform.

Another common technique for separating regions in an image is thresholding. Crisp thresholding uses a fixed intensity value to segment pixels. It can be applied in task like the separation of cherry bean pixels from the background in Rodríguez et al. (2020). The threshold value is often manually selected, but the Otsu method allows the automation of the calculation of the optimum threshold by minimising the intra-class variance of pixel intensity (Goh et al., 2018). For more complex images with gradual transitions or noise, fuzzy thresholding offers a more flexible approach. Fuzzy logic is a mathematical method designed to handle the concept of partial truth, where values are not limited to strict binary classifications. This flexibility allows fuzzy logic to model uncertainty, making it highly effective for solving complex problems where precise boundaries or classifications are difficult to define. For example, fuzzy logic has been employed in applications such as wine quality classification, where human interpretation plays a significant role (Petropoulos et al., 2017). In the context of image segmentation, fuzzy logic addresses challenges such as ambiguity and gradual transitions by assigning degrees of membership to pixels rather than rigidly categorising them. This results in smoother transitions between regions and more accurate segmentation of images with complex patterns. An illustration of this was the use of fuzzy c-means clustering for detecting diseased regions in the leaves of cucumber and pumpkin species (Sekulska-Nalewajko and Goclawski, 2011), and a fuzzy approach for estimating disease severity in grapevine leaves (Nagi and Tripathy, 2021).

Despite their limitations, classic methods like Hough transform, watershed, and thresholding, remain widely used due to their computational efficiency and ease of implementation. These methods are particularly valuable in scenarios with controlled conditions, such as laboratory experiments.

1.3.3 Machine learning in computer vision

Computer vision techniques have established a robust foundation for the preparation of data for the application of machine learning models, which has in turn led to the development of more sophisticated and efficient systems for the solution of complex visual tasks (Smith et al., 2021). For example, Mwebaze and Owomugisha (2016) illustrated how the integration of computer vision and machine learning can be utilised to assess the severity of cassava leaf diseases. In this study, image features were classified using machine learning algorithms, including the Linear Support Vector Classifier, KNN, and Extremely Randomized Trees. Similarly, Mukherjee (2020) used Gray-Level Co-Ocurrence Matrix (GLCM) to extract disease symptoms in potato leaf images and classified the images into two diseases using SVM.

Furthermore, the rise of deep learning has significantly transformed the field of computer vision, automating the feature extraction process and thereby facilitating substantial enhancements in accuracy and scalability (Voulodimos et al., 2018). Deep learning models are capable of learning hierarchical representations of data, which has resulted in notable advancements in tasks such as classification, segmentation, and object detection. As an example, deep learning has demonstrated its efficacy in optimising feature engineering in agricultural fields, generalising simple problems such as fruit counting, or developing robust models that consider challenging conditions such as illumination, background noise or different image resolutions (Kamilaris and Prenafeta-Boldú, 2018).

1.3.4 Image classification with deep learning

The evolution of deep learning in image classification has been marked by pivotal milestones that transformed computer vision by enabling methods to automatically extract meaningful features from visual data and classify them based on their content. One of the earliest breakthroughs was the introduction of convolutional neural networks (CNNs), with LeNet by LeCun et al. (1995) demonstrating their potential for tasks like digit recognition. A paradigm shift occurred with the advent of AlexNet (Krizhevsky et al., 2012), which leveraged GPUs and ReLU activation functions to achieve unprecedented accuracy on the ImageNet dataset, setting a new standard for performance. Subsequent architectures such as VGGNet (Simonyan and Zisserman, 2015), with its deep yet straightforward stacked convolutional layers, and GoogLeNet (Szegedy et al., 2015), featuring Inception modules for improved computational efficiency, further refined CNN capabilities. The introduction of ResNet (He et al., 2016) addressed the vanishing gradient problem by incorporating skip connections, enabling the training of much deeper networks. More recently, Vision Transformers (ViTs) have revolutionised the domain by adopting attention mechanisms (Dosovitskiy et al., 2020), previously dominant in natural language processing (Vaswani et al., 2017), to model global dependencies in images. In addition, as deep learning models grow in complexity and computational demand, the development of lightweight architectures has emerged as a critical area of research, particularly for deployment on resource-constrained devices such as smartphones, drones, and IoT (Internet of Things) systems. Models such as MobileNet (Howard et al., 2019), which introduced depthwise separable convolutions to reduce the number of parameters and computations, and EfficientNet (Tan and Le, 2021), which optimally balances model depth, width, and resolution through compound scaling, exemplify this trend. Similarly, lightweight adaptations of ViTs, such as MobileViT (Mehta and Rastegari, 2021), integrate attention mechanisms into compact architectures for low-resource environments. These lightweight models achieve competitive accuracy while significantly reducing memory and processing requirements, making them indispensable for real-time applications and scenarios requiring edge computing. The design of such models underscores the increasing emphasis on sustainability and accessibility in the deployment of deep learning systems across diverse environments. Together with advances in large-scale datasets, optimisation techniques, and hardware acceleration, these innovations have firmly established deep learning as the cornerstone of modern image classification.

Convolutional Neural Networks (CNNs) are deep learning architectures biologically inspired by the working of the cat's visual cortex, in which specific portions of the visual field seemed to excite particular neurons

(Aggarwal, 2018). This inspiration drives the design of convolutional neural networks to efficiently process and analyse the spatial structure of images by leveraging pixel-level spatial relationships (Figure 1.8). The convolution operation, which applies learnable filters to localised regions of an image, constitutes the fundamental operation of a convolutional neural network. This mechanism enables convolutional neural networks to extract both low-level and high-level features. Low-level features, such as edges and corners, are extracted in the initial layers of the network, while more complex patterns, such as shapes and textures, are extracted in the deeper layers. A significant advantage of convolutional neural networks is their capacity to discern local dependencies and spatial hierarchies within images. The translational equivariance of the convolution operation guarantees that the features learned by the network remain robust to positional changes, thereby enhancing the network's capacity for generalisation across tasks. To further refine the representation of features, pooling layers (such as max pooling or average pooling) reduce the spatial dimensions of feature maps, focusing on the most relevant features while reducing the computational overhead and mitigating overfitting by introducing spatial abstraction. In order to enhance the efficiency and performance of the training process, convolutional neural networks (CNNs) incorporate additional techniques, such as batch normalisation and dropout. Batch normalisation serves to reduce internal covariate shift by standardising the inputs to each layer, thereby accelerating convergence during training. Dropout, a regularisation method, randomly disables a fraction of neurons during training, encouraging the network to learn robust and distributed representations and reducing the risk of overfitting. These enhancements contribute to the generalisability and efficiency of CNNs across diverse datasets and applications. For instance, CNNs have revolutionised image classification, with applications such as medical diagnosis, where they can be used to detect diseases in images using transfer learning from non-medical images (Shin et al., 2016).

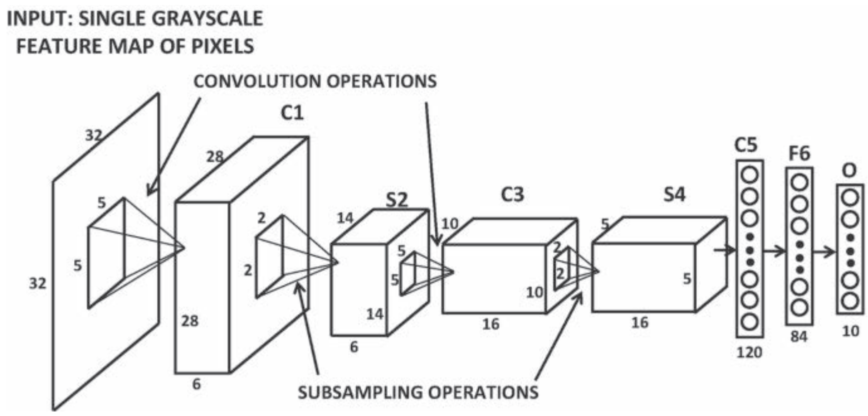


Figure 1.8: Convolutional Neural Network architecture. Source: Aggarwal (2018)

On the other hand, Vision Transformers (ViTs) have introduced a transformative approach to image classification by relying on self-attention mechanisms instead of convolutions (Dosovitskiy et al., 2020). In ViTs, images are divided into patches, which are treated as individual tokens in a manner similar to that employed in a Natural Language Processing (NLP) application (Figure 1.9). The self-attention mechanism enables the model to learn relationships between these tokens, capturing both local details and global dependencies. This allows ViTs to excel in tasks requiring a comprehensive understanding of image context, often outperforming CNNs on large datasets. However, ViTs typically have higher computational requirements and rely on extensive pretraining due to the absence of inductive biases like locality and translational equivariance, which are inherent in CNNs.

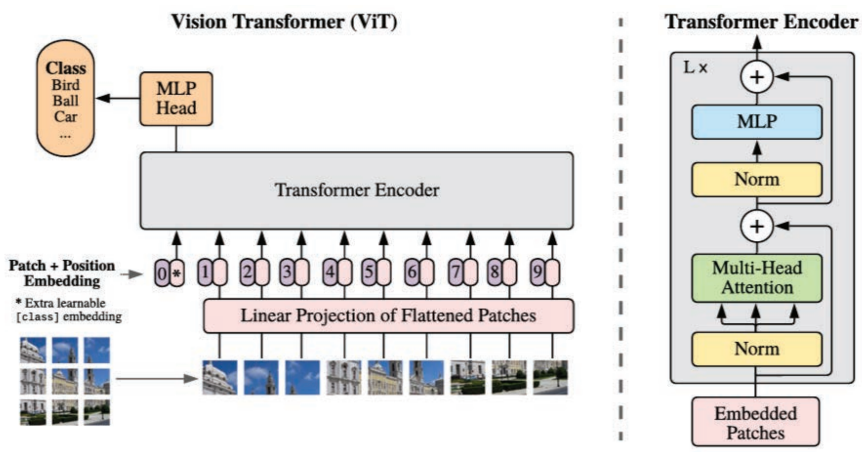


Figure 1.9: Vision Transformer architecture. Source: Dosovitskiy et al. (2020)

In general, CNNs and ViTs have demonstrated considerable potential in agricultural applications. CNNs have been widely applied to tasks such as crop yield prediction and disease detection (Akkem et al., 2023). For example, Sharma et al. (2021) implemented a CNN to classify the severity of mustard downy mildew. Similarly, Thai et al. (2021) highlighted the advances of vision transformers over CNNs for classifying cassava leaf diseases. However, ViTs often face challenges with limited data sets. To address this, Zhou et al. (2023) developed a residual distillation transformer for rice leaf disease identification that outperformed CNNs. In addition, some studies have explored innovative integrations for the analysis of high-resolution images. For example, (Li et al., 2019) combined CNNs with a multi-scale sliding window method for accurate oil palm detection and geolocation in satellite images. These studies demonstrated the potential of deep learning for image classification, providing scalable solutions to agricultural challenges.

1.3.5 Image segmentation with deep learning

The rise of deep learning, particularly convolutional neural networks (CNNs), has revolutionised segmentation by enabling highly accurate, automated approaches to process large and diverse datasets. Over the years, several specialised architectures have emerged, each addressing specific challenges such as spatial information preservation, multi-scale feature integration, and computational efficiency.

- **FPN (Feature Pyramid Network)** takes a top-down approach, integrating high-level semantic features from deeper layers with spatially detailed features from shallower layers via lateral connections. This architecture excels at multi-scale feature representation (Krizhevsky et al., 2012).
- **UNet** builds upon this concept with a symmetric U-shaped structure and skip connections that directly link encoder and decoder layers. These connections allow fine-grained details to propagate across the network, making UNet especially effective in domains like biomedical imaging, where high precision is required. U-Net has demonstrated particular efficacy in biomedical imaging, with its capability to accurately segment neuronal structures or cells (Ronneberger et al., 2015).
- **PSPNet (Pyramid Scene Parsing Network)** introduces a pyramid pooling module that extracts contextual information from multiple receptive fields (Zhao et al., 2016). By fusing global and local features, PSPNet achieves robust segmentation even in images with scale variation and complex scenes.

- **SegNet**, one of the earlier deep learning-based segmentation models, employs an encoder-decoder architecture. The encoder compresses the spatial dimensions of the input image, while the decoder restores them using unpooling layers that leverage max-pooling indices from the encoder. This design reduces computational complexity while maintaining spatial detail. SegNet has shown potential in autonomous driving, which its capacity to segment objects such as pedestrians and vehicles in road scenes (Badrinarayanan et al., 2017). Additionally, it has been employed in agricultural applications, including the detection of the number of flowers in grapevines (Palacios et al., 2020).
- **DeepLab**, with its multiple iterations (DeepLabv1-v3+), enhances segmentation accuracy using atrous (dilated) convolutions, which expand the receptive field without increasing the number of parameters (Chen et al., 2017). Some versions integrated Conditional Random Fields (CRFs) to refine segment boundaries by enforcing consistency along edges. Additionally, DeepLabv2 included atrous spatial pyramid pooling (ASPP) for capturing multi-scale information.
- **MANet (Multi-Attention Network)** leverages attention mechanisms to focus on salient features while suppressing irrelevant information. By integrating spatial and channel attention modules, MANet improves feature representation, making it suitable for complex scenarios such as the segmentation of images containing near-infrared and RGB data captured by different satellite sensors (Li et al., 2022).

Deep learning-based semantic segmentation represents a transformative tool in the field of agriculture, facilitating precise identification and localisation of crop features for a range of applications. Casado-García et al. (2022) demonstrated the efficacy of deep learning for plant segmentation, showcasing the effectiveness of DeepLabV3+ and MANet architectures in accurately analysing grapevine canopy components, including bunches, leaves, and wood. Tong et al. (2021) employed the PSSNet architecture, which was based on an encoder-decoder network, for the detection and counting of trees. Furthermore, disease evaluation might be conducted using deep semantic segmentation, as demonstrated in the study by Li et al. (2023), where a Multi-fusion U-Net was employed for the segmentation of grapevine leaf images captured using an unmanned aerial vehicle (UAV). Similarly, Gonçalves et al. (2021) further expanded the application scope by comparing six convolutional neural network (CNN) architectures to segment disease-affected areas in crops like coffee, soybean, and wheat. This work demonstrated the efficacy of architectures like DeepLabV3+ in capturing multi-scale features, while simpler models like U-Net and SegNet proved effective in less complex scenarios. In addition, Gao et al. (2021) developed a method for segmenting field images of potato crops, localising late bight symptoms using a SegNet-based architecture. These developments illustrate the capacity of deep learning to automate labour-intensive procedures, facilitate precision agriculture, and promote sustainable crop management.

1.3.6 Challenges in image analysis with deep learning

Despite the advances made by deep learning, one of the primary challenges in computer vision remains the necessity for a substantial quantity of quality labelled data to effectively train these models and facilitate the acquisition of intricate visual patterns. The time and effort required to obtain these annotations varies depending on the task, with the complexity of the labels influencing the level of work needed. This can range from classification to segmentation, ranging from the classification of a whole image to the classification of each of its pixels. This issue is particularly evident in specialised domains such as medical imaging, where expert knowledge is essential to guarantee accurate labelling (Litjens et al., 2017). In order to overcome the limitations posed by scarce data and the intricacy of labelling, techniques such as transfer learning, data augmentation and test-time augmentation (TTA) may be utilised. Data augmentation entails the generation of supplementary training examples through the application of arbitrary transformations, including rotations, scaling, flipping, cropping, blurring, shifting or contrast and brightness changes, to the images of the original dataset. This approach expands the size of the dataset without necessitating additional labelled examples (Shorten and Khoshgoftaar, 2019). Additionally, more sophisticated data augmentation techniques may be applied to enhance the complexity of the dataset. This could entail combining images and labels with MixUp (Ethiraj and Bolla, 2022) or generating synthetic data with Generative Adversarial Networks (GANs) (Gutiérrez and Tardaguila, 2023). Moreover, TTA can be applied during the inference phase to improve the robustness of the predictions. It applies several simple transformations to the input image at test time, running the model on each transformed version and averaging the predictions. This technique reduces the sensitivity of the model to variations in the input data, such as changes in orientation or illumination, and thus helps to improve generalisation (Ahamed et al., 2023).

Chapter 2

Objectives

The **main objective** of this PhD thesis was to develop new artificial intelligence and computer vision-based methods for the early assessment of downy mildew in grapevine using non-invasive sensing technologies under laboratory and field conditions.

The **specific objectives** of this research work were:

- To explore the application of artificial intelligence and non-invasive technologies for assessing downy mildew in grapevine using RGB and hyperspectral images under laboratory conditions.
- To develop and validate a new method to automatically assess the severity of downy mildew disease in grapevine by combining fuzzy logic and computer vision techniques under laboratory conditions
- To use convolutional neural networks and explainable artificial intelligence for early detection of downy mildew in grapevine under laboratory conditions.
- In-field downy mildew detection and localisation in grapevine using explainable deep learning using RGB images.
- To employ deep semantic segmentation for the severity assessment of downy mildew in grapevine using RGB images taken in commercial vineyards.

Chapter 3

Artificial intelligence and non-invasive sensing technologies for downy mildew evaluation

The assessment of grapevine downy mildew is essential for advancing effective treatment development and mitigating disease spread. Traditional methods rely on visual assessment by trained experts and are time-consuming. The integration of artificial intelligence and non-invasive sensing technologies can facilitate the rapid and accurate assessment of plant health without damaging the plants. Hyperspectral imaging (HSI) captures spectral data that can assist in identifying biochemical alterations in plants, whereas RGB imaging provides visual information that can facilitate the preliminary assessment of symptoms and monitoring. The present study explored the use of artificial intelligence and non-invasive sensing technologies for assessing grapevine downy mildew under laboratory conditions. Two innovative approaches were employed for assessing grapevine leaf discs infected with downy mildew. The first method utilised spectral processing, classic computer vision and machine learning to identify downy mildew with HSI. The second method employed classic computer vision to localise downy mildew symptoms in RGB images. The results demonstrated the potential of artificial intelligence and non-invasive technologies to automate and optimise downy mildew assessment in grapevine. HSI proved effective for early disease detection, while RGB imaging facilitated severity assessment. The combination of these methodologies offers a promising framework for accelerating the development and evaluation of treatments and advancing the study of plant-pathogen interactions.

Paper information

- **Title of the publication**: Artificial intelligence and novel sensing technologies for assessing downy mildew in grapevine
- **Authors**: Inés Hernández, Salvador Gutiérrez, Sara Ceballos, Rubén Iñíguez, Ignacio Barrio and Javier Tardaguila
- **Published in**: Horticulturae
- **DOI**: 10.3390/horticulturae7050103

Contributions of the PhD thesis' author: Inés Hernández Casado participated in preparing and acquiring RGB and hyperspectral images of grapevine leaf discs for the analysis of downy mildew disease. She designed and implemented the algorithms applying machine learning and computer vision techniques for severity estimation using RGB images and early detection using hyperspectral images. She was also in charge of the validation and analysis of the results. Finally, she wrote the first draft of the article and participated in the review process.

Chapter 4

Fuzzy logic and computer vision for the evaluation of downy mildew severity

The potential of classic computer vision to estimate grapevine downy mildew severity using RGB images taken in the laboratory was demonstrated in Chapter 3. These techniques could facilitate the development of disease treatments or the analysis of plant-pathogen interactions through the analysis of RGB images of plant infections. However, traditional visual evaluation of symptoms made by experts relies on their capability to discriminate and rate disease symptoms, taking into account the intensity and distribution of the symptoms in the leaves. On this basis, the following work developed and validated a computer vision-based method combined with fuzzy logic to automate disease severity assessment in the laboratory. This method employed classic computer vision to localise downy mildew symptoms in grapevine leaf discs. A fuzzy threshold was used for rating symptoms based on their intensity and estimating the disease severity. To evaluate its effectiveness, this approach was compared to a traditional method that used crisp thresholding for disease evaluation, as developed in Chapter 3. The robustness of the method was evaluated across two grapevine varieties, thereby ensuring its adaptability and reliability in diverse conditions. The proposed approach offered a precise, objective and rapid solution for assessing grapevine downy mildew severity, optimising the monitoring of downy mildew infection under laboratory conditions. The reliability of the method was increased by offering the visualization of symptom intensities on leaf discs. Additionally, the capability of the method to extract valuable insights from limited datasets provides a key advantage, supporting its application in adapting to new conditions or diseases.

Paper information

- **Title of the publication**: Assessment of downy mildew in grapevine using computer vision and fuzzy logic. Development and validation of a new method
- **Authors**: Inés Hernández, Salvador Gutiérrez, Sara Ceballos, Fernando Palacios, Silvia L. Toffolatti, Giuliana Maddalena, María P. Diago and Javier Tardaguila
- **Published in**: Oeno One
- **DOI**: 10.20870/oeno-one.2022.56.3.5359

Contributions of the PhD thesis' author: Inés Hernández Casado collaborated in the creation of the two datasets of RGB images of grapevine leaf discs infected with Plasmopara viticola fungus. The PhD student participated in the acquisition of the dataset taken in the Institute of Grapevine and Wine Sciences (ICVV) and coordinated experts in severity assessment. She designed and implemented an algorithm to estimate the disease severity using fuzzy logic and classic computer vision techniques, and analysed the resulting outcomes. She also wrote the first draft of the manuscript and participated in all the review stages.

Chapter 5

Deep learning for early detection of downy mildew

The timely identification of plant diseases at their early stages is of paramount importance for the effective implementation of intervention strategies and for gaining a comprehensive understanding of the progression of the disease. The potential of HSI for early detection was demonstrated in Chapter 3, while RGB sensors were highlighted as cost-effective. Furthermore, classic computer vision techniques have demonstrated in Chapters 3 and 4 their effectiveness in localising symptoms in RGB images under laboratory conditions. The present study aimed to use explainable deep learning to facilitate early detection of grapevine downy mildew and to classify infection stages under laboratory conditions. The study compared the use of different computer vision techniques for feature extraction to simplify disease detection in grapevine leaf discs. Disease symptoms were localised with fuzzy and crisp thresholds (in the same way as in Chapter 4), the colour space was transformed to HSV and the original RGB images were used. The detection of downy mildew in the discs was performed using convolutional neural networks (CNN). Grad-CAM was used to interpret model predictions. Finally, CNNs were used to identify disease stages in the images of symptomatic leaf discs. The findings demonstrated that the integration of deep learning with XAI can facilitate objective, accurate and rapid plant disease monitoring with cost-effective methods that can be readily interpreted and utilised by farmers or researchers. The work demonstrated the efficacy of CNNs in the early detection of disease symptoms using raw data, thereby reducing the time required for the development of specific image processing techniques and facilitating adaptation to new crops or diseases. Furthermore, CNNs enabled the accurate categorisation of infection stages into early, intermediate and high.

Paper information

- **Title of the publication**: Image analysis with deep learning for early detection of downy mildew in grapevine
- **Authors**: Inés Hernández, Salvador Gutiérrez and Javier Tardaguila
- **Published in**: Scientia Horticulturae
- **DOI**: 10.1016/j.scienta.2024.113155

Contributions of the PhD thesis' author: Inés Hernández Casado contributed to the acquisition of RGB images of grapevine leaf discs and the development of an algorithm for monitoring downy mildew infection using deep learning and exploring the use of explainable artificial intelligence techniques. She also led the analysis and discussion of the results, writing the first draft of the manuscript and leading the revision process.

Chapter 6

In-field downy mildew detection using explainable deep learning

The detection of plant diseases in the field is a complex process influenced by natural conditions, such as daylight or plant damage. The detection of small downy mildew symptoms in grapevine could allow for timely and targeted interventions, which can mitigate the spread of the disease and prevent significant crop losses. Traditional detection is conducted visually by trained personnel in the field, which is a time-consuming process when evaluating large areas of crops. As demonstrated in the previous chapters, computer vision and deep learning might detect downy mildew in RGB images. The objective of this study was to develop an automated and interpretable system for the detection and localisation of downy mildew under field conditions. This work presented a novel approach that integrates deep learning models with a sliding window method for the evaluation of high-resolution RGB images of the grapevine canopy. The collection of images was conducted in 14 plots, manually and using a mobile platform, with consideration given to the impact of different lighting and grapevine variety. Convolutional neural networks and vision transformer models were used for the identification of symptomatic regions on the plant. Transfer learning, fine-tuning and data augmentation were employed to mitigate overfitting and develop a robust method. Furthermore, Grad-CAM and attention maps were employed to interpret the neural network predictions. The method demonstrated the capacity of deep learning to detect grapevine downy mildew under challenging field conditions and its potential for extensive crop assessment using ground vehicles such as tractors, thereby paving the way for more sustainable and efficient farming practices. The adaptability of the method to new conditions, crops, or diseases was a key advantage, achieved through straightforward labelling by classifying plant areas. Furthermore, the integration of explainable artificial intelligence offered an understandable approach for agricultural professionals, facilitating the validation of the decisions made by the AI system.

Paper information

- **Title of the publication**: In-field disease symptom detection and localisation using explainable deep learning: Use case for downy mildew in grapevine
- **Authors**: Inés Hernández, Ignacio Barrio, Rubén Íñiguez, Salvador Gutiérrez and Javier Tardáguila
- **Published in**: Computers and Electronics in Agriculture
- **DOI**: 10.1016/j.compag.2024.109478

Contributions of the PhD thesis' author: Inés Hernández Casado collaborated in the image collection in different vineyards in the North of Spain and led their preparation and labelling for downy mildew detection considering the complexity of the field conditions. She designed and implemented an interpretable and robust algorithm for the detection of symptomatic regions in plants using deep neural networks and explainability methods. She also played a significant role in writing and reviewing the manuscript.

Chapter 7

Deep semantic segmentation for severity estimation of downy mildew under field conditions

The precise localisation of downy mildew symptoms in the vineyard enables detailed monitoring, aiding in the effective management of the disease, and reducing the use of chemical treatments or helping to test new treatments. Traditionally, the evaluation of the disease severity is carried out visually by trained experts. However, as demonstrated in Chapter 6, deep learning offers a powerful alternative by extracting infection-related information from field-captured images. The aim of this work was to assess the severity of downy mildew in grapevine under field conditions by employing deep semantic segmentation for localising visual symptoms. State-of-the-art semantic segmentation architectures were utilised to detect the symptoms in the grapevine canopy. The data collected in Chapter 6 was utilised, considering the variability present in the field. The study explored different approaches to address data imbalance caused by the small size of the symptoms, including the use of simple data augmentation, oversampling, undersampling, and the MixUp method. Furthermore, Test-Time Augmentation (TTA) was employed to make the results robust to brightness changes, minimising the occurrence of false positive values. This method presented a robust and objective solution for the assessment of downy mildew severity in the vineyard. The models were developed for real-time application, using efficient neural networks to facilitate their deployment on mobile platforms for disease monitoring in the field. Furthermore, the strategies to address data imbalance further enhanced sensitivity to minor and early symptoms, thus marking a significant advance in non-invasive crop health assessment.

Paper information

- **Title of the publication**: Early detection of downy mildew in vineyards using deep neural networks for semantic segmentation
- **Authors**: Inés Hernández, Rui Silva, Pedro Melo-Pinto, Salvador Gutiérrez and Javier Tardáguila
- **Submitted in**: Biosystems Engineering (under review)

Contributions of the PhD thesis' author: During her internship at the Centre for the Research and Technology of Agro-Environmental and Biological Sciences, in the Universidade de Trás-os-Montes e Alto Douro, Inés Hernández Casado contributed to the development of an algorithm for the segmentation of grapevine canopy images localising downy mildew symptoms and addressing the issue of imbalance due to the small size of the symptoms. In addition, she was responsible for writing the initial draft of the manuscript and its revision.

Chapter 8

Conclusions

Main conclusion: This PhD thesis demonstrates the potential of artificial intelligence, particularly machine learning, deep learning, and computer vision, combined with proximal sensing technologies, to enable non-invasive, rapid, objective, and accurate detection of early symptoms of grapevine downy mildew under both laboratory and field conditions.

The specific conclusions of this PhD thesis were:

Artificial intelligence and non-invasive sensing technologies for downy mildew evaluation

1.1. Artificial intelligence methods and proximal sensing technologies have demonstrated their capability to automate the non-invasive evaluation of downy mildew in grapevine under laboratory conditions.

1.2. Classic computer vision techniques applied to RGB images have demonstrated their capability for cost-effective, accurate and rapid estimation of downy mildew severity, providing an interpretable localisation of the symptoms that supported disease evaluation.

1.3. The combination of machine learning models, particularly artificial neural networks such as MLP and CNN, with hyperspectral imaging (HSI) has facilitated the analysis of complex spectral data, enabling the early detection of downy mildew.

Fuzzy logic and computer vision for the evaluation of downy mildew severity

2.1. Classic computer vision and fuzzy logic proved their capability for automatic, rapid, accurate, and objective estimation of grapevine downy mildew severity under controlled laboratory conditions.

2.2. The use of a fuzzy threshold to estimate disease severity outperformed the use of a crisp threshold, creating a method strongly related to expert visual assessment. In addition, it provided a comprehensive visual representation of the disease symptoms showing their intensity in the leaves.

2.3. The robustness of the method was demonstrated by its capability to accurately estimating downy mildew severity in different grapevine varieties acquired in different conditions, which showed the potential of its adaptability.

Deep learning for early detection of downy mildew

3.1. The use of convolutional neural networks, combined with transfer learning, enabled the automatic and effective early detection of downy mildew in grapevine in the laboratory, even when the symptoms were barely visible to humans. The Grad-CAM method demonstrated the focus of the neural networks on the downy mildew symptoms of the leaves.

3.2. Image thresholding, especially using fuzzy logic, facilitated the convergence of the neural networks focusing the detection on the symptoms of the disease. On the other hand, the use of raw images helped to achieve an accurate detection in early stages of the infection, avoiding false negative values caused by the thresholding.

3.3. The use of convolutional neural networks also helped to identify the stage of downy mildew infection in the laboratory, differentiating early, intermediate and late stages.

In-field downy mildew detection using explainable deep learning

4.1. The application of convolutional neural networks combined with fine-tuning allowed the accurate detection and localisation of grapevine downy mildew under field conditions, outperforming vision transformers. In addition, explainable artificial intelligence (XAI) offered results that could be comprehensible to farmers.

4.2. The sliding-window method enabled the analysis of grapevine canopy, detecting regions containing small downy mildew symptoms in high-resolution images of plants, allowing straightforward adaptation to new crops or diseases through the annotation of image regions.

4.3. The use of data collected from different vineyards and in different daylight conditions manually and with a mobile platform demonstrated the robustness of the method and opened the possibility for the integration of the algorithm into crop treatment machines like tractors, enhancing in-field disease detection and management.

Deep semantic segmentation for severity estimation of downy mildew under field conditions

5.1. The utilisation of deep semantic segmentation enabled the efficient and objective evaluation of disease severity in images taken in commercial vineyards, localising downy mildew symptoms in the grapevine canopy.

5.2. The combination of models such as U-Net with lightweight encoders like MobileVit-S and transfer learning exhibited superior performance in comparison to smaller architectures like adhoc SegNet, providing a rapid and precise solution that could be adaptable to compact devices or mobile platforms for real-time disease management.

5.3. The localisation of small symptoms was enhanced through background reduction and augmentation of the samples with downy mildew symptoms, while training the models with a dice loss function focused on the symptom. Furthermore, the employment of Test-Time-Augmentation, in conjunction with a restrictive prediction threshold, has been demonstrated to enhance disease detection, thereby reducing errors caused by leaf defects, background variability, and natural light conditions.

8.1 Future work

This research has opened the door to several promising avenues, that collectively aim to enhance the scalability, precision, and practical utility of AI-based disease detection systems in agriculture:

- The adaptation of the developed methodologies to new crops, diseases, and pests would expand their applicability and provide a broader impact in agriculture.
- Further exploration of hyperspectral imaging (HSI), incorporating the use of spectral wavelengths greater than 1000 nm or the consideration of spatial information, could offer a promising approach for the asymptomatic detection of downy mildew, with the potential to improve early assessment in the laboratory or in the field.
- The development of real-time detection systems for field conditions that could enable timely interventions and minimise disease progression.

- The improvement of deep learning methods incorporating advanced data augmentation techniques, such as Generative Adversarial Networks (GANs), could address the difficulty of data collection and improve model robustness and accuracy.

References

Abdelghafour, F., Keresztes, B., Germain, C., and Costa, J.-P. D. (2020). In field detection of downy mildew symptoms with proximal colour imaging. *Sensors*, 20:4380.

Abdullah, H. M., Mohana, N. T., Khan, B. M., Ahmed, S. M., Hossain, M., Islam, K. S., Redoy, M. H., Ferdush, J., Bhuiyan, M., Hossain, M. M., and Ahamed, T. (2023). Present and future scopes and challenges of plant pest and disease (p&d) monitoring: Remote sensing, image processing, and artificial intelligence perspectives. *Remote Sensing Applications: Society and Environment*, 32:100996.

Adem, K., Ozguven, M. M., and Altas, Z. (2023). A sugar beet leaf disease classification method based on image processing and deep learning. *Multimedia Tools and Applications*, 82:12577–12594.

Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. Springer International Publishing.

Ahamed, M. F., Syfullah, M. K., Sarkar, O., Islam, M. T., Nahiduzzaman, M., Islam, M. R., Khandakar, A., Ayari, M. A., and Chowdhury, M. E. H. (2023). Irv2-net: A deep learning framework for enhanced polyp segmentation performance integrating inceptionresnetv2 and unet architecture with test time augmentation techniques. *Sensors*, 23:7724.

Akkem, Y., Biswas, S. K., and Varanasi, A. (2023). Smart farming using artificial intelligence: A review. *Engineering Applications of Artificial Intelligence*, 120:105899.

Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., and Herrera, F. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, 39:2481–2495.

Ball, D., Upcroft, B., Wyeth, G., Corke, P., English, A., Ross, P., Patten, T., Fitch, R., Sukkarieh, S., and Bate, A. (2016). Vision-based obstacle detection and navigation for an agricultural robot. *Journal of Field Robotics*, 33:1107 – 1130.

Bansal, P., Kumar, R., and Kumar, S. (2021). Disease detection in apple leaves using deep convolutional neural network. *Agriculture*, 11:617.

Barbedo, J. G. A. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. *SpringerPlus*, 2:660.

Barker, M. and Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, 17:166–173.

Bijalwan, V., Kumar, V., Kumari, P., and Pascual, J. (2014). Knn based machine learning approach for text and document mining. *International Journal of Database Theory and Application*, 7:61 – 70.

Bock, C. H., Barbedo, J. G. A., Ponte, E. M. D., Bohnenkamp, D., and Mahlein, A.-K. (2020). From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy. *Phytopathology Research*, 2:9.

Bock, C. H., Poole, G. H., Parker, P. E., and Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in Plant Sciences*, 29:59–107.

Boso, S., Santiago, J. L., and Martínez, M. C. (2005). A method to evaluate downy mildew resistance in grapevine. *Agronomy for Sustainable Development*, 25:163–165.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Buonassisi, D., Colombo, M., Migliaro, D., Dolzani, C., Peressotti, E., Mizzotti, C., Velasco, R., Masiero, S., Perazzolli, M., and Vezzulli, S. (2017). Breeding for grapevine downy mildew resistance: a review of "omics" approaches. *Euphytica*, 213:1–21.

Casado-García, A., Heras, J., Milella, A., and Marani, R. (2022). Semi-supervised deep learning and low-cost cameras for the semantic segmentation of natural images in viticulture. *Precision Agriculture*, pages 1–26.

Chen, L. C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv*.

Cortes, C., Vapnik, V., and Saitta, L. (1995). Support-vector networks. *Machine Learning 1995 20:3*, 20:273–297.

Cover, T. M. and Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21–27.

da Cunha, V. A. G., Hariharan, J., Ampatzidis, Y., and Roberts, P. D. (2023). Early detection of tomato bacterial spot disease in transplant tomato seedlings utilising remote sensing and artificial intelligence. *Biosystems Engineering*, 234:172–186.

Diago, M. P., Tardaguila, J., Aleixos, N., Millan, B., Prats-Montalban, J. M., Cubero, S., and Blasco, J. (2015). Assessment of cluster yield components by image analysis. *Journal of the Science of Food and Agriculture*, 95:1274–1282.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv*.

Ethiraj, S. and Bolla, B. K. (2022). Training efficient cnns: Tweaking the nuts and bolts of neural networks for lighter, faster and robust models. *arXiv*.

Fuentes, S., Tongson, E., and Viejo, C. G. (2024). *Artificial intelligence and Big Data revolution in the agrifood sector*, pages 171–188. Elsevier.

Gamarra, M., Zurek, E., Escalante, H. J., Hurtado, L., and San-Juan-Vergara, H. (2019). Split and merge watershed: A two-step method for cell segmentation in fluorescence microscopy images. *Biomedical Signal Processing and Control*, 53:101575.

Gao, J., Westergaard, J. C., Sundmark, E. H. R., Bagge, M., Liljeroth, E., and Alexandersson, E. (2021). Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning. *Knowledge-Based Systems*, 214:106723.

Gao, Z., Khot, L. R., Naidu, R. A., and Zhang, Q. (2020). Early detection of grapevine leafroll disease in a red-berried wine grape cultivar using hyperspectral imaging. *Computers and Electronics in Agriculture*, 179:105807.

García, S., Luengo, J., and Herrera, F. (2015). *Data Preprocessing in Data Mining*, volume 72. Springer International Publishing.

Goh, T. Y., Basah, S. N., Yazid, H., Safar, M. J. A., and Saad, F. S. A. (2018). Performance analysis of image thresholding: Otsu technique. *Measurement*, 114:298–307.

Gonçalves, J. P., Pinto, F. A., Queiroz, D. M., Villar, F. M., Barbedo, J. G., and Ponte, E. M. D. (2021). Deep learning architectures for semantic segmentation and automatic estimation of severity of foliar symptoms caused by diseases or pests. *Biosystems Engineering*, 210:129–142.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks. *Science Robotics*, 3:2672–2680.

Gutiérrez, S., Hernández, I., Ceballos, S., Barrio, I., Díez-Navajas, A. M., and Tardaguila, J. (2021). Deep learning for the differentiation of downy mildew and spider mite in grapevine under field conditions. *Computers and Electronics in Agriculture*, 182:105991.

Gutiérrez, S. and Tardaguila, J. (2023). Evolutionary conditional gans for supervised data augmentation: The case of assessing berry number per cluster in grapevine. *Applied Soft Computing*, 147:110805.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46:389–422.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2016-Decem, pages 770–778. IEEE.

Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence*, 40:185–234.

Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., and Le, Q. (2019). Searching for mobilenetv3. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, volume 2019-Octob, pages 1314–1324. IEEE.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., and Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, 2:230–243.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349:255–260.

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

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60:84–90.

Kumar, M. and Jindal, S. R. (2019). Fusion of rgb and hsv colour space for foggy image quality enhancement. *Multimedia Tools and Applications*, 78:9791–9799.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–444.

LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C., Denker, J. S., Drucker, H., Guyon, I. M., Muller, U., Sackinger, E., Simard, P. Y., and Vapnik, V. N. (1995). Learning algorithms for classification: A comparison on handwritten digit recognition.

Lee, W. S. and Tardaguila, J. (2023). *Pest and Disease Management*, pages 93–118. Springer, Cham.

Li, L., Zhang, S., and Wang, B. (2021). Apple leaf disease identification with a small and imbalanced dataset based on lightweight convolutional networks. *Sensors*, 22:173.

Li, R., Zheng, S., Zhang, C., Duan, C., Su, J., Wang, L., and Atkinson, P. M. (2022). Multi-attention-network for semantic segmentation of fine-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13.

Li, W., Dong, R., Fu, H., and Yu, L. (2019). Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks. *Remote Sensing*, 11.

Li, W., Yu, X., Chen, C., and Gong, Q. (2023). Identification and localization of grape diseased leaf images captured by uav based on cnn. *Computers and Electronics in Agriculture*, 214:108277.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88.

Madden, L. V., Hughes, G., and van den Bosch, F. (2017). *The Study of Plant Disease Epidemics*. The American Phytopathological Society.

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease*, 100:241–251.

Mahlein, A.-K., Kuska, M. T., Thomas, S., Wahabzada, M., Behmann, J., Rascher, U., and Kersting, K. (2019). Quantitative and qualitative phenotyping of disease resistance of crops by hyperspectral sensors: seamless interlocking of phytopathology, sensors, and machine learning is needed! *Current Opinion in Plant Biology*, 50:156–162.

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L. R., Davis, C. E., and Dandekar, A. M. (2015). Advanced methods of plant disease detection. a review. *Agronomy for Sustainable Development*, 35:1–25.

McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. (1995). A proposal for the dartmouth summer research project on artificial intelligence. *AI Magazine*, 27:12–12.

Mehta, S. and Rastegari, M. (2021). Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. In *ICLR 2022 - 10th International Conference on Learning Representations*. International Conference on Learning Representations, ICLR.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill.

Mukherjee, A. (2020). Analysis of diseased leaf images using digital image processing techniques and svm classifier and disease severity measurements using fuzzy logic. *International Journal of Scientific & Engineering Research*, 11:1905–1912.

Mwebaze, E. and Owomugisha, G. (2016). Machine learning for plant disease incidence and severity measurements from leaf images. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 158–163. IEEE.

Nagi, R. and Tripathy, S. S. (2021). Severity estimation of grapevine diseases from leaf images using fuzzy inference system. *Agricultural Research*, pages 1–11.

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., and Walsh, J. (2020). *Deep Learning vs. Traditional Computer Vision*, volume 943, pages 128–144. Springer Verlag.

Palacios, F., Bueno, G., Salido, J., Diago, M. P., Hernández, I., and Tardaguila, J. (2020). Automated grapevine flower detection and quantification method based on computer vision and deep learning from on-the-go imaging using a mobile sensing platform under field conditions. *Computers and Electronics in Agriculture*, 178:105796.

Paleyes, A., Urma, R.-G., and Lawrence, N. D. (2023). Challenges in deploying machine learning: A survey of case studies. *ACM Computing Surveys*, 55:1–29.

Pan, T., Chyngyz, E., Sun, D. W., Paliwal, J., and Pu, H. (2019). Pathogenetic process monitoring and early detection of pear black spot disease caused by alternaria alternata using hyperspectral imaging. *Postharvest Biology and Technology*, 154:96–104.

Paulus, I., Busscher, R. D., and Schrevens, E. (1997). Use of image analysis to investigate human quality classification of apples. *Journal of Agricultural Engineering Research*, 68:341–353.

Petropoulos, S., Karavas, C. S., Balafoutis, A. T., Paraskevopoulos, I., Kallithraka, S., and Kotseridis, Y. (2017). Fuzzy logic tool for wine quality classification. *Computers and Electronics in Agriculture*, 142:552–562.

Rehman, T. U., Mahmud, M. S., Chang, Y. K., Jin, J., and Shin, J. (2019). Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Computers and Electronics in Agriculture*, 156:585–605.

Rodríguez, J. P., Corrales, D. C., Aubertot, J.-N., and Corrales, J. C. (2020). A computer vision system for automatic cherry beans detection on coffee trees. *Pattern Recognition Letters*, 136:142–153.

Román, C., Llorens, J., Uribeetxebarria, A., Sanz, R., Planas, S., and Arnó, J. (2020). Spatially variable pesticide application in vineyards: Part ii, field comparison of uniform and map-based variable dose treatments. *Biosystems Engineering*, 195:42–53.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351:234–241.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115:211–252.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2:160.

Sekulska-Nalewajko, J. and Goclawski, J. (2011). A semi-automatic method for the discrimination of diseased regions in detached leaf images using fuzzy c-means clustering. In *Perspective Technologies and Methods in MEMS Design*, pages 172–175. IEEE.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128:336–359.

Serre, T. (2019). Deep learning: The good, the bad, and the ugly. *Annual Review of Vision Science*, 5:399–426.

Sharma, R., Kukreja, V., and Sakshi (2021). Mustard downy mildew disease severity detection using deep learning model. In *2021 International Conference on Decision Aid Sciences and Application, DASA 2021*, pages 466–470. IEEE.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35:1285–1298.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6:60.

Shortliffe, E. H. (1977). Mycin: A knowledge-based computer program applied to infectious diseases. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pages 66–69.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR.

Smith, M. L., Smith, L. N., and Hansen, M. F. (2021). The quiet revolution in machine vision - a state-of-the-art survey paper, including historical review, perspectives, and future directions. *Computers in Industry*, 130:103472.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 07-12-June-2015, pages 1–9. IEEE.

Szeliski, R. (2022). *Computer Vision*. Springer International Publishing.

Tan, M. and Le, Q. V. (2021). Efficientnetv2: Smaller models and faster training. *arXiv*.

Tardaguila, J., Stoll, M., Gutiérrez, S., Proffitt, T., and Diago, M. P. (2021). Smart applications and digital technologies in viticulture: A review. *Smart Agricultural Technology*, 1:100005.

Thai, H.-T., Tran-Van, N.-Y., and Le, K.-H. (2021). Artificial cognition for early leaf disease detection using vision transformers. In *2021 International Conference on Advanced Technologies for Communications (ATC)*, volume 2021-Octob, pages 33–38. IEEE.

Toffolatti, S. L., Russo, G., Campia, P., Bianco, P. A., Borsa, P., Coatti, M., Torriani, S. F., and Sierotzki, H. (2018). A time-course investigation of resistance to the carboxylic acid amide mandipropamid in field populations of plasmopara viticola treated with anti-resistance strategies. *Pest Management Science*, 74:2822–2834.

Tong, P., Han, P., Li, S., Li, N., Bu, S., Li, Q., and Li, K. (2021). Counting trees with point-wise supervised segmentation network. *Engineering Applications of Artificial Intelligence*, 100:104172.

Vallejo-Pérez, M. R., Sosa-Herrera, J. A., Navarro-Contreras, H. R., Álvarez Preciado, L. G., Ángel Gabriel Rodríguez-Vázquez, and Lara-ávila, J. P. (2021). Raman spectroscopy and machine-learning for early detection of bacterial canker of tomato: The asymptomatic disease condition. *Plants*, 10:1–13.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Łukasz Kaiser, and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-December:5999–6009.

Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018:1–13.

Zhang, B. H., Li, J. B., Zheng, L., Huang, W. Q., Fan, S. X., Zhao, C. J., and Meng, Q. D. (2015). Development of a hyperspectral imaging system for the early detection of apple rottenness caused by penicillium. *Journal of Food Process Engineering*, 38:499–509.

Zhang, X., Xun, Y., and Chen, Y. (2022). Automated identification of citrus diseases in orchards using deep learning. *Biosystems Engineering*, 223:249–258.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2016). Pyramid scene parsing network. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January:6230–6239.

Zhou, C., Zhong, Y., Zhou, S., Song, J., and Xiang, W. (2023). Rice leaf disease identification by residual-distilled transformer. *Engineering Applications of Artificial Intelligence*, 121:106020.

Íñiguez, R., Palacios, F., Barrio, I., Hernández, I., Gutiérrez, S., and Tardaguila, J. (2021). Impact of leaf occlusions on yield assessment by computer vision in commercial vineyards. *Agronomy*, 11:1003.