2. Artificial intelligence and protected cultivation

Greenhouse production processes are already highly automated and controlled but, similar to what is occurring in many sectors, AI systems are now taking control to unprecedented levels. Because of their potential ability to process large amounts of data and make tiny continuous adjustments, AI systems are beginning to provide greenhouse operators with myriads of production-related benefits (Treena Hein, 2021)

AI in protected horticulture can predict yield, ensure product quality from starting material to harvest, help decide on the planning of time-to-market and resources used and improve efficiency. It can, therefore, contribute to the economic profit of growers and the sustainability of their production. Both are important factors in industrialised production processes with large greenhouse compartments at different locations, a lack of skilled labour and increased demand for high-value food close to urban areas. In addition, the link between growing conditions and shelf-life processes needs to be elaborated, such that information from the end point (the consumer acceptance) is used as feedback to alter growing conditions.

2.1. Plant phenotyping of horticultural crops and the use of crop sensors

Plant phenotyping can be defined as the set of methodologies and protocols used to measure plant growth, architecture and composition with a certain accuracy and precision at different scales of organisation, from plant organs to complete crop canopies. The term is often restricted to plant breeding purposes, but it can also be used for plant production, specifically where measured plant features are used for precise crop maintenance and crop control in a controlled environment, such as (autonomous) greenhouses and vertical farms.

2.1.1. Digitalisation and artificial intelligence for crop morphology measurements

The shape and morphology of plants is related to variety, the underlying genetics and environmental factors (light, temperature, irrigation). Digital plant phenotyping refers to the use of computers for plant phenotyping where digital sensors are used to measure plant characteristics.

One of the most common digital phenotyping methodologies is image analysis, where cameras are used to record images and software is used to automatically extract the measurements from the images to access plant morphology (the shape of a plant), in a reproducible and accurate way (Van der Heijden & Polder, 2015).

Currently, many different types of cameras are available for measuring important plant features to characterise plant morphology. The most used camera is the RGB colour camera, which produces images in the visible spectrum, mimicking the human eye. To relate the images to real dimensions, 3D information is often needed, which resulted in RGB camera-based 3D sensors. The Intel Real-sense RGBD sensor is an affordable example of a RGB 3D sensor and is often used in horticultural phenotyping, e.g., for tomato fruit detection and counting (Afonso et al., 2020; Fonteijn et al., 2021). Other examples are LiDAR sensors. All of these might become low cost because of the development of smart phone cameras for consumers.

In greenhouse crop production, the plants may be intertwined, and so they cannot be easily imaged from all sides. This leads to occlusion and hampers the possibility of imaging important plant traits with a 3D camera. To overcome this problem, more advanced imaging solutions are needed. This can either be achieved by a moving trolley system with a mounted camera, flying drones inside the greenhouse or a robot that scans the plant with a 3D camera from many viewpoints. Using artificial intelligence algorithms, the point clouds from different single viewpoints are converted into a robust representation of the crop (Boogaard et al., 2020).

2.1.2. Digitalisation and artificial intelligence for crop physiology performance

Next to plant morphology, plant physiological processes are important for crop monitoring. In crop production, photosynthesis in the leaves yields important biochemicals, such as sugars, starch, chlorophyll and nutrients, that are transferred to the plant organs, flowers and fruits (Dieleman et al., 2018a).

Therefore, measuring the efficiency of plant photosynthesis directly and non-destructively is a desirable way for obtaining information on crop performance and for the early detection of deviations from optimal physiological conditions. Technologies like chlorophyll fluorescence imaging and thermal imaging are promising, especially if they can be applied to other parts of crop canopies, as well as individual leaves.

The chemical composition of the crop can be determined by sampling leaves or fruits, sending them to a laboratory and waiting for the analysis. Recent imaging spectroscopy was tested on a laboratory scale, to determine the composition of biochemicals in crops, with promising results (Dieleman et al., 2018b).

Imaging spectroscopy is an imaging technique for images taken using many narrow wavelength bands over a range extending across the visible spectrum (from ultraviolet to shortwave infrared) and compared to a standard camera, which only records red, green and blue light. In doing so, it creates an extremely detailed image of the reflection of light on plants or other objects. Imaging spectroscopy provides a lot of information on plant pigments, sugars, proteins, fats and water, as well as their distribution over the leaves or organs.

Regions of interest, such as the fruits or leaves, can be automatically extracted from the image. This opens the possibility of using this technique on mobile platforms (Mishra et al., 2020). Currently, a lot of research and development effort is going into the development of spectral cameras, making them less bulky, more robust, faster, and less costly.

Currently, AI techniques are explored to extract useful information from the massive amount of data collected by the spectral cameras (Mishra et al., 2021; Signoroni et al., 2019).

These developments suggest an outlook for the future, providing more information on different important plant features. Until now, most plant features could only be measured manually, destructively and/or very locally with scarce datapoints. Digitalisation of the measurement and use of modern sensors and camera systems will help to collect more datapoints. AI methods will largely help in the interpretation of variable data output. AI algorithms will also help to transform and combine the output of multiple sensors into useful information for growers.

2.2. Autonomous growing and the use of AI

Greenhouse horticulture is characterised by relatively high operational efficiency involving powerful managerial skills. However, demand for high vitamin and mineral food is increasing rapidly (Rabobank, 2018; Tilman et al., 2011). The volatile market demands, resource prices, scarcity of experienced labour (Brian, 2018), as well as uncertain weather and environmental conditions, make greenhouse farming a complex and risky endeavour. While encountering an environmental crisis (United Nations, 2019), food production systems need to become more productive, resource-efficient, and environmentally sustainable (Willett et al., 2019). The development of advanced and autonomous greenhouse production systems aims at realising the best possible production outcomes, considering quality and sustainability targets, with the uncertainties of resource availability, weather or market demand. The realisation of fully autonomous and intelligent horticulture entails three major components: remote sensing, remote control, and hands-free practices with robotisation.

2.2.1. Data for autonomous growing and production

Data regarding greenhouse production systems are becoming of increasing importance and are a means of deeper understanding and efficient management of the complex biological dynamic processes. Large and meaningful datasets about all growing aspects are sparse. The greenhouse climate is relatively well-monitored, resulting in a time series with short intervals. However, manual, subjective, time-consuming, often invasive, and costly measurements of traits of crop growth, development, pests, and pathogens result in fragmented weekly or bi-weekly data points (Bouzembrak et al., 2020). This implies considerable data uncertainty as a result of noise, missing data, inconsistent formats, and non-standard collection protocols, among others (Lezoche et al., 2020). Investment into integrating diverse and unstructured data is required before any additional meaningful insights are possible (Osinga et al., 2022).

Ongoing technological developments, computational power, and high-fidelity sensors offer new opportunities for automated, remote, and non-invasive sensing of growing parameters. The higher spatial and temporal resolution in the measurements and in the growing conditions allows for interpretation of the system's variability at coarser and granular levels and offers opportunities for sufficient information extraction towards more efficient adaptation of horticultural practices.

AI and machine learning can deal with the larger datasets and capture the nonlinear relationships present in the heterogeneous data sources in greenhouses.

2.2.2. Machine learning for yield prediction and resource use efficiency

Scalable and generic machine learning analytics are currently used to complement expert-based approaches for supporting yield predictions. Implementations of intelligent algorithms focus on predictions of indoor climate, microclimate (Ali & Hassanein, 2020; Taki et al., 2016, 2018), yield and quality aspects of vegetable crops and flowers (Alhnaity et al., 2020; Reissig et al., 2021; Xiao et al., 2021), as well as growth and development indicators. Descriptive and predictive models (Partial Least Squares (PLS)) (Li et al., 2016), Support Vector Machines (SVMs) (Fandel et al., 2021; Yan et al., 2010), Random Forests (RF) (Amir et al., 2021), Artificial Neural Networks (ANNs) (Ullah et al., 2020), and k-nearest Neighbours (KNNs) have developed yield forecasting tools and decision support systems (DSS) using predictors and outcomes from experience. In addition to ML, deep learning (Long Short-Term Memory (LSTM)) (Alhnaity et al., n.d.; Ali & Hassanein, 2020; Moon et al., 2020), Temporal Convolutional Networks (TCN) (Gong et al., 2021), and Multilayer Perceptron Neural Networks (MLP-NN) (Petrakis et al., 2022) have also shown significant advantages in processing time-series data to yield higher precision and better performance than other machine learning methods.

Reinforcement learning finds applications in selecting actions, based on continuous feedback, to maximise the system's performance. Current applications are aimed at learning the best operational decision for day-to-day climate optimisation, with fewer being aimed at irrigation control and crop management planning. Experiments for greenhouse control at a distance, using state-of-the-art artificial intelligence algorithms, yielded promising results in the cultivation of cucumbers (Hemming et al., 2019) and cherry tomatoes (Hemming et al., 2020), compared to references of experience-based manual growing. Different AI technologies have been shown to have the potential to contribute to predicting yield, as well as increasing yield and product quality and, at the same time, save resources such as energy, water and nutrients.

2.2.3. Deep learning for pest and pathogen management

In the future, the detection of plant pathogens and pests will become extremely important. Unless it is known what a plant is suffering from, nothing can be done about it. The earlier pests and pathogens are identified, the easier it is to control them. Automated systems are starting to play a greater part in this (Bauriegel et al., 2011; Polder et al., 2014; Rumpf et al., 2010).

Automatic detection of pathogens in plants, as early as possible and without damaging the plant, is an approach that is gaining ever more attention in horticulture. In automatic detection, the basic assumption is that a diseased plant looks different from a healthy one. For example, leaves can have subtle colour differences, which are often invisible to the human eye but can be captured using techniques such as spectral imaging. Spectral imaging, combined with deep learning techniques (described in the previous section), has the potential to become a powerful tool in pathogen detection in greenhouses and vertical farms.

Pest detection is often challenging because pests and their eggs are often located underneath the plant canopy and are, therefore, difficult to detect. They are often very small and show a very local distribution. Crops in general might suffer from multiple pests at the same time. Therefore, not only high-resolution detection but also local and organism specific detection is required. High-resolution imaging, in combination with deep learning techniques might have the potential for precision farming in greenhouses and vertical farms.

In both cases, large amounts of labelled images are required from different situations (locations, seasons, crop varieties) to sufficiently train the deep learning algorithms. More smart training is needed to overcome the lack of such real data and labelled images.

2.3. Digital twins and decision support for market-oriented production

Today's high-tech greenhouses are equipped with different standard sensors for monitoring light, temperature, humidity, and CO_2 and for actively controlling different actuators (e.g. lighting, screening, heating, ventilation, cooling, CO_2 dosing, fogging, dehumidification, irrigation, and fertiliser dosing) in order to control all growth factors important for crop production at every moment. Today's growers determine the climate, irrigation and crop management strategies based on experience and define the setpoints for climate and irrigation control manually. Actuators then operate based on the setpoints configured in a processing computer, while sensors give feedback on measured data for the control loop (Hemming et al., 2020).

The rapid pace of technological advancements, AI, cloud computing, and the uptake of the IoT produces an increasing data stream at high spatial and temporal resolution, almost in real-time.

In smart horticulture, the greenhouse grower can monitor and control operations at a distance, based on real-time digital information instead of direct observations and tasks on-site.

Large amounts of data can be leveraged for the design and implementation of advanced models, known as digital twins. A digital twin is equivalent to real-life objects mirroring the behaviour and states over its lifetime in a virtual space (C. Verdouw et al., 2021). As a digital representation of actual physical systems and technology integrators, digital twins offer a solution for complex systems analysis and can act as decision support tools (Pylianidis et al., 2021). Digital twins are increasingly adopted in the manufacturing, automotive, and energy industries (Caputo et al., 2019; Kritzinger et al., 2018; Sivalingam et al., 2018).

2.3.1. Digital twin of the greenhouse system

Dynamic climate models have been developed (Vanthoor et al., 2011) which act as digital twins of real greenhouses. An overview of today's greenhouse climate models was given in a previous study (López-Cruz et al., 2018). Since greenhouses differ from each other, an appropriate parameter determination or calibration is necessary for each model, to act as a digital twin of an existing greenhouse. These mechanistic digital twin models can be used to assist intelligent decision support on climate control actions. Simulations of past or future scenarios provide information on how different climate control in the past could have improved crop production and which actions are required to reach a certain crop production goal in the future. These models can also be coupled with intelligent algorithms to automatically determine climate setpoints, an action that is currently performed manually by the grower. In order to control crop production by an automated algorithm, mechanistic greenhouse climate and crop models are coupled to resemble a real greenhouse. The effects of changing set points can be tested on the digital twin and then, sent automatically to a processing computer to control the different actuators (Hemming et al., 2020).

2.3.2. Digital twin of the crop

The crop has a central role in every greenhouse production system. Crop management decisions and actions are mostly taken by the greenhouse staff. Since experienced and well-trained crop managers are scarce, crop simulation models can play a role in decision making. An overview of greenhouse crop models and modelling approaches are given in other studies (Kuijpers et al., 2019; Sarlikioti et al., 2011). Crop models can be used as virtual representations of reality (Marshall-Colon et al., 2017). They can be used to simulate different growing conditions and crop management strategies and to predict their effect on crop development and yield, as well as on fruit quality. Crop models can help to understand the crop behaviour under different growing conditions and can support the grower in making decisions. Additional sensors, monitoring crop status, can provide the grower with further information as described in the previous chapter. While automated greenhouse climate control algorithms have already been developed and are widely introduced in modern high-tech greenhouses, automated control procedures for crop status are still in their infancy (Hemming et al., 2020).

The available digital twins do not yet include all aspects for crop production. Typically, water and nutrient management could be described in more detail. Crop quality aspects are not described well and pest and pathogen management is lacking. More attention needs to be paid to the completion of mechanistic digital twins in future research.

2.3.3. Digital twins for decision support and AI

In general, complete digital twins (including greenhouse twins, the physical environment and crop twins) can facilitate operational and tactical management decisions, strategic design decisions, and predictive maintenance information. Preventive and corrective actions can be simulated and evaluated in the digital environment before the final actual intervention. Such complete digital twins are highly suitable for capturing available 'horticultural/green' knowledge and obtaining artificial training datasets for future system design and operation.

Convergence between digital and physical greenhouse production systems has been pursued as an essential goal for data-driven horticulture. In the domain of process systems engineering, Reinforcement Learning (RL) has been applied to resolve stochastic optimal control challenges with the uncertainties of the highly non-linear and complex processes. As real-world data is augmented in mechanistic algorithms that comprise the digital twin, the virtual environment can act as a learning environment that generates adaptive control actions with statistical significance, instead of the conventional hardcoded control logics of deterministic conditions.

Deep RL networks require finite learning iterations. To explore the potential of such data-greedy networks for horticultural challenges in a practical, timely and economically feasible manner, data from the digital twins can be used as it is repeatable, inexpensive, and clean. In view of conditional, highly automated and high-fidelity twins, interventions suggested in the digital twin can be directly implemented without the grower's inspection or physical proximity. The twins are able to self-diagnose and adapt to users' preferences (C. N. Verdouw et al., 2016). The benefits can result in cost savings of recourses, improved product quality, faster actions with lower risks, and increased production (Pylianidis et al., 2021; C. Verdouw et al., 2021).