Hungarian University of Agriculture and Life Sciences

Doctoral School of Plant Science

Ph.D. Dissertation

Integrative Approaches in Enhancing Processing Tomato (*Solanum lycopersicum L.*) Cultivation: From Root Development and Machine Learning Predictions to Genetic Diversity Evaluation

DOI: 10.54598/004840

By

Oussama M’hamdi

Gödöllő, Hungary

2024

The Doctoral School: Doctoral School of Plant Sciences

Name: Oussama M’hamdi

Discipline: Agricultural Sciences

Head: *Professor Dr. Lajos Helyes*
Head of Doctoral schools of the Hungarian University of Agricultural and Life Sciences
Director of Institute of Horticultural Sciences
Hungarian University of Agriculture and Life Sciences

Supervisors: *Professor Dr. Zoltán Pek*
Institute of Horticultural Sciences
Hungarian University of Agriculture and Life Sciences

Dr. Sándor Takács
Institute of Horticultural Sciences
Hungarian University of Agriculture and Life Sciences

………………………………………

Approval of the Head of Doctoral School

……………………………… ………………………………

Approval of the Supervisor Approval of the Supervisor

Table of contents

1 INTRODUCTION

The tomato is not only a fundamental component of global diets but also a cornerstone of a multi-billion-dollar international industry. Tomatoes (Solanum lycopersicum L.), belonging to the Solanaceae family, are key players in both global diets and agricultural research due to their nutritional value, economic importance, and health benefits. Their widespread cultivation is supported by advanced breeding techniques and sustainable farming practices, which have significantly increased yield and quality. In 2021, the global tomato production reached approximately 186 million tons across 4.9 million hectares (FAO, 2022). This widespread cultivation highlights the tomato's role as a fundamental component in various global cuisines and as a crucial element in the processed food sector (J. Liu et al., 2021; Nemeskéri et al., 2019; X. Wang et al., 2019). Furthermore, the integration of advanced agricultural technologies and sustainable farming practices is playing a significant role in boosting tomato yields globally. Beyond their culinary adaptability, tomatoes are praised for being rich in nutrients and bioactive compounds. They contain a diverse range of sugars, acids, vitamins, and other vital nutrients (Agbemafle et al., 2014; Ali et al., 2021; Collins et al., 2022; Daood et al., 2014; Takács et al., 2020; Vélez-Terreros et al., 2021). Their distinctive organoleptic qualities, characterized by a balanced sweet-sour taste, a striking pink-red hue, and an alluring aroma, are primarily attributed to their high sugar and acid content (Gharezi et al., 2012).

The cultivation of processing tomatoes, faces numerous challenges. These challenges encompass a broad array of environmental stressors, notably the variability in water availability, such fluctuations in water supply significantly affect the quality and yield of tomatoes, impacting each phenological stage of the plant's growth in distinct ways (Nemeskéri & Helyes, 2019; Takács et al., 2020). Adding to these challenges is the necessity to consistently adhere to stringent fruit quality standards, a task that becomes increasingly challenging under diverse climatic conditions (Giuliani et al., 2019). Furthermore, the ripening process of tomatoes is yet another critical aspect, marked by significant alterations in metabolic pathways (Zhu et al., 2022). These changes are crucial in determining the fruit's external appearance, internal quality parameters such as Brix value and Lycopene content, and the distinctive colour indicative of ripeness. Therefore, predicting quality in diverse climatic conditions becomes paramount, as uniformity in quality and appearance is essential for consumer acceptance. Additionally, the genetic diversity inherent in different tomato genotypes presents both challenges and opportunities. While this diversity demands adaptation in cultivation practices for each genotype, it also provides significant opportunities for agricultural advancements and the breeding of varieties more resilient to specific environmental stressors (Udriște et al., 2022). Therefore, the cultivation of processing tomatoes encompasses a wide spectrum of scientific and practical considerations, each playing a crucial role in ensuring the sustained production and availability of this globally important crop.

Current research exhibits a significant gap in fully understanding how environmental factors affect tomato root development and fruit quality. The potential of using advanced predictive techniques, such as machine learning, in determining fruit quality in relation to varying environmental and genetic factors is also yet to be extensively explored. This thesis investigates the intricate relationship between environmental conditions and root development, it addresses the integration of advanced machine learning methodologies in the prediction of fruit quality attributes and it examine the interaction between genetic makeup and environmental factors.

Objectives to achieve

The primary aim of this research is to comprehensively understand the growth, development, and quality of processing tomato plants in response to various environmental conditions and genetic factors. Specifically, this study seeks to achieve the following objectives:

To Investigate Root Development in Processing Tomato Plants under Different Water Supply Levels

Investigate the impact of differential water supply on the root system architecture of processing tomato plants (Solanum lycopersicum). This includes monitoring changes in root count, length, and overall development using non-destructive methods. The goal is to understand how water stress or abundance affects root growth patterns, potentially impacting overall plant health and yield

To Perform a Comparative Analysis of Machine Learning Models in Predicting Tomato Fruit Quality

Utilize two advanced machine learning techniques, eXtreme Gradient Boosting (XGBoost) and Artificial Neural Network (ANN), to predict key quality attributes of processing tomato fruits. These attributes include Brix, Lycopene content, and a/b ratio. The data for this analysis comprises variables like different cultivars, planting locations, years, and climatic factors, aiming to establish robust predictive models for fruit quality assessment.

To Evaluate the Genetic Resources of Processing Tomato Plants in Diverse Environments

Conduct a comprehensive assessment of the genetic diversity among different processing tomato genotypes and their response to various environmental conditions using GGE biplot analysis. This objective focuses on understanding how genetic variation influences important quality traits like Brix and Lycopene content across different years and locations, providing insights into genotype-environment interactions.

2 LITERATURE OVERVIEW

2.1 Tomato

2.1.1 Economic importance

The tomato is a prominent agricultural product with a significant role in the global agricultural sector, both in terms of economic impact and widespread consumption (Agbemafle et al., 2014). Its versatility, being utilized in both fresh and processed forms, has made it a staple in various cultures and cuisines around the world (Bergougnoux, 2014). the market value of tomatoes was remarkably high, estimated at 181.74 billion US dollars. This economic importance is underscored by the volume of tomato production: an estimated 186.1 million metric tons were harvested, covering a cultivation area of nearly 4.9 million hectares (FAO, 2022). These statistics not only highlight the extensive scale of tomato agriculture but also indicate a progressive increase in the cultivation and consumption of tomatoes globally. Looking at national statistics provides a more detailed perspective. For instance, Hungary, in the year 2022, produced approximately 137.8 thousand metric tons of tomatoes, cultivated over an area of around 1.6 thousand hectares (FAO, 2022). This mirrors the global trend of increasing tomato cultivation, thereby emphasizing the substantial role of tomatoes in both domestic and international markets. The escalating demand for tomatoes can be attributed to their multifaceted utility in various culinary forms, including soups, sauces, juices, and concentrated powders. In the broader context of global crop production, tomatoes hold a significant position, ranking eighth after major crops like sugar cane, maize, wheat; rice, potatoes, soybeans, and cassava. According to 2022 data from the Food and Agriculture Organization of the United Nations Statistics Division (FAO, 2022), these statistics not only illustrate the robust and expanding nature of the tomato industry but also underscore its critical role in the global agricultural economy.

2.1.2 Nutritional value

Tomatoes have been the subject of extensive research due to their nutritional components and associated health benefits. Their nutritional profile is rich and varied, including a range of vitamins, minerals, and bioactive compounds. Among these, lycopene stands out as a potent antioxidant and one of the most studied components in tomatoes. It is recognized for its potential to reduce the risk of certain types of cancers and cardiovascular diseases. For instance, Giovannucci (1999) highlighted an inverse relationship between lycopene consumption and the risk of prostate cancer. Similarly, Burton-Freeman & Sesso (2014) underscored the role of lycopene in reducing oxidative stress and improving heart health.

In addition to lycopene, tomatoes are also a significant source of essential nutrients such as Vitamin C, Vitamin K, potassium, and folate. The research of Burton-Freeman & Reimers (2011) highlights the significance of these nutrients in promoting overall health. They are particularly noted for enhancing immune function, aiding in blood pressure regulation, and supporting bone health. Additionally, as Slavin & Lloyd (2012) discuss, the dietary fiber content in tomatoes plays a crucial role in digestive health and assists in regulating blood glucose levels. This makes tomatoes a beneficial dietary component in the management of diabetes. Additionally, recent studies have increasingly focused on the potential of tomatoes in mitigating the risk of neurodegenerative diseases. Meeusen (2014) investigated the presence of carotenoids and flavonoids in tomatoes, which have been linked to improved cognitive function and a lower risk of neurodegenerative diseases. This research is particularly relevant considering the aging global population and the rising prevalence of conditions like Alzheimer's disease.

Moreover, the impact of cultivation methods and tomato varieties on nutritional value has been a topic of research. Oliveira et al. (2013) found that organically grown tomatoes tend to have higher levels of phenolic compounds than conventionally grown varieties, suggesting variations in health benefits. Additionally, the processing of tomatoes into products like sauces and juices also influences their nutritional content. Basu & Imrhan (2007) noted that processing can enhance the bioavailability of certain nutrients, notably lycopene.

2.2 Root Development under Different Water Supply Levels

2.2.1 Overview of Root Development in Plants

The study of root systems, which are crucial for plant growth, nutrient uptake, and overall health, is a well-established and significant area in plant sciences. The literature on this topic is extensive, with several significant studies providing foundational knowledge (Bellini et al., 2014; Kafkafi, 2008; Osmont et al., 2007; H. Wang et al., 2006; Y. Zhang et al., 2022). This overview aims to synthesize key concepts and findings from these sources, offering a comprehensive understanding of root development in plants.

2.2.1.1 Root Types and Classification Systems

The vast morphological diversity of root systems in the plant kingdom requires a detailed and comprehensive framework for classification. Such categorization, pivotal in botanical studies, predominantly hinges on parameters such as the root's size, quantity, and ontogenetic origin. Notable typologies include fibrous roots, characterized by a mesh-like structure; storage roots, which primarily function as nutrient reservoirs; and tap roots, distinguished by their singular, central prominence (Gliński & Lipiec, 1990; Guerrero-Campo & Fitter, 2001). Additionally, the developmental classification approach, detailed by Waisel et al. (1996), has gained attention for its analytical value in understanding root evolution and functionality. This approach not only offers a deeper insight into root system architecture but also sheds light on the adaptive strategies different plant species employ in various environmental settings.

2.2.1.2 Root Structure and Function

The morphological complexity of plant root systems is characterized by a diverse array of structures. Seminal roots, which develop directly from the seed, and adventitious roots, emerging from internodal regions, are predominant features. Additionally, lateral roots, branching off from existing root structures, add to this intricate system (Bellini et al., 2014; Gonin et al., 2019). This variation in root size and form is often a phenotypic response to environmental factors, serving functions such as water and mineral absorption, anchorage, nutrient storage, and propagation (Givnish, 2002; Ristova & Busch, 2014). Root growth is also influenced by a range of external factors, including biological (such as old wormholes), chemical (like nutrient availability, soil pH, and redox conditions), hydrological, microclimatic (soil temperature), and physical elements (soil texture and structure). This dynamic interaction between root architecture and environmental factors plays a crucial role in plant ecology and adaptation, underscoring the importance of roots in both resource acquisition and environmental responsiveness.

2.2.1.3 Root-to-Shoot Ratio

The root-to-shoot ratio (R:S) serves as a critical indicator of a plant's balance. This ratio changes depending on various factors, including the plant's development stage, growth conditions, plant type, and genotype (Bláha, 2019). Typically, most plants show a logarithmic increase in R:S during the vegetative growth phase under stable conditions. Root system patterns and sizes are also influenced by factors like Raunkiaer’s life forms, ecological strategies, and seed size (Guerrero-Campo & Fitter, 2001). The adaptive significance of the relationship between root and shoot growth has also been highlighted, Chen et al. (2016) emphasize that the root-to-shoot ratio is ultimately regulated by maintaining a balance between the root's functions and the shoot's, which is crucial for optimal carbon fixation. Further, the root-to-shoot ratio serves as an indicator of plant responses to environmental stressors and is instrumental in efficient crop management and resource allocation (Agathokleous et al., 2019), thereby linking plant biology with broader ecological and agricultural practices.

2.2.1.4 Roots in Crop Breeding

Historically, crop breeding programs have primarily focused on enhancing shoot-related traits, often overlooking the importance of root systems in these efforts (McGrail et al., 2020; Tracy et al., 2020). It has been observed that modern cultivars tend to have smaller root systems and lower root-to-shoot ratios compared to their historical counterparts (Waines & Ehdaie, 2007; Y. H. Zhu et al., 2019). Recognizing the crucial role of roots in plant establishment and performance, contemporary research is increasingly aiming to improve yields through root system optimization, a movement often referred to as the "second green revolution" (White et al., 2013). This paradigm shifts towards a more holistic understanding of plant morphology, which includes both shoots and roots, marks a significant advancement in crop breeding. It heralds a new era that promises increased agricultural productivity and sustainability.

2.2.2 Impact of Water Supply on Root Growth

2.2.2.1 Hydrotropic Response and Root Growth

Plants have developed sophisticated tropic responses, such as hydrotropism, to navigate soil layers and adapt to varying environmental conditions variations in water potential gradients and the influence of gravity (Wyatt & Kiss, 2013). Terrestrial plant roots are equipped with numerous sensors in the root cap for detecting stimuli including gravity, humidity, light, mechanical pressure, temperature, and oxygen levels. While root growth direction is primarily governed by gravity, a constant factor, other fluctuating factors, like water potential gradients, can significantly alter this directional growth (Cassab et al., 2013). In areas with lower precipitation, roots exhibit a stronger hydrotropic response, growing towards deeper soil layers where water is more readily available. This is due to gravitational infiltration and reduced evaporation at these depths. Such an adaptive mechanism is vital for plants to efficiently acquire water, particularly in environments with variable moisture conditions (Takahashi et al., 2003). Furthermore, studies, such as those by Hernandez-Espinoza & Barrios-Masias (2020), indicate that under drought conditions, young roots tend to mature closer to the root tip and the soil layer with the highest root density becomes more critical for managing water stress than the maximum rooting depth. This highlights the importance of understanding these mechanisms to optimize yield and water management in stress conditions, highlighting that adaptation can vary within species.

2.2.2.2 Impact of Water Scarcity on Crop Yields and Root Architecture

Water scarcity in the root zone is a critical factor in reducing agricultural yields and poses a significant challenge to global food security, a situation further aggravated by climate change (Salekdeh et al., 2009). The impacts of climate change, including prolonged droughts increased rates of evapotranspiration, and more intense precipitation events, exacerbate this water scarcity and have profound effects on agriculture (Wang et al., 2017). These climatic changes negatively impact essential plant physiological processes such as photosynthesis, growth, nutrient assimilation, and osmoregulation, leading to diminished crop yield and quality (McDowell et al., 2011). In response to these challenges, the architecture of root systems becomes crucial for plant adaptation, especially under conditions of water stress. Drought conditions trigger notable changes in root morphology and function, essential for maintaining plant health and productivity.

Plants adapt by modifying their root systems, which includes changes in root depth, distribution, and density, to improve water and nutrient acquisition from deeper soil layers (Cassab et al., 2013). These adaptations, influenced by both genetic and environmental factors, are critical for root growth. For example, research on crops like rice has identified specific root traits, such as shallower root angles and increased lateral root proliferation, that are genetically associated with improved water and nutrient acquisition under drought conditions (Fry et al., 2018; Guimarães et al., 2020). Other studies have indicated that root depth was reported to increase upon exposure to water limitation (Vadez et al., 2013). Additionally, water scarcity affects root exudation patterns and interactions with soil microbiota, influencing nutrient availability and uptake. Understanding these root-centric responses and the broader implications of Root System Architecture is vital for developing targeted agricultural strategies to mitigate the adverse effects of water scarcity (Rogers & Benfey, 2015).

2.2.2.3 Tomato Plants' Response to Water Stress

Tomato plants, known for their high-water demand, are particularly sensitive to fluctuations in water availability, which significantly affects their yield and quality (Patanè et al., 2011). Many studies have focused on how tomato plant roots respond to water stress, revealing that under water-deficient conditions, these plants adapt by developing deeper roots to reach regions of higher soil moisture (Takahashi et al., 2003). This adaptive behaviour allows the plants to maintain a higher tissue water potential, a crucial strategy for survival (C. Singh et al., 2017). The literature also indicates that while tomato plants are capable of surviving prolonged periods of low soil moisture, severe water stress can have a significant impact on their biomass and yield (Katerji et al., 2013). A larger root system, which is a typical response to drought stress, enhances the efficacy of water uptake (Lee et al., 2016). Further studies have revealed that irrigation practices, specifically the frequency and rate of watering, influence root system development. Less frequent and lower rates of irrigation encourage the growth of a deeper root system (Marouelli & Silva, 2007). These adaptive responses, including structural changes in root morphology and reallocation of resources from shoots to roots, are essential for plant survival under water stress conditions.

2.2.3 Non-destructive Methods for Root Monitoring

Effective phenotyping techniques are essential for understanding root systems, which play a crucial role in plant performance (Zhu et al., 2011). Laboratory-based methodologies have been instrumental in the development of field applications, which integrate automation with advanced imaging and image processing techniques. Image analysis has emerged as a primary method for fast and reliable root phenotyping, utilizing software such as EZ-Rhizo (Lobet et al., 2011), Root

System Analyzer (Clark et al., 2013), and WinRhizo (Abràmoff et al., 2004). To facilitate root observation, these systems often use soil-less media. Methods include growing plants in paper rolls (Le Marié et al., 2014; Lopes & Reynolds, 2010), gels (Downie et al., 2012), or hydroponic systems with transparent Plexiglas (Ma et al., 2019). High-resolution cameras and scanners are employed to capture detailed images of the roots. Software like WinRhizo (Pfeifer et al., 2014), Image J (Pound et al., 2013) and Root System Analyzer (Clark et al., 2013)is then used to analyze root system architecture, providing valuable insights into root traits.

The transition to using soil-based mediums for studying root systems and their interactions with soil presents several challenges. In an effort to replicate natural growth conditions, researchers have employed soil-filled rhizotrons or columns. In this context, X-ray Computed Tomography (CT) has emerged as a pivotal technology, offering unparalleled visualization of three-dimensional root configurations in these simulated environments. This technological advancement facilitates a more in-depth understanding of root systems in undisturbed soil. The process involves excavation, washing, and detailed imaging, followed by analysis using specialized software (Herrera et al., 2007; Perkons et al., 2014). The utility of CT scanning lies particularly in its ability to produce detailed, three-dimensional images of root systems. This allows for precise examination and analysis of root architecture (Grift et al., 2011; Mairhofer et al., 2013). Despite these technological advances, applying X-ray Computed Tomography in field conditions faces significant limitations. These include high operational costs and a general restriction to small-scale studies in controlled environments, limiting its widespread use in field-based research. As a result, researchers often resort to more conventional methodologies, such as the trench profile technique or the soil core method, for investigating root density and structure in actual field conditions. While CT offers detailed insights into soil-root interactions and root system architecture (RSA), surpassing other destructive methods, it is important to recognize the inherent limitations that accompany its use, particularly in terms of scale and cost (Thompson et al., 2016).

Shovelomics, a popular method for analyzing root systems in field research, involves the careful excavation of soil to isolate a single plant, leaving its root system exposed on the soil surface. This process is followed by a gentle cleansing of the roots, enabling the examination of various root characteristics, including density and angles. The evaluation of fundamental root traits, such as dimensions, structure, and branching, employs various techniques ranging from basic counting methods to advanced imaging and specialized image analysis software (Bucksch et al., 2014; Trachsel et al., 2011). While shovelomics is effective for certain aspects of root system analysis, it does have limitations. One significant drawback is the potential for root damage during the excavation process, which can alter the natural structure and orientation of the roots. This disturbance may lead to inaccurate measurements or misinterpretation of the root system's natural state. Additionally, the technique may not capture the entire root system, particularly deeper roots, thus providing a partial view that might not fully represent the plant's root architecture (De Bauw et al., 2020; Slack et al., 2018).

The use of electrical capacitance measurement for assessing total root mass has become increasingly prominent, particularly in studying soil-root interactions. This technique involves applying an electrical current to the plant via electrodes in the rooting medium and then measuring the plant’s response (Dietrich et al., 2013). Known for its speed and cost-effectiveness, this method presents an innovative way to monitor root biomass (Dietrich et al., 2012). However, there are concerns about its precision, especially in analyzing detailed root architecture. Additionally, the reliability of this technique is subject to debate. Factors such as soil moisture and texture can potentially affect the measurements, leading to questions about its effectiveness in providing a comprehensive analysis of root function (Singh et al., 2020).

Magnetic Resonance Imaging (MRI) is renowned for its superior resolution, which allows for clear differentiation between various root tissues and the surrounding soil matrix. However, MRI has its limitations (Metzner et al., 2015). The high costs associated with this technology, along with its sensitivity to metallic elements in soil that can cause image distortion, pose significant challenges to its widespread use in root system analysis (Van Dusschoten et al., 2016).

In contrast, Ground-Penetrating Radar (GPR) provides a more practical option for in situ studies of larger root systems. GPR is particularly valuable for its ability to offer insights into the spatial distribution of roots (Lantini et al., 2018). Despite this advantage, GPR faces limitations due to its relatively lower resolution, especially in dense soil conditions. Additionally, GPR cannot distinguish between living and dead root structures, which can be a critical drawback in certain studies (Zhang et al., 2019).

Near-Infrared Spectroscopy (NIRS) represents another non-invasive technique, well-regarded for its quick and non-destructive assessment of root traits. The primary advantage of NIRS is its effectiveness in determining the composition of root systems, providing essential data on their chemical and biological properties. This technique has proven particularly useful in agriculture and horticulture for evaluating the quality of fruits and vegetables, owing to its ability to measure essential attributes like soluble solids content, dry matter, and acidity (Luypaert et al., 2007; Nicolaï et al., 2007). However, its capability in conducting detailed architectural analysis of root systems is somewhat limited. This is partly due to the technique's reliance on surface-level measurements, which may not fully capture the complex structural characteristics of roots (Picon-Cochard et al., 2009).

Thermal Imaging is a technique that capitalizes on the detection of root activity by measuring temperature variances in soil. It stands out for its non-invasive nature and its suitability for large-scale monitoring, making it an advantageous tool in root system studies (Shi et al., 2021). However, a notable limitation of this method is its susceptibility to fluctuations in environmental conditions. Such variations can affect the precision of the temperature measurements, potentially compromising the accuracy of the results. Therefore, while Thermal Imaging offers significant benefits for non-invasive monitoring, careful consideration must be given to environmental factors that may impact its effectiveness (Costa et al., 2013).

The reviewed literature underscores the significant impact of water supply on root system development in processing tomato plants, affecting aspects like root length and density. While these studies provide comprehensive insights into root development, they also prompt questions regarding the effects on fruit quality, a crucial aspect for processing tomatoes. This situation necessitates the exploration of predictive modeling as a tool that can bridge the gap between root development and fruit quality assessment.

2.3 Predictive Modeling of Tomato Fruit Quality

Given the established importance of root health, it becomes imperative to understand how these below-ground traits translate into tangible fruit quality parameters. Here, predictive modeling, specifically using XGBoost and ANN, emerges as a powerful approach to correlate root development conditions with key quality metrics like Brix, Lycopene, and a/b ratio in tomato fruits.

2.3.1 Quality Parameters in Tomato Fruits

The Brix degree (°Brix), a critical measurement unit, is employed to ascertain the soluble solids content (SSC), a subset of total soluble solids (TSS) primarily representing sugar concentration in juice. This metric quantifies the sugar percentage in a solution, offering insights into the sweetness and overall flavor profile of the fruit. High °Brix values typically correspond to sweeter taste and significantly influence overall flavor intensity, which aligns with consumer preferences in both commercial and domestic tomato strains (Agius et al., 2018; Baldwin et al., 2008; Barickman et al., 2016). Furthermore, the Brix level in tomatoes serves as a comprehensive indicator of the fruit's palatability and market value. It not only marks sweetness but also indirectly indicates other quality attributes like acidity and aroma (Jaywant et al., 2022; Kuhn & Johnson, 2013). The Brix can be easily measured by a refractometer in Brix units, but estimating it solely based on maturity varies with the cultivar (Baltazar et al., 2008). This measurement is particularly valuable in the food industry for quality control, ensuring consistency in products like wines and sauces, where sugar content is critical for taste and preservation (Jaywant et al., 2022).

Lycopene, a significant carotenoid contributing to the characteristic red hue of ripe tomatoes, has been the focus of numerous studies due to its potent antioxidant properties and health benefits. As highlighted by research from Giovannucci, (1999), Jürkenbeck et al. (2020) and Rao et al. (2018), the intake of Lycopene is associated with reduced risks of chronic diseases such as cardiovascular diseases and certain cancers. This correlation is further supported by findings from Dorais et al. (2008) and Ilahy et al. (2018), emphasizing its role in nutritional value and potential health advantages. The concentration of Lycopene in tomatoes, intensifying during ripening, is not merely an indicator of fruit maturity and flavor but also a measure of its nutritional content, as detailed by Gascuel et al. (2017). Moreover, the variation in Lycopene content, as discussed by Helyes et al. (2007, 2012), is mainly attributed to factors like tomato variety, environmental conditions (temperature, light, and water supply), and the interaction between genetic makeup and environmental influences (GEI), as per Csambalik et al. (2019) and Bhandari et al. (2022). While the Lycopene content can be roughly estimated based on fruit colour (Kim et al., 2020; Petropoulos et al., 2020), accurate quantification necessitates more sophisticated laboratory analyses, (Deák et al., 2015; Goisser et al., 2019, 2020). This comprehensive understanding of Lycopene's role in fruit quality and its health implications highlights its importance in both consumer acceptance and nutritional science.

The colour and uniformity of tomato fruit are fundamental factors that consumers prioritize when assessing fruit quality. This visual attribute serves as the main determinant in tomato purchasing decisions, as the ever-evolving shade of tomatoes, transitioning from green to deep red or even yellow based on the cultivar, act as vital visual cues. These cues subsequently influence consumer selection, quality evaluations, and market dynamics (Adalid et al., 2010; Arias et al., 2000). Consumers often associate specific colours with superior taste, higher nutritional value, and freshness. In addition to these consumer perceptions, the colour of tomatoes is also indicative of their stage of ripening, nutritional content, and suitability for processing in various food products. In general, fruit colour can be measured by visual analysis or different instrumental methods such as colorimetry, spectrophotometry, or a computer vision system. Among these, computer vision systems are gaining prominence for their ability to provide rapid and non-destructive evaluation of fruit quality, which is pivotal for large-scale commercial applications (Palumbo et al., 2021). Furthermore, the Chromaticity Ratio (a/b ratio) derived from colorimetric data provides a quantifiable measure of the tomato’s colour balance, offering an objective method to assess the shift in hue as tomatoes ripen, which is crucial for quality control and breeding programs (Thole et al., 2020). The precision of this method allows for a more nuanced understanding of the ripening process, potentially leading to improved storage and transportation strategies that can enhance the shelf life and overall quality of tomatoes.

2.3.2 Use of Machine Learning in Agriculture

A central objective in agriculture is to decrease production costs without compromising yield or quality (Chlingaryan et al., 2018). Advancements in computer science have popularized Machine Learning (ML) techniques, which utilize features extracted from these datasets (You et al., 2017). Particularly in crop management for yield prediction and quality assessment, the integration of ML has shown promising results. These advancements underscore the evolving landscape of agricultural technology, where ML emerges as a crucial tool in addressing the complex challenges of modern farming. The progression in ML algorithms, coupled with advanced data acquisition techniques such as remote sensing and UAVs, has opened new avenues for precision agriculture, specifically in fruit quality prediction. This convergence of ML and agriculture is demonstrating its potential in enhancing productivity, optimizing resource utilization, and ensuring sustainable farming practices, as highlighted in studies by Behmann et al. (2015) and Mehra et al. (2016). Such ML-driven insights are poised to revolutionize farming practices, making them more efficient and sustainable.

Yield prediction, a cornerstone of precision agriculture, leverages ML to enhance yield mapping, estimation, and crop management strategies. A diverse range of studies exemplify the application of ML in this field. For instance, Ramos et al. (2017) showcased a low-cost, non-destructive ML approach for classifying coffee fruits into various maturation stages, aiming to enhance the economic returns for coffee growers. Similarly, Amatya et al. (2016) introduced a machine vision system, driven by ML, to automate cherry harvesting, particularly focusing on the detection of occluded branches, thereby reducing the reliance on manual labor. Extending ML's use in yield prediction, Sengupta & Lee, (2014) introduced an avant-garde system designed for the precocious detection of immature green citrus within orchards. This system provides essential information specific to yields, which is of significant utility to citrus growers.

In the context of grassland management, Ali et al. (2017) employed an Artificial Neural Network (ANN)-based model to estimate biomass by harnessing remote sensing data. The accuracy of yield predictions experienced further enhancement by (Pantazi et al., 2016), through the utilization of satellite imagery coupled with soil data to forecast wheat yields with heightened precision. In the sector of vegetable crop management, Senthilnath et al., (2016) employed a novel methodology which integrates electromagnetic (EM) and RGB imagery, captured via Unmanned Aerial Vehicles (UAVs), for the effective detection of tomato crops. Pertaining to rice cultivation, Su et al. (2017) utilized Support Vector Machines (SVM) in conjunction with geographic data obtained from meteorological stations, thereby enabling the prediction of the developmental stages of rice crops in China.

Kung et al. (2016) developed a generalized model using Evolutionary Neural Networks (ENNs) for agricultural yield prediction, assisting Taiwanese farmers in market supply and demand management. Moreover, Darwin et al. (2021) reviewed the application of deep learning in smart farming, particularly in crop yield prediction, achieving an impressive average accuracy of 92.51%. This study demonstrates the efficacy of these models across various crops, including fruits. Similarly, Umrani et al. (2021) focused on the use of deep learning in smart agriculture, highlighting robust fruits counting and yield prediction. Their work illustrates the integration of machine learning with Internet of Things (IoT) and robotics in modern agricultural practices.

These studies collectively underscore the significant role of ML and deep learning in advancing smart farming techniques, offering innovative solutions for improved agricultural productivity and sustainability.

2.3.3 Comparative Analysis of Models

The literature reflects a growing trend towards employing various machine learning models in agriculture, with a focus on improving accuracy, efficiency, and applicability. A study by Mao et al. (2019) conducted a comparative analysis of five distinct machine learning regression algorithms - Artificial Neural Networks (ANN), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Random Forest (RF), and Gradient Boosting Regression Trees (GBRT). This research, which focused on Cotton Leaf Area Index retrieval, identified the GBRT model as the most precise. Additionally, it recognized the computational efficiency of the SVR, emphasizing its potential for real-time operational applications in crop management.

In the domain of smart farming, Venkatesan et al. (2022) evaluated a predictive model for peak energy usage, employing various machine learning algorithms including ANN, SVR, and Random Forest. Their findings indicated that the Random Forest-based model demonstrated superior accuracy, reaching approximately 92%. This result underscores the model's potential utility in smart agricultural practices. Radočaj et al. (2021) presented a novel machine learning approach for assessing cropland suitability, demonstrating enhanced computational efficiency and accuracy compared to traditional methodologies, where the Random Forest algorithm outperformed the Support Vector Machine in prediction accuracy. Paola Carrieri et al. (2019)

introduced an innovative machine learning workflow for rapid phenotype prediction from whole shotgun metagenomes, emphasizing the efficiency and accuracy of this method in real data classification tasks, relevant for soil health and fertility assessment in precision agriculture. Furthermore, a comprehensive review by Yadav et al. (2020) encapsulated the multitude of machine learning applications within the agricultural sector. This review critically discusses various computational methods, including neural networks and k-means clustering, highlighting their precision and analytical capabilities. This comprehensive overview significantly contributes to the understanding of the breadth and depth of machine learning techniques applicable in agricultural contexts.

These studies and more suggest that machine learning model are highly effective for various agricultural applications, ranging from crop classification to energy management in smart farms. Two of these techniques have prominently emerged as viable contenders particulary for agricultural data processing: eXtreme Gradient Boosting (XGBoost) and Artificial Neural Networks (ANNs). XGBoost, a highly efficient gradient boosting framework, excels in both classification and regression tasks (Chen & Guestrin, 2016; Ge et al., 2022). It stands out as an advanced gradient boosting decision tree algorithm. Recognized for top performance, XGBoost is an open-source boosted tree toolkit, appreciated for its ability to combine multiple tree models into a powerful learning framework. Its proficiency in handling large-dimensional datasets, especially in gene expression research, highlights its significance (Friedman, 2001; Haq Chowdhury et al., 2023; Song et al., 2020; P. Zhang et al., 2022). Concurrently, ANNs have gained widespread recognition in the deep learning domain for their ability to process high-dimensional data and extract meaningful features (Bishop, 1995; You et al., 2017). These features offer transformative insights, potentially reshaping agricultural practices towards sustainability. Particularly in the field of remote sensing, ANNs are routinely employed to forecast vegetation parameters and crop yields (Farifteh et al., 2007; Kaul et al., 2005; Kuwata & Shibasaki, 2015). However, the deployment of ANNs presents certain challenges, such as optimizing the number and size of hidden layers, determining the appropriate learning rate, the need for expansive training datasets, and confronting issues like overfitting.

The application of machine learning models has demonstrated promising potential in predicting the quality of tomato fruits, based on various environmental and cultivation factors. These predictive models, while providing valuable insights into the environmental and management factors impacting fruit quality, also suggest the potential influence of genetic diversity, a variable remains to be thoroughly investigated in relation to its interaction with environmental factors.

2.4 Evaluation of Tomato Plant Genetic Resources

Building upon the insights gained from predictive modeling, the role of genetic diversity in tomato plants becomes a focal point. This section delves into how different genetic makeups of processing tomato plants respond to environmental factors, impacting not only root development but also the attributes predicted by our models, such as Brix and Lycopene content.

2.4.1 Genetic Diversity in Tomato Plants

Genetic diversity is essential in crop plants, serving as a basis for breeding programs and adaptation to varying environmental conditions. In the case of tomato plants, this diversity is critical for ensuring sustainable production and food security. The genetic variability within tomato plants not only facilitates the improvement of qualitative and quantitative traits through selective breeding but also provides a buffer against the unpredictability of climate change and potential pest outbreaks (Dixit & Pandey, 2017; Udriște et al., 2022).

Tomato cultivation, through extensive breeding and domestication processes, has led to notable improvements in traits such as yield and size. However, these advancements have inadvertently caused a considerable decline in the genetic diversity of tomato plants (Kulus, 2018). This reduction in genetic variation presents several challenges: it heightens the susceptibility of tomato plants to diseases and pests, diminishes their ability to adapt to evolving climatic conditions, and imposes constraints on future breeding endeavors aimed at enhancing other desirable attributes (Jaiswal et al., 2020). Consequently, the long-term sustainability and resilience of tomato crops are potentially compromised, underscoring the need for strategies that balance agricultural productivity with genetic conservation (Gliessman, 2022).

Studies have shown that wild tomato species and heirloom varieties possess a greater range of genetic variation compared to commercial cultivars. This variation includes traits for disease resistance, tolerance to environmental stresses, and nutritional quality. By tapping into this genetic pool, breeders can introduce these beneficial traits into commercial tomato varieties, enhancing their resilience and adaptability (Aflitos et al., 2014).

The role of genetic diversity in adapting to different environments is particularly noteworthy. Tomatoes face challenges like fluctuating temperatures, varying soil conditions, and water availability. Genetic diversity enables these plants to withstand such environmental stresses, ensuring stable yields. For instance, genes from wild varieties have been used to develop tomato plants that can tolerate drought and saline conditions, vital traits as climate change impacts become more pronounced (Fischer et al., 2011).

Modern biotechnological tools, such as genomics and CRISPR, are revolutionizing how this diversity is understood and utilized. These tools allow for more precise identification of beneficial genes and traits within the tomato genome (Corbin et al., 2020; Zsögön et al., 2018). Consequently, they enable the development of new tomato varieties with desired characteristics more efficiently and accurately than traditional breeding techniques. Furthermore, they open new avenues for addressing challenges in agriculture, such as improving crop resilience to environmental stresses and enhancing nutritional profiles, thereby contributing substantially to global food security and sustainable agricultural practices (D. P. Roberts & Mattoo, 2019).

Despite these advancements, challenges remain. The narrow genetic base of most commercial tomato cultivars limits the scope of improvement. Furthermore, the reliance on a few commercially dominant varieties exacerbates the loss of genetic diversity. There is a pressing need for conservation efforts, both in situ and ex situ, to preserve the wide array of genetic resources available in tomato plants. This includes not only the wild relatives but also landraces and heirloom varieties that are repositories of unique and potentially valuable genetic traits (Gao et al., 2019).

Future research directions include exploring the untapped genetic potential in wild and traditional tomato varieties, understanding the genetic basis of traits relevant to climate resilience and nutritional quality, and developing sustainable breeding practices that maintain or increase genetic diversity. These efforts are crucial for the continued success and improvement of tomato cultivation in the face of global environmental challenges.

2.4.2 GGE Biplot Analysis

The concept of Genotype × Environment Interaction (GEI) holds significant importance in agricultural research, especially in the context of crop adaptation. This phenomenon embodies the response of different genotypes under various environmental conditions, a critical aspect in the field of agronomy (Kwabena Osei et al., 2019). To comprehensively assess GEI, multilocation trials have been identified as indispensable. These trials involve testing crop varieties across diverse environmental settings and seasons, with a primary aim to evaluate the adaptability and stability of genetic materials. This approach is not only essential for understanding the stability of phenotypic traits but also serves as a basis to identify patterns in genotype responses across varied environments (Allard & Bradshaw, 1964; Crossa, 1990; Happ et al., 2021).

Within this framework, the Genotype main effect plus genotype × environment interaction (GGE) biplot analysis, as introduced by Yan, (2001), emerges as a pivotal tool. This statistical method is highly versatile, allowing for a multi-perspective evaluation of gene bank accessions and breeding lines. The GGE biplot analysis plays a crucial role in the efficient selection and development of crop varieties, contributing significantly to the advancement of agricultural research. This methodology has gained prominence for its ability to aid breeders and agronomists in developing crop varieties that are not only high-yielding but also resilient to environmental changes (Fayeun et al., 2018).

The GGE biplot methodology is an analytical tool used for visualizing the results of Site Regression (SREG) analysis in multi-environment trial (MET) data. This method combines the concept of a biplot, as introduced by Gabriel in 1971, with the GGE (Genotype main effect plus Genotype by Environment interaction) framework developed by Yan et al. (2000). It adeptly displays the two primary factors affecting genotype evaluation in MET data: the genotype effect (G) and the genotype-environment interaction effect (GE). The utility of the GGE biplot is particularly evident in several core areas of agricultural research. It is effective for mega-environment analysis, demonstrated by the 'Which-won-where' pattern, and for genotype evaluation, which involves a juxtaposition of mean performance and stability. Moreover, it is valuable for evaluating test environments by providing a comparative assessment of their discriminating power and representativeness. The flexibility of the GGE biplot method and its capacity to handle various data types with a two-way structure have contributed to its widespread adoption in agricultural research (Amira et al., 2013; Atnaf, 2013; Yan et al., 2007).

In the specific context of tomato crops, these analytical tools and methods provide a foundation for understanding the complex interactions between genetic resources and environmental factors. This understanding is crucial for developing tomato varieties that can thrive in diverse climates and conditions, ensuring food security and sustainability in agricultural practices.

2.4.3 GGE and Influence of Environment on Genetic Expression

The exploration of the influence of environmental factors on the genetic expression of tomato plants, especially in relation to fruit quality, has been an area of extensive research. Many studies provide valuable insights into how different environmental conditions impact the quality attributes of tomato fruits. Nguyen et al. (2014) made a significant discovery in understanding the role of chloroplast distribution in tomato fruits. They found that the distribution and abundance of chloroplasts, which are regulated by GOLDEN2-LIKE (GLK) transcription factors, are crucial in determining the nutritional quality of ripe tomatoes, including their sugar content. Additionally, the interaction between genetic and environmental cues in the ripening process of tomato fruits was highlighted by Moore et al. (2002). They emphasized the influence of environmental factors on ethylene biosynthesis and cell wall metabolism, which are critical processes in the maturation of tomato fruits. Similarly, D’Esposito et al. (2017) demonstrated that the genetic makeup of different tomato genotypes dictates their specific responses, at both transcriptomic and metabolomic levels, to environmental conditions, thereby influencing fruit quality. The impact of light quality on sugar metabolism in tomato fruit was another area of interest explored by Dong et al. (2019). Their research showed that varying combinations of light quality significantly affect the stages of fruit development and sugar content in tomatoes. Moya et al. (2017) explored a different aspect of environmental influence by studying how the management of nutrient solutions, particularly electrical conductivity, impacts the dietary and organoleptic qualities of soilless culture tomatoes. Their findings revealed significant changes in fruit quality parameters. Furthermore, the work of Prudent et al. (2010) underscored the genotype-dependent response of tomato fruit to variations in carbon availability, affecting fruit growth and composition.

When it came to the GGE biplot analysis to evaluate genetic resources of tomatoes, the technique has proven instrumental in understanding genotype performance across diverse environments. Joshi et al. (2011) employed the GGE biplot technique to evaluate five F1 tomato hybrids over six years in North Carolina. They highlighted substantial genotype * year interactions, marking the technique's effectiveness in identifying high-yielding and stable genotypes across different years. They found 'Mountain Fresh' to be an ideal hybrid, showcasing the benefits of GGE biplot analysis in crop improvement efforts. Frutos et al. (2014) presented an interactive computational implementation of classical and GGE biplot methods in R language. Their work, encapsulated in the GGEBiplotGUI package, enables a comprehensive visual analysis of multi-environment trials, which is crucial for plant breeders and geneticists studying genotype and environment interactions. Lastly, the study by Divéky-Ertsey et al. (2022) on the phytochemical evaluation of tomato landraces under different production systems employed GGE biplot analysis to understand how open-field versus protected production affects the phytonutrient content of tomato plant genetic resources. Their findings revealed that the phytonutrient content of certain genetic resources is more influenced by microclimatic conditions rather than location, thus enriching the nutritional datasets of tomato landraces and supporting their conservation and utilization in breeding programs.

GGE biplot analysis is a robust tool in crop improvement, particularly in evaluating genetic resources in crops like tomatoes. It aids in identifying high-yielding and stable genotypes across different environments and understanding the influence of environmental factors on crop characteristics, thereby guiding breeding programs towards developing better crop varieties.

3 MATERIALS AND METHODS

This section delineates the comprehensive methodologies employed to investigate the development of roots under different water supply levels, to predict the quality of tomato fruits utilizing machine learning algorithms, and to assess the genetic diversity within populations of processing tomato plants. The selection of these methodologies was predicated on their robustness, precision, and ability to provide insightful data across different study dimensions.

3.1 Part 1: Root Development Monitoring under Different Water Supply Levels

This section examines the effects of different irrigation treatments on root development and physiological responses of tomato plants. By understanding how root systems adapt to varying water supply levels, we can gain insights into the resilience and efficiency of tomato plants in different environmental conditions.

3.1.1 Plant Material and Experimental Set-Up

The experiment was conducted at the Horticultural Experimental Farm of the Hungarian University of Agriculture and Life Sciences (GPS: 47°34′51.6″ N 19°22′39.0″ E), in Gödöllő, Hungary. The field features a loamy soil, characterized by 47.5% silt, 41% sand, and 11.5% clay, with 1.6% humus content in the upper 0–30 cm layer. Processing tomato plants generally have a determined growth type and thick fruit skin. A processing tomato hybrid, H1015, was used in the experiment, which is a variety with a determinate growth type that is widely used by growers in our region. It has good adaptability to the growing conditions, and its maturity is favorable under our climate. This variety has been used in irrigation deficit experiments in other studies as well (Arbex de Castro Vilas Boas et al., 2017).

The row spacing was 140 cm, and the plant spacing was 20 cm, resulting in a plant density of 3.57 plants per square meter. Seedlings were transplanted on 14 May 2020 and 15 May 2021. The fertilization schedule was adjusted to the phenological phase of the tomato plants. Specific quantities of N, P_2O_5, and K_2O were administered in each phase based on the plant’s needs. The total distribution for the growing season was 129 kg ha^{-1} of N, 89 kg ha^{-1} of P_2O_5, and 317 kg ha^{-1} of K_2O in 2020 148 kg ha^{-1} of N, 67 kg ha^{-1} of P_2O_5, and 121 kg ha^{-1} of K_2O in 2021. Plant protection interventions were applied when they were needed, according to a local expert, uniformly to all treatments.

In 2020, fertilization was carried out using a solid fertilizer, whereas in 2021, a solution was applied via drip tape with 10 cm emitter spacing in all treatments. The fertigation circuit was separate from the clean irrigation water supply, which ensured equal and uniform fertilization across all treatments.

Clean irrigation water was supplied through a drip irrigation system. Three distinct treatments were implemented, supplying 100% of crop evapotranspiration (I100), 50% of I100 (I50), and a non-irrigated control (K). Each treatment was replicated thrice to ensure the robustness of our data. As depicted in Figure 1, the experimental design ensured an even application of the different treatments. The I100 treatment received 241mmof water, the I50 received 201 mm, and the non-irrigated K received 159 mm during the monitoring period of the first growing season. The crop evapotranspiration was 231 mm, averaging 4.2 mm per day in 2020. In 2021, the total water supplies for the I100, I50, and K treatments were 206 mm, 129 mm, and 51 mm, respectively, with 180 mm evapotranspiration, averaging 5 mm per day during the monitoring period. These data indicate that the 2021 observations occurred during a drier period.

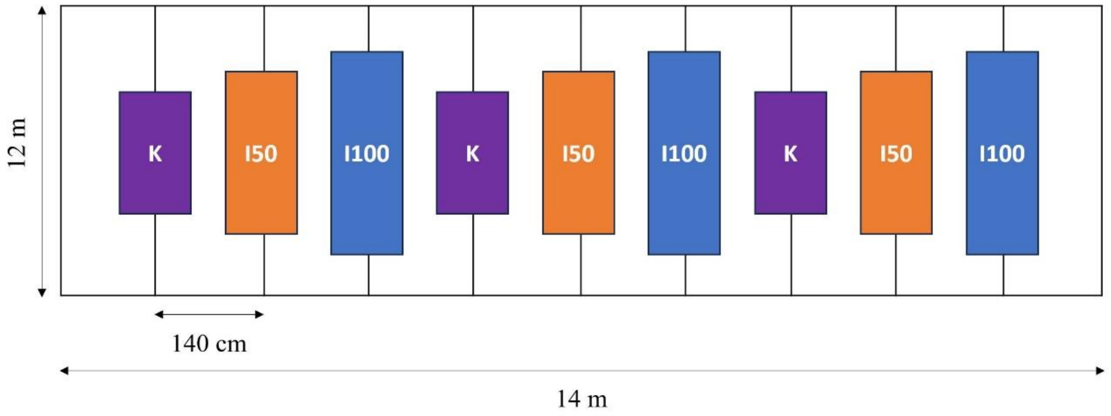


Figure 1. The experimental design for both experimental

Different types of drip tapes were employed to supply varying water quantities. A 10 cm emitter spacing tape delivering 10.6 L per hour per meter was used for the I100 treatment, whereas a 15 cm emitter spacing tape providing 5.3 L per hour per meter was used for the I50 treatment (Irritec S.p.A., Rocca di Caprileone, Italy). The positioning of the drip lines, monitoring tubes, and tomato plants remained constant across the treatments. The FAO Penman–Monteith method was used, utilizing the AquaCrop 6.1 software to calculate crop water demand, incorporating meteorological data from a nearby station (Takács et al., 2021). The irrigation treatment in the 2020 growing season began on 8 June but was followed up only on 29 June and concluded on 3 August. In 2021, the irrigation treatment started on 3 June and ended on 2 August.

3.1.2 Image Acquisition

Images of the roots were taken using a CI-600 In-Situ Root Imager (CID Bio-Science Inc., Camas, WA, USA) (Figure 2). This device allows for non-destructive imaging of living roots. The working principle of the CI-600 involves a cylindrical scanner that rotates within a transparent tube inserted into the soil. The scanner head, which contains a digital camera and lighting system, captures images of the root system through the transparent wall of the tube as it rotates 360 degrees. The captured images are then stitched together to create a comprehensive view of the root environment. The standard tube measures 105 cm in length, with an inner diameter of 6.35 cm and an outer diameter of 7 cm. The scanning head is 34.3 cm long, with a diameter of 6.35 cm. The resultant images measure 21.6 × 19.6 cm. Images were captured in 300 dpi resolution at three different depths: 10–30 cm, 30–50 cm, and 50–70 cm. On each monitoring date, we took root pictures at each layer of 3 plants in each irrigation treatment. We calibrated the scanner with a white calibration tube at the start of each measurement event. CI-600 Root Scanner software was employed to conduct the scans. Monitoring was carried out from June 25 to August 18, taking images once a week in 2020, and from June 9 to July 14 in 2021. Root monitoring was chiefly tied to this period when deficit irrigation treatments were continuously applied.

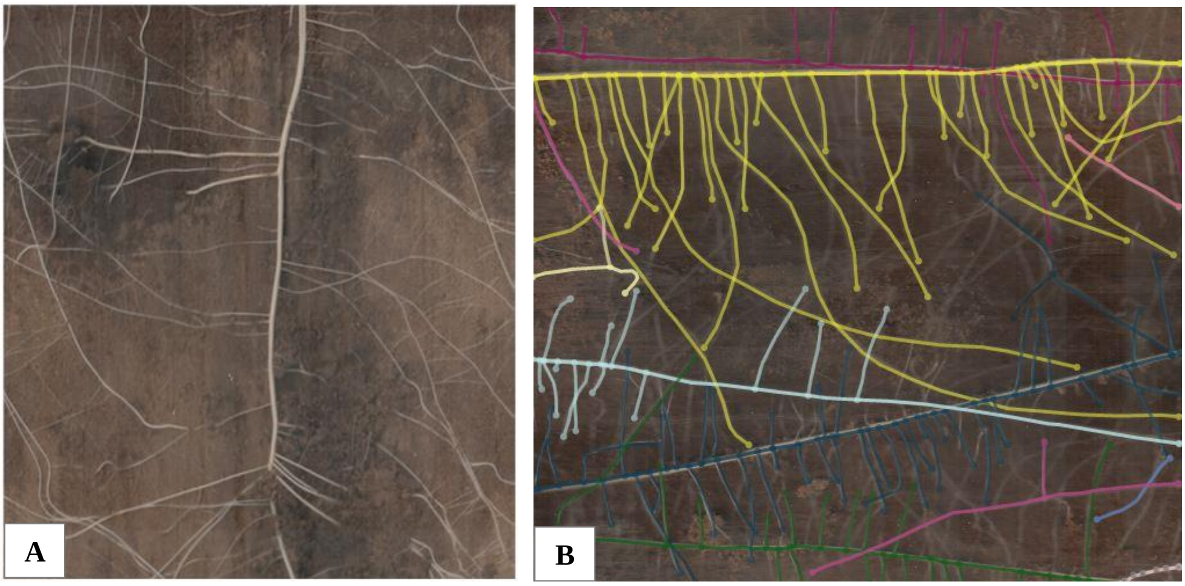


Figure 2. Image taken by CI-600 in situ root imager (A) and an image taken during the analysis process (B).

The scanner tubes were installed in three replications per treatment immediately following the transplantation of tomato seedlings. These tubes were set up next to randomly selected plants within the treatments. This installation method did not cause harm to the roots, as the tubes were in place prior to root development in the field.

3.1.3 Image Processing

Each loaded image was defined by the tube it was captured in, the depth in the soil, and the date of acquisition. The brightness, contrast, and gamma of the images were adjusted to achieve the highest clarity. Root distribution mapping was performed manually by tracing the cursor along the length of the root adding points, and branches were connected to the parent root. Data from RootSnap 1.4 software, concerning tubes, windows, or sessions, were exported to a spreadsheet.

3.1.4 Relative Chlorophyll Content and Photosynthetic Activity

Measurements were conducted on randomly selected plants within each treatment at around 12:00 on each measurement date. The SPAD index was measured using a SPAD 502 chlorophyll meter (Konica Minolta, Warrington, UK). The SPAD meter operates by emitting light at two wavelengths (650 nm and 940 nm) and measuring the transmitted light through the leaf. The difference in light transmission, affected by chlorophyll absorption, is used to calculate the SPAD index, estimating leaf chlorophyll content. A PAM 2500 fluorometer device (Heinz Walz GmgH, Effeltrich, Germany) was used to measure chlorophyll fluorescence. Data were acquired from the device with PamWin-4 4.01 software. DLC-8 leaf clips were applied to leaves 30 min before measurement to provide dark acclimation before taking the fluorescence measurement, in which Fv/Fm values were recorded. In total, 16 measurements per treatment were taken, which included 4 measurements per treatment in each repeated block. All measurements were carried out non-destructively on healthy, fully developed leaves.

3.2 Part 2: A Comparative Analysis of XGBoost and Neural Network Models for Predicting Tomato Fruit Quality

This section focuses on the application of machine learning models to predict tomato fruit quality traits. By comparing the performance of XGBoost and Neural Network models, we aim to determine the most effective approach for accurately prediciting key quality parameters under various environmental conditions.

3.2.1 Dataset Description

In this study, a comprehensive dataset was utilized encompassing physicochemical characteristics and environmental factors across a diverse selection of tomato cultivars over five consecutive growing seasons from 2017 to 2021. The dataset included observations of 48 cultivars and 28 locations (Loc) within Hungary.

The selection and distribution of cultivars varied annually, with 25 cultivars at 7 locations in 2017, 22 cultivars at 18 locations in 2018, 27 cultivars at 19 locations in 2019, 27 cultivars at 18 locations in 2020, and 26 cultivars at 15 locations in 2021. This variability provided a rich dataset for analysing tomato quality traits under diverse environmental conditions. For each cultivar–location combination within a given year, multiple measurements were conducted on a random sample selection after harvesting on the same day to assess the quality traits of the tomatoes, ensuring the robustness of the dataset. A total of 28,474 individual measurements were recorded for each of the three main variables of interest which were the ◦Brix (denoting water-soluble solids content (Brix)), lycopene concentration, and fruit colour (quantified through the a/b ratio as a measure of colour balance in the Hunter Lab colour space).

To understand the impact of meteorological factors on tomato cultivation, meticulous records were conducted over growing seasons characterized by various climatic factors. These records were sourced from the Operational Drought and Water Scarcity Management System in Hungary (General Directorate of Water Management, Budapest, Hungary). This database provided a comprehensive overview of the conditions for each growing season, defined specifically as the period from 30 May to 30 August of each year, which is the favorable and usual growing period for tomatoes in Hungary mostly covering the period from intensive vegetative growth to harvest. The number of days with temperatures ideally between 21°C and 27°C (T21_27) was noted, as this range is optimal for tomato growth. Total precipitation (TotPrecip) during the growing season and the number of rainy days (RainDays) were recorded to understand moisture availability. Additionally, the average relative humidity (AvgRH) was monitored to assess the overall moisture content in the air. The count of days with relative humidity within the 40% to 70% range (RH40_70) was also tracked, being the ideal range for tomato cultivation. Furthermore, instances of high humidity were observed, specifically the days when the average daily relative humidity exceeded 90% (RH_90+), as such conditions could adversely affect plant health. Alongside these climatic factors, the soil type (SoilTyp) at each location was classified according to the USDA soil classification system.

3.2.2 Measurement of tomato quality traits

The physicochemical properties of tomatoes were assessed using state-of-the-art automated stations. Brix was measured by the SV01 from the Maselli Misure Quality Station (2020 Maselli Misure S.p.A, Parma, Italy), which first processed the tomatoes into juice followed by an automatic refractometric analysis to determine the sugar content, presented on a temperature-compensated scale with a range from 0 to 10 Brix and accuracy within ±0.15 Brix, adhering to the nD/Bx ICUMSA (1974) standard. Lycopene content was quantified via an automated spectrophotometric analysis, reporting concentration levels in mg/100 g with measurement limits of 0 to 80 mg/100 g, an accuracy up to 0.5 mg/100 g, and a repeatability ±0.25 mg/100 g. Additionally, fruit colour was assessed through spectrophotometric analysis measuring the colorimetric coordinates L, a, b, from which the chromaticity ratio (a/b ratio) was derived to evaluate the balance between red and yellow hues, with a repeatability for X, Y, Z coordinates less than 0.07, ensuring consistency in the colour assessment of the tomatoes.

3.2.3 Data Preprocessing

The dataset was undergoing several preprocessing steps to ensure data quality and facilitate exploratory analysis. The initial preprocessing involved the transformation of categorical attributes such as 'Loc', 'Cultivar', and 'SoilTyp'. Each of these attributes were transformed into one-hot encoded vectors to convert them into numeric representations suitable for machine learning algorithms (Goodfellow et al., 2016; Lecun et al., 2015). Then the dataset's integrity was assessed by quantifying missing entries within each column. Missing values within numerical columns were imputed using the respective column's mean, while those in categorical columns were replaced with the mode. This approach helped maintaining the original distribution of the data and minimize the distortion introduced by imputation. Post-cleanup, an in-depth exploration into the relationships between the different variables was conducted using a correlation matrix visualized on a heatmap, utilizing the seaborn library.

3.2.4 Machine learning models

3.2.4.1 XGBoost Model

The XGBoost (eXtreme Gradient Boosting) model, is known for its efficiency in handling missing values and evaluating feature importance, based on gradient-boosted decision trees. This model iteratively refines predictions by adding trees that minimize error (Chen & Guestrin, 2016). To augment the dataset for time series prediction, lag features for the 'Predicted Variable' (i.e., Brix, Lycopene, a/b ratio) column were engineered, considering lag values from the previous 1 to 3-time steps. A rolling mean (moving average) feature was computed for the 'Predicted variable' column, with a window of three time points to capture temporal patterns and to smoothen out short-term fluctuations (Box et al., 2015). The dataset was split into training and test subsets employing a 5-fold Time Series Split method, partitioning the dataset into five sequential time-based segments. Each segment is once utilized as the test set, while all previous segments form the training set. This approach enables iterative training and validation of the model on distinct portions of the dataset, thereby maintaining the integrity of temporal sequences and avoiding leakage of future information during model training (Roberts et al., 2017). Each feature subset underwent standardization using the StandardScaler method, ensuring zero mean and unit variance. The XGBoost Regression model was employed for the prediction task. The model's hyperparameters were optimized through grid search coupled with 3-fold cross-validation. The hyperparameter grid encompassed various combinations of 'n_estimators', 'max_depth', 'learning_rate', 'colsample_bytree', and 'gamma' to minimize the squared error. Once the optimal hyperparameters were identified, the model was trained on the entirety of the training dataset and subsequently evaluated on the test set. The performance was assessed using the R-squared value, root mean squared error (RMSE) (1), and magnitude relative error (MRE) (2) where:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \text{ and} \tag{1}$$

$$\text{MRE} = \frac{|y_i - \hat{y}_i|}{|y_i|} \tag{2}$$

n is the number of observations in the dataset,
y_i is the actual value for the i-th observation,
$\hat{y}_i$ is the predicted value for the i-th observation.

3.2.4.2 ANN Model

Artificial Neural Networks (ANNs), inspired by the human brain's neural network, excel in modeling complex non-linear data relationships (Goodfellow et al., 2016). For the ANN model, data was sorted chronologically based on the 'Year' column. To capture potential temporal patterns, lag features were generated for the 'Predicted variable' measurements spanning three previous time points. Additionally, a three-time point rolling average was computed to smoothen short-term fluctuations. The architecture of the model was determined through hyperparameter tuning, which included the number of neurons, dropout rates, and learning rates (Bergstra et al., 2012). The network featured two hidden layers with a variable number of units, dropout layers for regularization, and an output layer for predictions. A random search, complemented by early stopping based on validation loss to prevent overfitting, facilitated systematic hyperparameter exploration. The data was split into training and test subsets using a 5-fold Time Series Split method, partitioning the dataset into five sequential time-based segments, ensuring a chronological division and preventing future data leakage during the training process. Both training and test datasets were standardized to have zero mean and unit variance using StandardScaler. The trained ANN was then evaluated on the test set, with model performance evaluated using the R-squared value, root mean squared error (RMSE) and magnitude relative error (MRE).

3.2.5 Feature Importance Analysis with SHAP

The SHAP (SHapley Additive exPlanations) value analysis, developed by (Lundberg & Lee, 2017), was utilized to elucidate the impact of individual features on the predictions of both XGBoost and ANN models. SHAP values measure each feature's contribution to the prediction by assessing their marginal contribution across all possible feature combinations. For the XGBoost model, following optimization, we conducted a SHAP analysis on features including

'Loc', 'Cultivar', 'SoilTyp', 'AvgT', 'T21_27', 'TotPrecip', 'RainDays', 'AvgRH', 'RH40_70', and 'RH_90+'. The data was standardized using the StandardScaler method before computing the SHAP values for the training set, thus showcasing the average contribution of each feature (Lundberg et al., 2018). In the case of the ANN model, the training data was adapted to be compatible with the SHAP library, employing the GradientExplainer method to compute SHAP values for the same features. For both models, categorical features such as 'Loc', 'Cultivar', and 'SoilTyp' required aggregation to assess their collective importance. We generated SHAP summary plots to visualize the relative importance and effect of each feature. These plots employed a dot plot format, where the x-axis represented the magnitude of SHAP values and the y-axis the features. A dual-colour scheme was used, with red and blue indicating high and low feature values, respectively, illustrating the directional influence of each feature on the model predictions.

The observed differences in the SHAP graphs between the XGBoost and ANN models can be primarily attributed to their intrinsic architectural differences and the specific methods used for SHAP value calculation. The XGBoost model, operating within a gradient boosting framework and utilizing decision trees, facilitates a more straightforward computation of SHAP values by assessing the impact of each feature across an ensemble of trees. In contrast, the ANN model, comprising a complex network of neurons with non-linear activations, necessitates the use of approximation methods such as the shap.GradientExplainer, making the calculation of SHAP values more intricate. This complexity contributes to the variations observed in the visual representations of feature importances in the SHAP graphs for each model.

3.3 Part 3: Evaluation of Tomato Plant Genetic Resources for Brix and Lycopene in Different Environments

3.3.1 Plant Material

In our study, we evaluated six distinct commercial tomato varieties. each with unique characteristics in terms of maturity, disease resistance, and physical properties. The H1015 variety, developed by Heinz, is remarkable for its Extended Field Storage (EFS™) capability, disease resistance, and adaptability to a wide range of climatic conditions. Classified as a second early maturity variety, it thrives in both humid and arid environments. The N6416 hybrid variety is notable for its early maturity, resistance to Tomato Spotted Wilt Virus (TSW), and high total acidity, making it an excellent choice for industrial applications. This variety is also known for producing a homogeneous plant population. The Prestomech F1 variety is characterized by very early maturity, square/round fruits, and disease resistance. It's well-suited for both industrial and garden cultivation, thanks to its high sugar content and resistance to overripening. UG11227, a high-yielding variety from United Genetics Seeds Co., is utilized in industrial-scale agriculture. Meanwhile, UG812J, primarily used in industrial processing and agricultural research in Hungary, is notable for its resistance to mechanical injuries, enhancing its suitability for industrial processing. Lastly, the Ussar variety stands out for its mild, juicy flavor and distinctive appearance, making it a versatile choice for various culinary applications. These varieties were selected for their diverse and advantageous attributes to provide a comprehensive analysis of their performance in different environments.

3.3.2 Environments

In this study, the comparative performance of six tomato varieties was assessed in three distinct locations in Hungary, with experiments conducted at various intervals within a five-year period from 2017 to 2021. The selected locations—Szarvas, Mezöberény, and Kocsér—each offer unique environmental conditions conducive to evaluating the adaptability and performance of the tomato varieties. In Szarvas, experiments were conducted annually over the entire five-year span, utilizing meadow chernozem soil known for its rich organic matter and nutrient content. In Mezöberény, trials were carried out in 2018 and 2019 on casting meadow soil, recognized for its texture and fertility. Lastly, the experiment in Kocsér took place in 2020 on soil that ranges from sand to humus sand, a composition affecting water retention and nutrient availability. The choice of these diverse environments, each with its specific soil type, was instrumental in evaluating the resilience and productivity of the varieties under varying agro-climatic conditions. This multi-year, multi-location approach provided a comprehensive understanding of the genotypic performance and environmental interactions, crucial for the GGE biplot analysis.

3.3.3 Instrumental Measurements

The physicochemical properties of tomatoes were assessed using state-of-the-art automated stations. The Brix value, indicating sugar content, was obtained through the Maselli Misure Quality Station's SV01 device. This process involved transforming tomatoes into juice and then conducting an automated refractometric analysis. The results, adjusted for temperature variations, were displayed on a scale ranging from 0 to 10 Brix with an accuracy of ±0.15 Brix, conforming to the nD/Bx ICUMSA (1974) standards. Additionally, the lycopene concentration was determined using an automated spectrophotometric method, yielding readings in mg/100 g. The measurement range for this analysis was from 0 to 80 mg/100 g, with a precision of up to 0.5 mg/100 g and a repeatability margin of ±0.25 mg/100 g.

3.3.4 GGE biplot

This GGE biplot is constructed through plotting first the principal component (PC1) scores of the genotypes and the environments against their respective scores for the second principal component (PC2) that result from “singular value decomposition” (SVD) of environment-centered data using the following formula (Yan et al., 2000):

$$Y_{ij} = \mu + e_j + \sum_{n=1}^{N} \lambda_n \gamma_{in} \delta_{jn} + \varepsilon_{ij}$$

where, Y_{ij} = mean incidence of ith genotype (i = 1, …, I) in the jth environment (j = 1, …, J), μ= grand mean, e_j = environment deviations from the grand mean, λ_n = The eigen value of PC analysis axis, γ_{in} and δ_{jn} = genotype and environment principal components scores for axis n, N = number of principal components retained in the model and, ε_{ij} = Residual effect ~ **N** (0, σ^2). The GEI interpretation was visualized using the GGE biplot (Yan 2007). This technique helps identify the 'which-won-where' pattern in genotype-environment interactions, ranks entries based on mean performance and stability, and compares genotypes against an ideal standard. The GGE biplot was analyzed using Genstat.v12 software (Roger W. Payne 2009). The applied model used row-metric preserving (SVP = 2), without data transformation (Transform = 0), not scaled (Scaling = 0), and environment centred (Centring = 2).

3.4 Statistical Analysis

A comprehensive analysis involving a two-way Analysis of Variance (ANOVA) was conducted to investigate the differences between root count and root length per irrigation treatment (differing levels of irrigation) and per layer (varying root depths). Then, the effect of different treatments was evaluated on the root count and total root length in different soil layers. Additionally, ANOVA also was employed to test the presence of genotype-environment interaction (GEI) across the eight environments. This was complemented by a linear regression to explore the relationships between SPAD values, chlorophyll fluorescence, and root data. Performed using R version 4.2.1, the study deemed results significant at $p < 0.05$ for the irrigation and soil layer study, utilizing the Tukey test, and at $P < 0.01$ for the GEI analysis, applying Tukey posthoc tests to ensure comprehensive statistical validation.

4 RESULTS AND DISCUSSION

4.1 Part 1: Root Development Monitoring under Different Water Supply Levels

4.1.1 Results

4.1.1.1 Meteorological Data

The meteorological data are presented in Figure 3. The average temperature in 2021 exhibited a decline compared to 2020. However, an intriguing contrast was observed in the extremes of the thermal range. Despite a lower average, 2021 reported both higher maximum and lower minimum temperatures, underscoring a significant increase in annual thermal amplitude compared to the previous year. Regarding precipitation, the total recorded amount of 236mm for 2021 was marginally less than that of 2020 (262 mm). Noteworthy, the temporal distribution of rainfall events throughout the growing season was different, suggesting that the main rainy periods differed between the 2020 and 2021 growing seasons as they were concentrated in the middle of June in 2020 and in the middle of July in 2021 (Figure 3). At the same time, the temporal distribution was more balanced during the 2020 growing season.

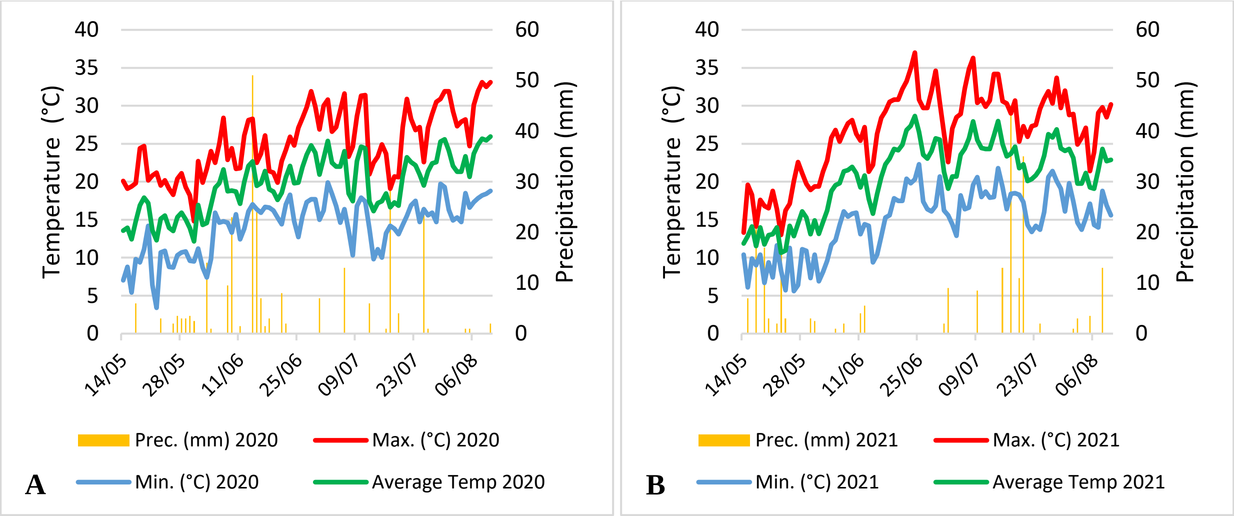


Figure 3. Temperature and precipitation in the 2020 (A) and in the 2021 (B) growing seasons.

4.1.1.2 General Results Regarding Root Count and Root Length

In 2020, a statistical analysis showed that the full irrigation treatment (I100) resulted in a smaller number of roots with less total length than in the water-stressed treatments, meaning 45% less root per plant and 40% less total length compared to the I50 treatment and control, respectively, with no significant difference between the stressed treatments. However, plants subjected to mild and severe stress treatments developed similar average root numbers, with no significant difference. (Figure 4A,B). The results of the 2021 analysis showed that plants under the control treatment produced the highest number of roots with the highest total length, followed by the I100 and then the I50 treatments. No significant difference was found between the two irrigated treatments in 2021. Overall, fewer roots were captured in 2021 than in 2020 (Figure 4). The reason for this difference can be attributed to the different periods when a long-term irrigation deficit could develop and the different irrigation treatments that were applied. This period was determined by the meteorological conditions of the given growing season which is discussed in the Materials and Methods section.

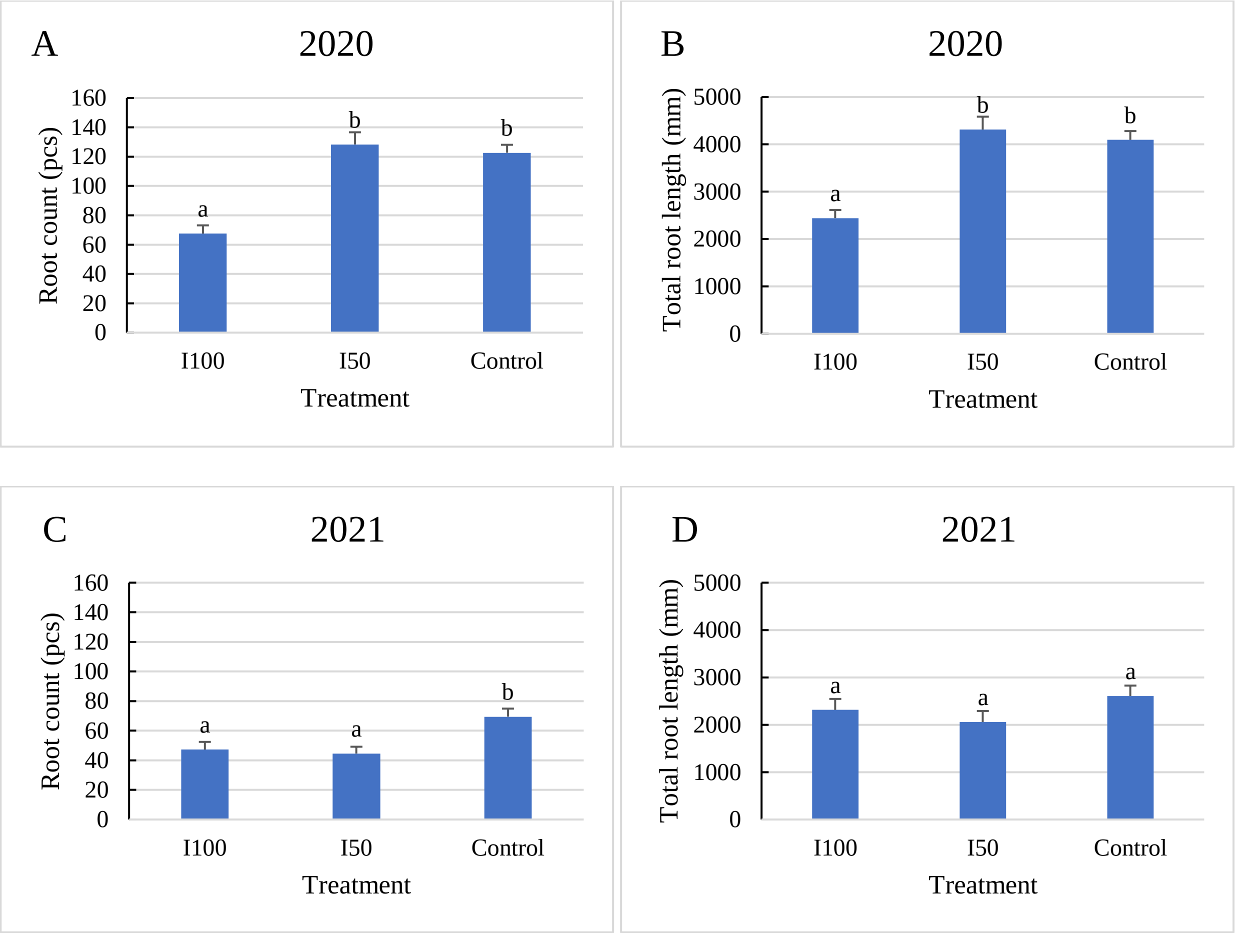


Figure 4. Average root count and total root length under different treatments during the monitoring period. (**A**) Average root count in 2020, (**B**) average root length in 2020, (**C**) average root count in 2021, (**D**) average root length in 2021. Error bars represent SD. Different letters indicate statistically significant difference at $p < 0.05$ level.

The statistical analysis of the 2020 growing season data from the perspective of layers revealed that the number and length of roots developed by tomato plants varied based on their depth in the soil. According to the results shown in Figure 5, plants generally grew more roots with a greater total root length in the middle and bottom layers (in our study, these are the 30–50 and 50–70 cm layers), featuring 127 and 122 detectable roots per plant and total lengths of 4251 and 4319 mm, respectively. In contrast, roots in the top layer did not exceed 71 roots with a 2081 mm total length. The 2021 results reinforced this pattern, indicating that root density in the soil increased with depth.

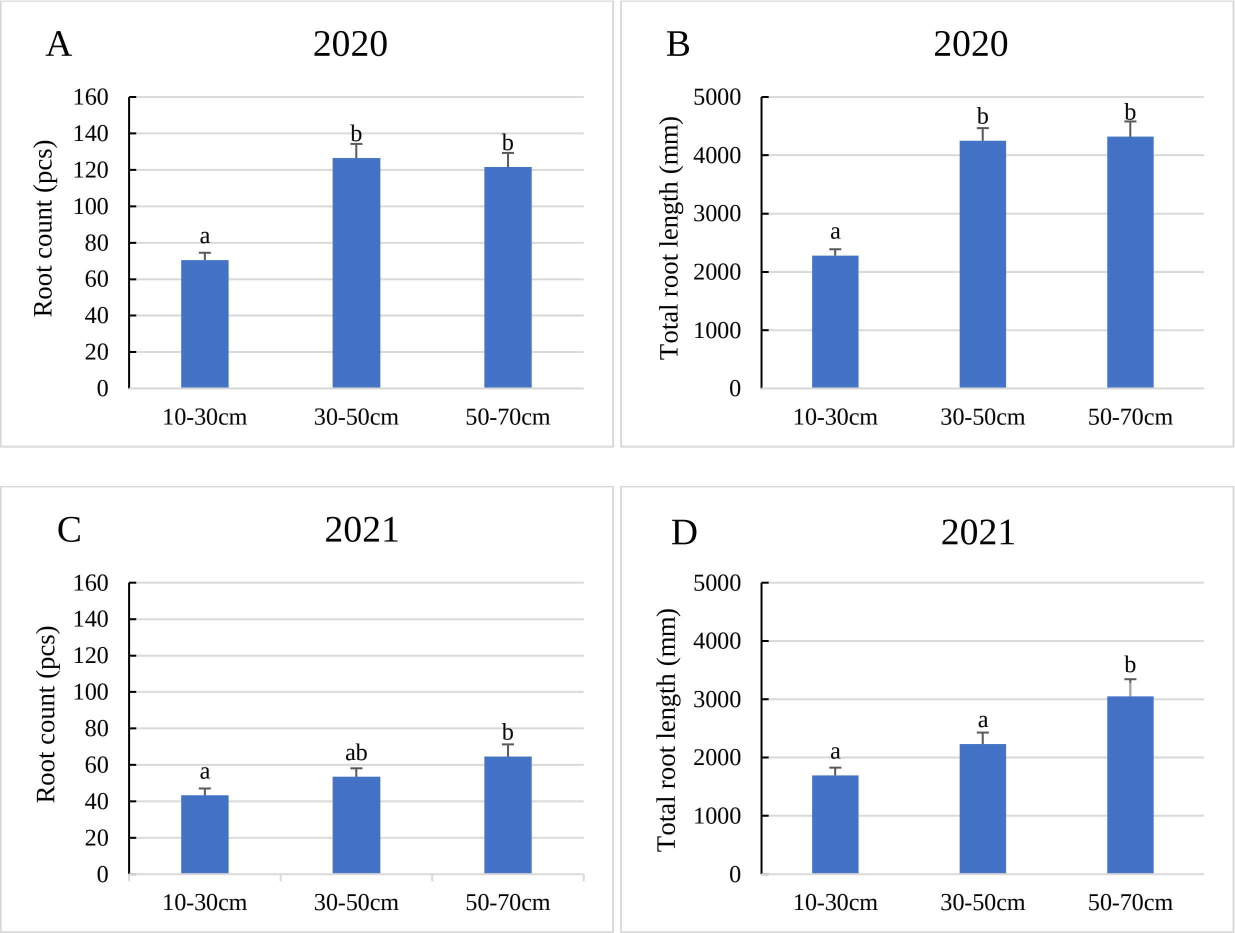


Figure 5. Average root count and total root length in different soil layers. (**A**) Average root count in 2020, (**B**) total root length in 2020, (**C**) average root count in 2021, (**D**) total root length in 2021. Error bars represent SD. Different letters indicate statistically significant difference at $p < 0.05$ level.

4.1.1.3 Evaluation of the Time Scale for the Monitored Root Zone

In 2020, the plants under the mild stress treatment (I50) exhibited significantly more roots with longer total lengths by the end of the monitoring period compared to the control and I100 treatments. Although the initial data suggested that the I50 and I100 treatments started on similar grounds, by the second week of monitoring, the rapid growth rate of the I50-treated plants led to a high root count comparable to the control. This observation could suggest that mild stress conditions stimulate the plants to develop more roots to absorb available water.

In contrast to 2020, the 2021 growing season demonstrated a reduced number of roots and total root length in all the treatment groups (Figure 6). Notably, the plants in the control group showed the most extensive root growth, developing 99 roots with a total length of 3689 mm by the end of the six-week monitoring. The irrigated treatments produced similar root counts during the experiment, and the three treatments barely differed in the final two weeks of the monitoring period in root length, while the higher number of roots in the control was continuous from the second week of the monitoring period.

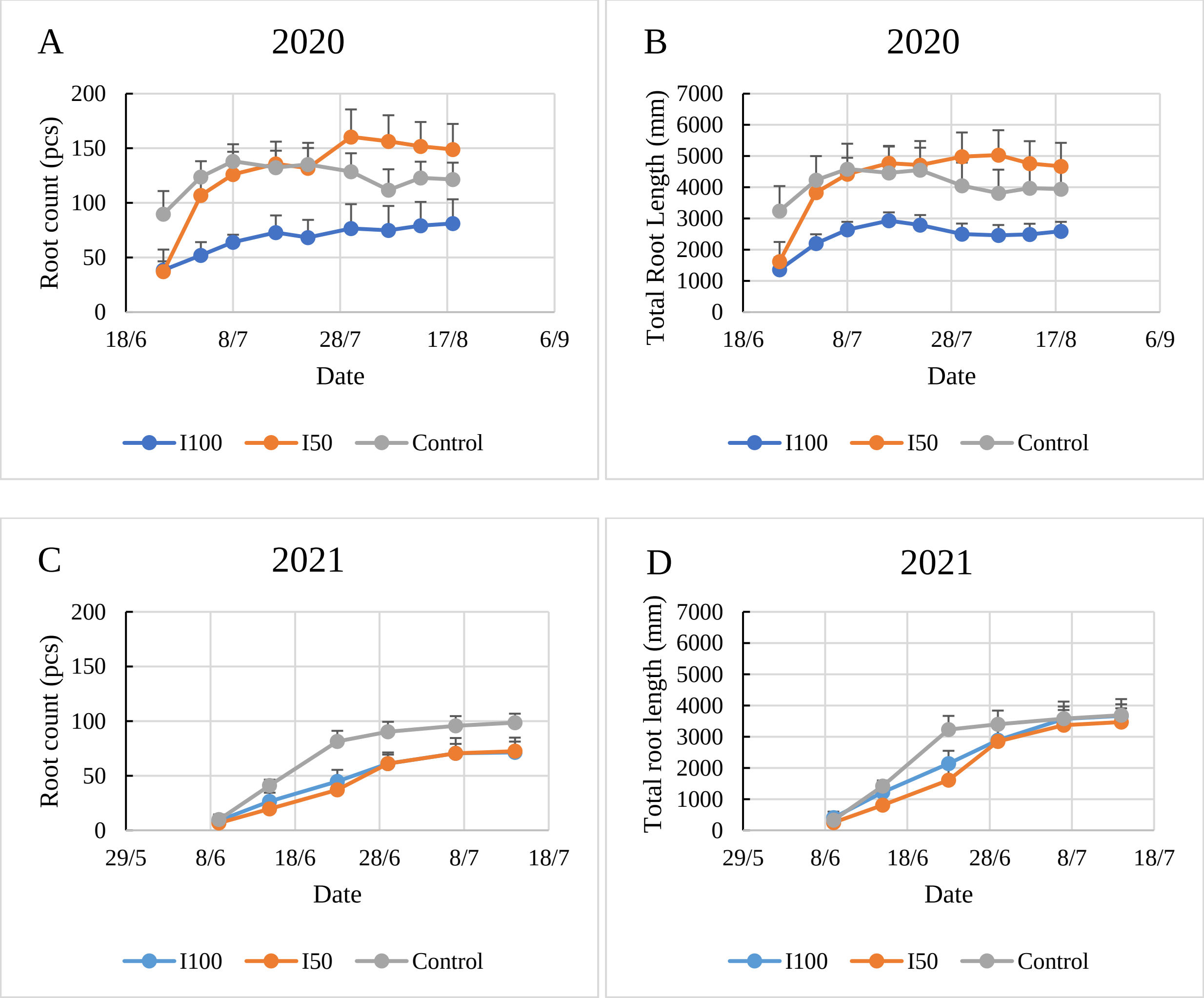


Figure 6. Evolution of root count and total root length in time under different water supply treatment. The numbers are summarized in the three observed root zone layers. (**A**) Evolution of root count in 2020, (**B**) evolution of root length in 2020, (**C**) evolution of root count in 2021, (**D**) evolution of root length in 2021.

4.1.1.4 Evaluation of the Layer Scale for the Monitored Root Zone

In the 2020 growing season, the statistical analysis indicated a statistically significant difference between the layers within each treatment (Figure 7A.B). Consequently, the distribution of roots was not uniform in the 10–70 cm rooting depth. The top layer developed a smaller number of roots with the least total length under all treatments (Figure 7). Regarding the top layer, the highest number and length were captured in the control, meaning 88 roots with a length of 2825 mm in the 10–30 cm layer. The plants that received full irrigation developed the highest number and longest roots in the middle layer compared to the other two soil layers, growing 90 roots with a 3454 mm root length. The 2020 growing season data revealed no significant difference in either root number or total length between the middle and bottom layers in the mild stress treatment. Furthermore, the statistical analysis indicated that the interaction effect between treatment and layer significantly affected root length but not root count.

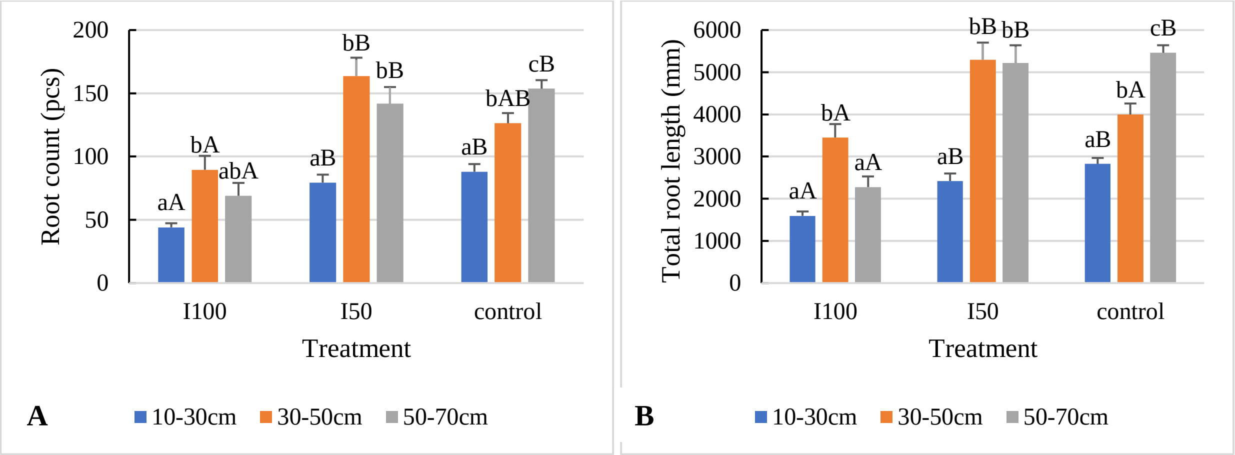


Figure 7. Effect of different water supply levels on root count (A) and total root length (B) across different soil layers in 2020. Error bars represent SD. Different capital letters above the columns represent statistical differences caused by different water supply levels within the same layer, while lowercase letters express differences within the same treatment between soil layers at $p < 0.05$ level.

Contrarily, the 2021 findings regarding vertical root distribution deviated from the 2020 results (Figure 8). In general, the deeper the soil layer was, the more and longer roots were developed, and the highest counts were observed in the control treatment in the 50–70 cm layer, represented by 85 roots with a total length of 3355 mm in 2021. However, the total length of roots was the lowest in the I50 treatment in the 10–30 cm layer (1272 mm), and the number of roots was equally low in the irrigated treatments in this layer (35 pcs). Comparing the number and length of roots, the data revealed that the plants in the control group developed significantly more roots in each layer compared to the irrigated treatments due to severe water stress. The results for the I50 treatment were inconsistent in the two growing seasons since the same level of root number was detected as in the I100 treatment; however, the detected roots in the 10–30 cm layer were shorter for the I50 treatment compared to the I100 treatment. This discrepancy was most remarkable in the middle layer of the I50-treated group between the two years. Additionally, the statistical analysis also showed that there is no significant interaction between treatment and layer, meaning that the impact of the water treatment on root count and total root length seems to be similar across different soil layers.

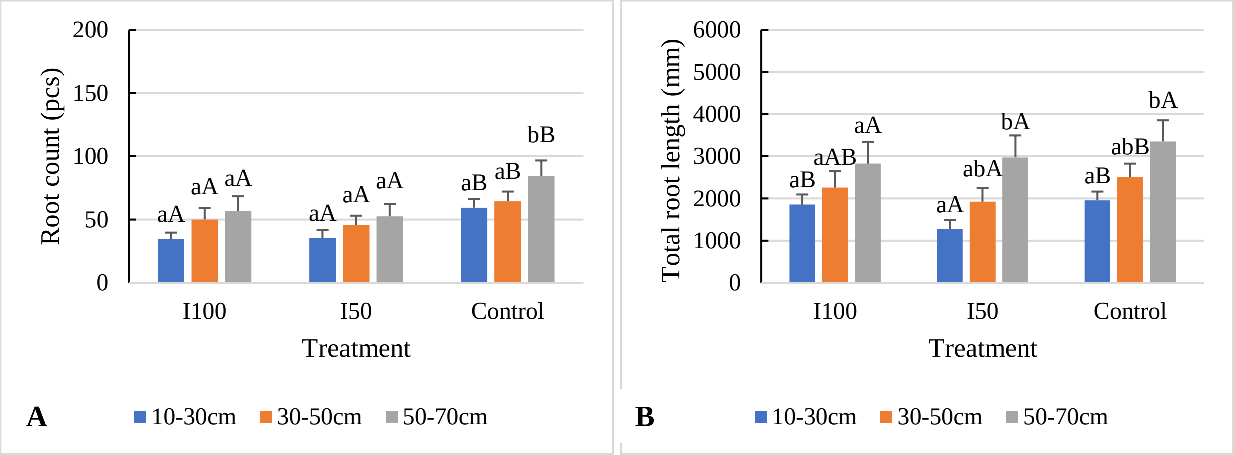


Figure 8. Effect of different water supply on root count (a) and total root length (b) between different soil layers in 2021. Error bars represent SD. Different capital letters above the columns represent statistical differences caused by different water supply within the same layer, while lowercase letters express differences within the same treatment between soil layers at $p < 0.05$ level.

4.1.1.5 Root Development during the Monitoring Period

According to the data in Figure 9, the top layer developed fewer roots under all treatments, except for the I100 treatment at the beginning of the monitoring period in the 2020 growing season. The maximum count of 104 roots in the 10–30 cm layer was found in the control group on 8 July. By the end of the monitoring period, it was found that the full irrigation treatment produced the most dense roots in the middle layer, expressed mostly in the total root length (Figure 9A,B). These graphs indicate that during the intensive root development period between 25 June and 8 July, the plants under full irrigation developed shorter root systems compared to the plants under severe and mild stress treatments. Under the mild stress treatment, the plants in the middle and bottom layers developed roots of nearly the same length throughout the entire monitoring period. The increment in root count was continuous and consistent until a later date of 30 July for the I50 treatment compared to the control and the I100 treatment, where root count stagnated after the intensive development period and even showed a decrease later in the monitoring period, especially in the control. The biggest growth in root count during the 2020 monitoring period was recorded in the 50–70 cm layer for the I50 treatment when the number of roots increased by 460% during one week (Figure 9C).

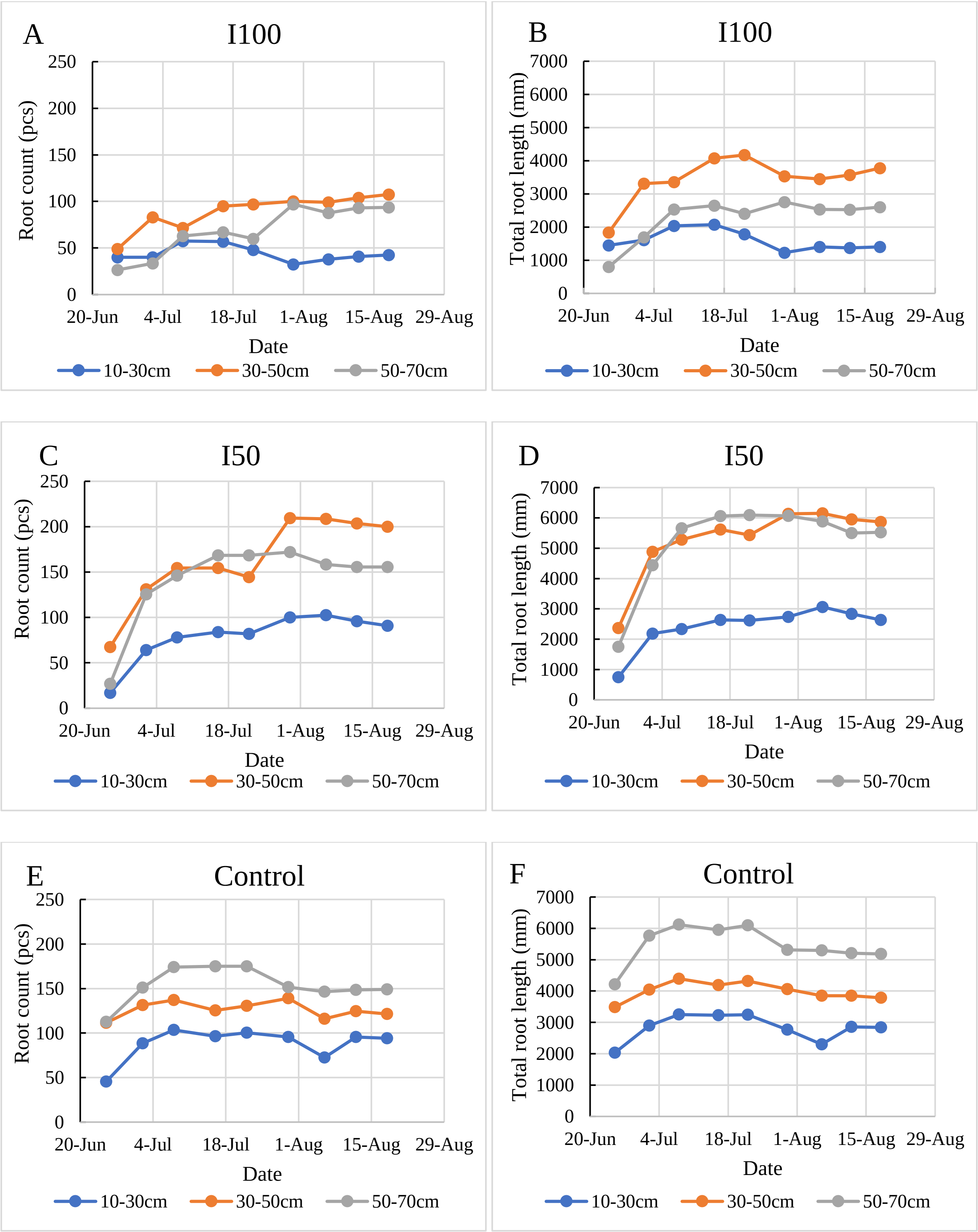
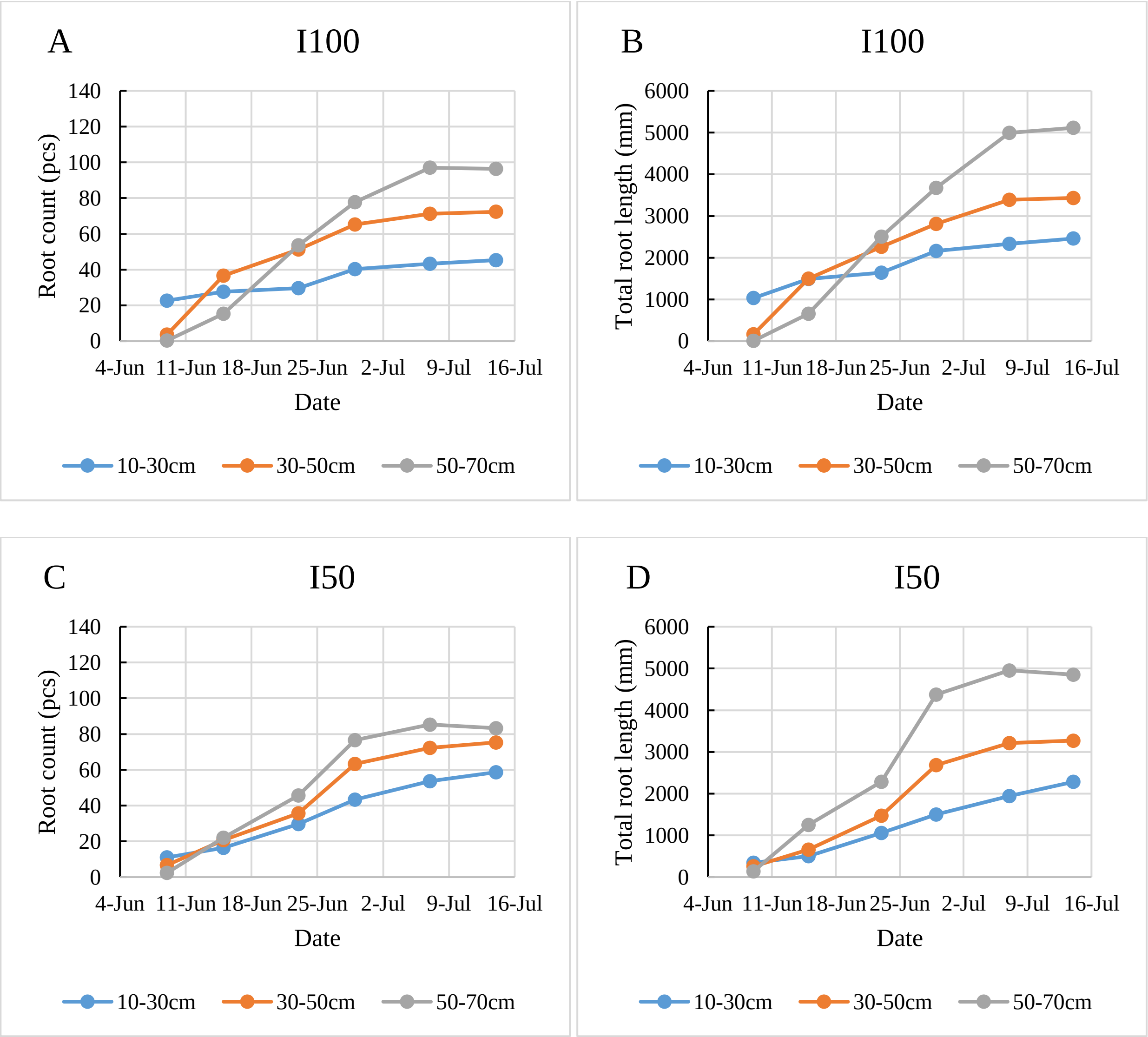


Figure 9. Evolution of root count (A,C,E) and the total root length (B,D,F) in different layers under the different water supply treatments in 2020.

In the initial week of monitoring in 2021, the root systems of the plants in all the treated groups were primarily concentrated in the top soil layer (Figure 10). Under the full irrigation treatment, roots showed the most robust growth in the bottom layer, reaching a maximum of 97 roots and a total length of 5113 mm. A shift was observed starting from the third week. Under all treatments, the bottom layer consistently exhibited the highest root count and total root length, followed by the middle and then the top layer. This trend became increasingly pronounced as the weeks progressed. The 50–70 cm layer became the most densely rooted layer following the period between 15 and 23 June, when the most intensive growth took place. This intensive growth period was most prominent in the control group, where the rapid development produced almost a 200% increase in root numbers and more than a 300% increase in root length. This result deviated from the 2020 findings and contradicted our expectations of denser root systems in the upper layers under frequent irrigation. Both the I50 treatment group and the control group exhibited a broader distribution of roots in the deeper soil layers sooner than in the I100 treatment group. The I50 treatment group reached a maximum of 85 roots with a total length of 4955 mm, while the control treatment group reached a maximum of 122 roots with a total length of 4835 mm. It is worth noting that, while the control treatment group demonstrated a slightly higher root count and total root length, these values seemed to stabilize in the final three weeks of the monitoring period.



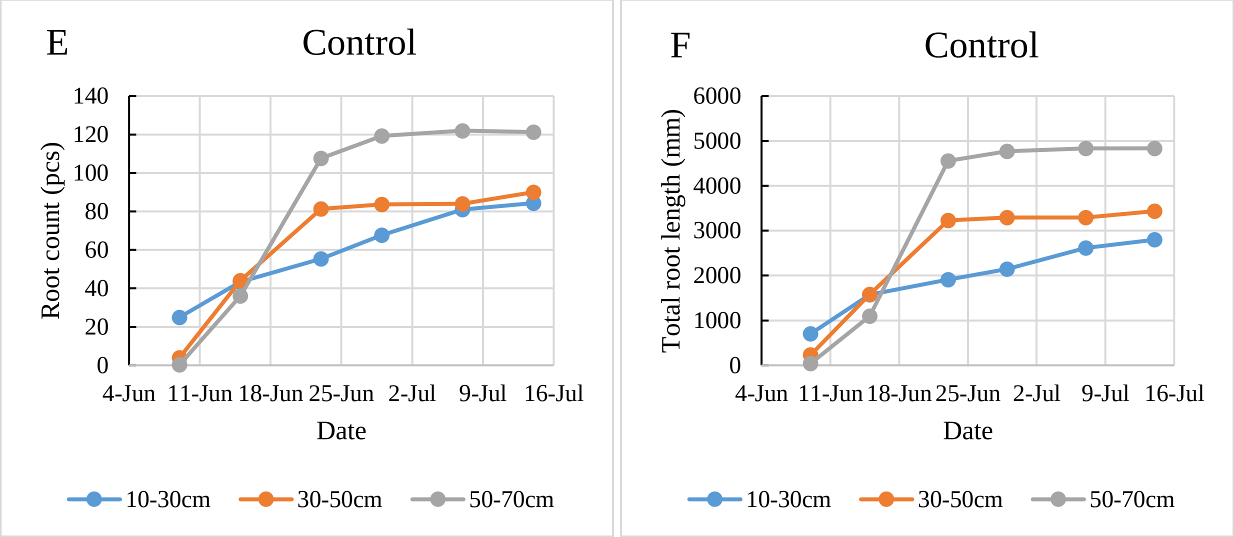


Figure 10. Evolution of root count (A,C,E) and the total root length (B,D,F) in different layers under the different water supply treatments in 2021.

4.1.1.6 Comparison of the Root Development in the Two Years

The comparative results demonstrate that the tomato plants cultivated in 2020 exhibited more substantial root growth and lengthier roots compared to those grown in 2021, regardless of the treatment applied (Figure 11). Consequently, the highest quantity and length of roots were observed in 2020 under the I50 treatment. The highest root count was recorded at 128 and 69, with corresponding total lengths of 4313 mm and 2607 mm for the years 2020 and 2021, respectively. Meanwhile, the minimum root count was observed in the 10–30 cm soil layer, with 70 and 41 roots and total lengths of 228 mm and 1610 mm, respectively, for 2020 and 2021. In 2020, the root counts were nearly equal in the 30–50 and 50–70 cm layers, whereas in 2021, both exhibited a consistent increase towards the bottom layer. The differences between treatments were less explicit in 2021.

The statistical analysis of the interaction effects between the year of measurement and the water treatment demonstrated a significant effect on both root count and root length, suggesting that the effectiveness of water treatments on root development varies depending on the year. On the other hand, the interaction between year and layer on root count shows significance (p-value = 0.0164), while it isn't significant for Root Length (p-value = 0.115), indicating that the influence of soil layer on root length is consistent across different years.

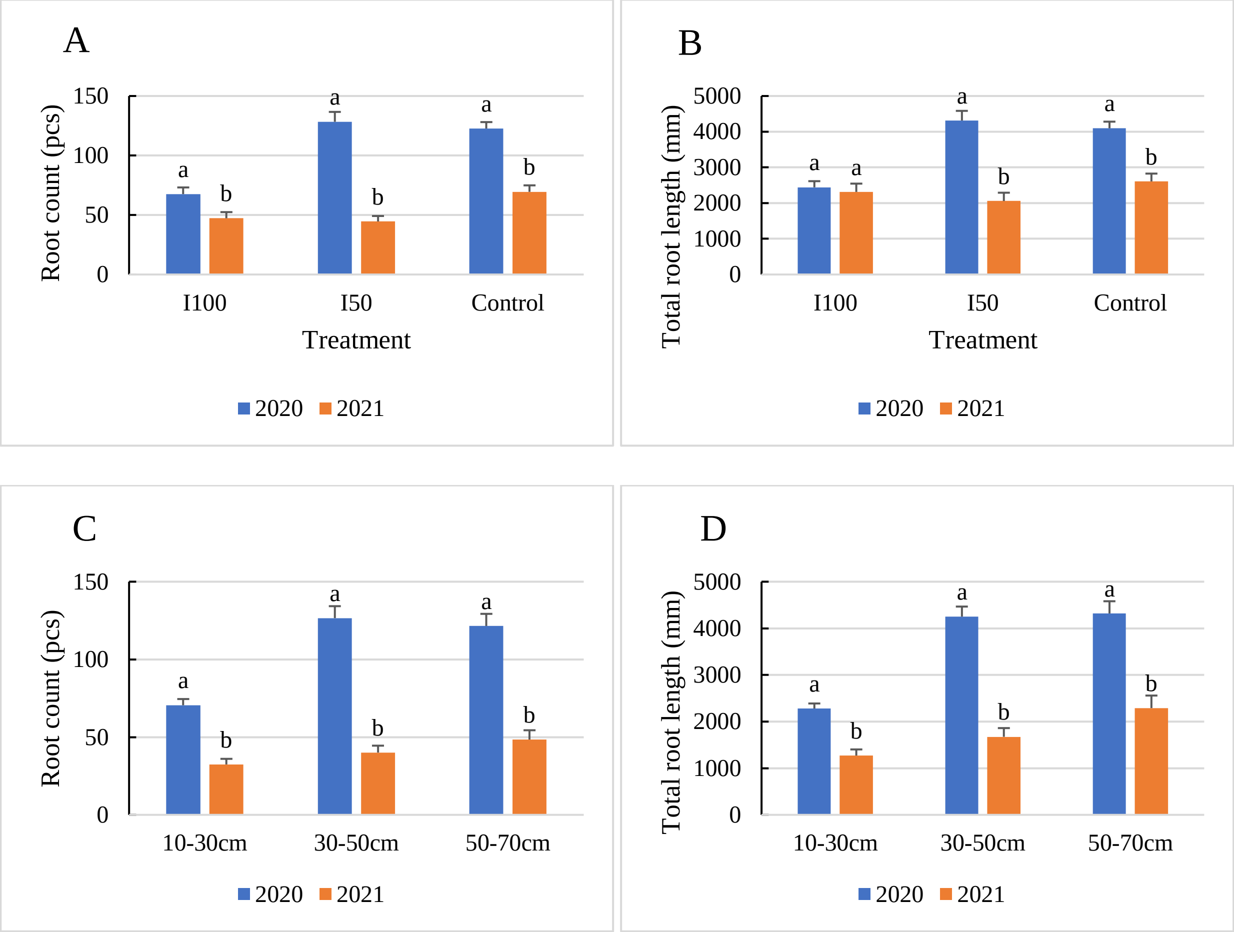


Figure 11. Comparison of average root count (A,C) and the average of total root length (B,D) under the different treatments (A,B) and in different soil layers (C,D) for the 2 growing seasons. Error bars represent SD. Different letters indicate statistically significant difference growing seasons at $p < 0.05$ level.

4.1.1.7 Effect of Different Treatments on Relative Chlorophyll Content (SPAD) and Chlorophyll Fluorescence (Fv/Fm)

The available data facilitate a comparison of the SPAD values of the tomato plants under the different irrigation treatments on each measurement date. In 2020, on 8 July, the I100 treatment displayed a lower SPAD value compared to both the I50 treatment and the control treatment (Figure 12A,B). From 15 July to 29 July, the I50 treatment's values generally surpassed those of the I100 treatment but fell short of the control treatment's values. The SPAD values for all three treatments diminished during this period. On both 6 August and 12 August, the statistical analysis showed no significant difference between treatments. In 2021, the SPAD values of the control treatment group were significantly higher than the I100 and I50 groups during the whole measurement period, except on 30 June, where all treatments showed comparable SPAD values, indicating similar chlorophyll content. On 7 July, the control treatment exhibited higher SPAD values compared to the irrigated treatments. By 14 July, the differences between the treatments became more pronounced and significant. The I100 treatment showed a slight increase in SPAD value to 57.4, while the I50 and the control treatments exhibited a larger increase to 61.4 and 70.2, respectively, indicating a greater chlorophyll content.

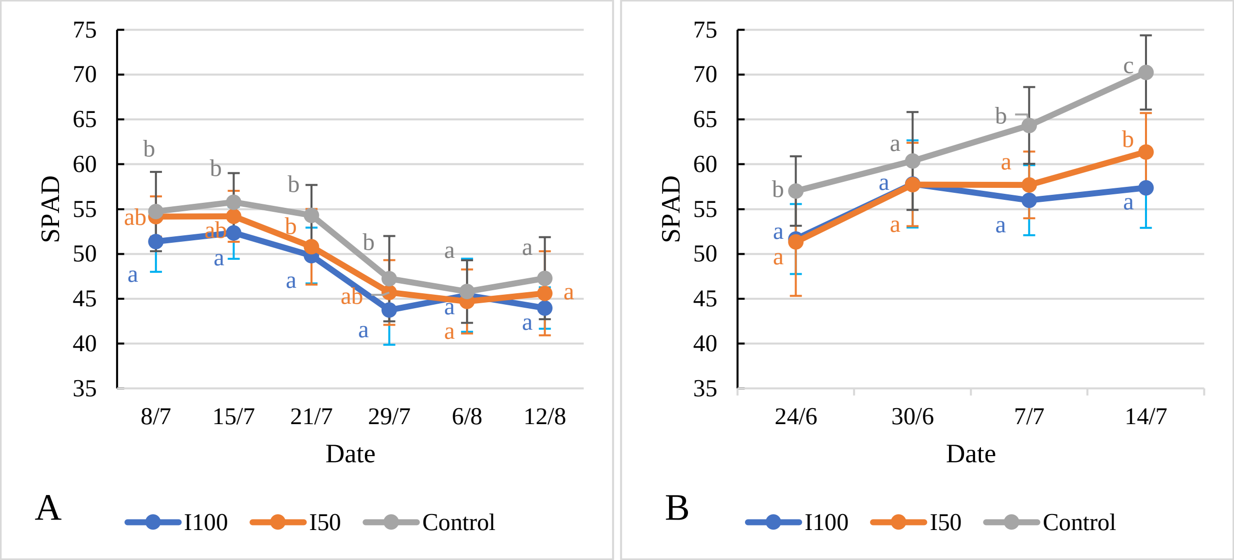


Figure 12. Evolution of chlorophyll content (SPAD) under different water supply treatment. (A) Evolution in 2020, (B) evolution in 2021. Error bars represent SD. Different letters indicate statistically significant difference at $p < 0.05$ level.

Linear regression tests were performed on the means of the given treatments regarding the SPAD and the number and length of the roots. A weak relationship was found between SPAD values and root development in the 2020 growing season, where the highest regression coefficient was $R^2 = 0.22$ for the root length in the 10–30 cm layer. On the other hand, in the 2021 growing season, the strongest relationship was found between SPAD values and root length in the 10–30 cm layer ($R^2 = 0.63$), and the regression coefficient became lower for deeper soil layers. The linear relationship between root count in the top layer and SPAD values was even stronger, $R^2 = 0.87$, and also slightly reduced with the deeper layers, but still indicated a good relationship in the deepest monitored layer, $R^2 = 0.62$.

The data show the chlorophyll fluorescence values of tomato plants under the different irrigation treatments on each measurement date (Figure 13). In 2020, the I100 plants initially exhibited lower values compared to the I50 and control plants. However, over time, the chlorophyll fluorescence values for the I100 plants gradually increased and eventually surpassed the values of the I50 and control plants by the time of the maturity period, after the irrigation had ended. The statistical analysis indicated that, except for the measurement taken on 29 July, there were no significant differences observed between the treatments. A similar observation was recorded in 2021, when the statistical analysis revealed that there were no significant differences between the treatments on all measurement dates, except for 29 July, where the control treatment had a higher value than the I50 treatment, which reported the lowest value. Both irrigated treatments displayed very similar values.

No significant relationship was found between chlorophyll fluorescence and root development data in 2020. On the other hand, in the 2021 growing season, a similar relationship was found between the Fv/Fm data and root count in all the monitored layers, ~ $R^2 = 0.35$. Root length showed a similar but slightly stronger relationship, where the middle layer was found to be the highest, $R^2 = 0.41$.

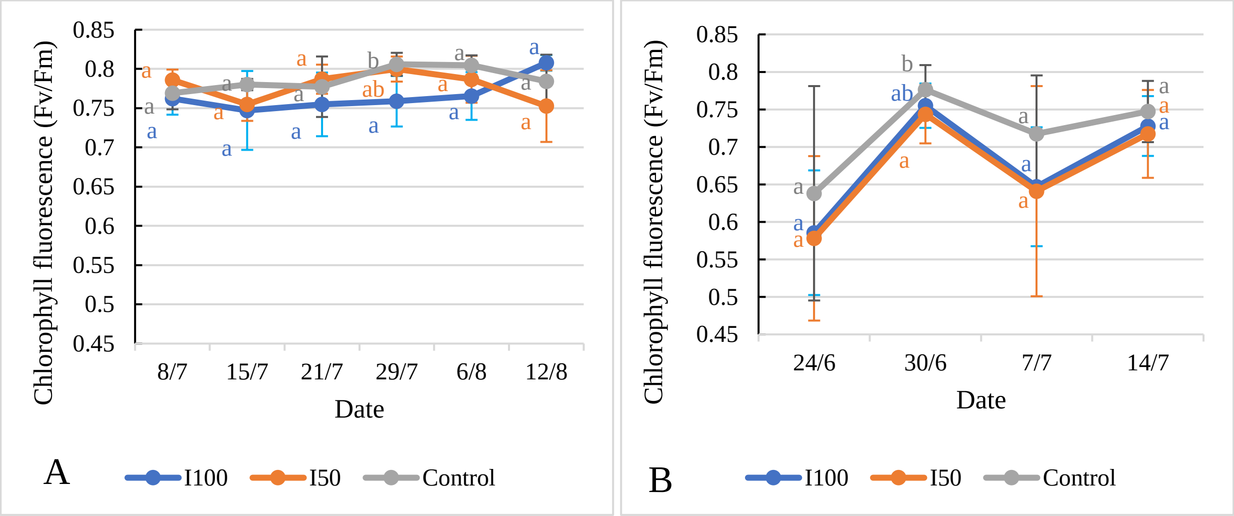


Figure 13. Evolution of chlorophyll fluorescence under different water supply treatment. (A) Evolution in 2020, (B) evolution in 2021. Error bars represent SD. Different letters indicate statistically significant difference at $p < 0.05$ level.

4.1.2 Discussion

The results obtained in 2020 were in agreement with the result of Scholander et al. (1965), who reported that plants, when faced with water scarcity, tend to develop deeper roots, accessing sections with higher soil moisture content. Plants can avoid water-deficient conditions by maintaining relatively high tissue water potential despite a shortage of soil moisture. To maintain turgor, plants employ adaptive strategies such as increasing root depth or evolving an efficient root system to maximize water uptake (Singh et al., 2017).

Tomato plants grown under open-field conditions can endure prolonged periods of low soil water content, but the severe stress affects biomass and yield significantly (Katerji et al., 2013), and a deeper root system with powerful suction force can better utilize deep soil moisture (Nemeskéri & Helyes, 2019). Our observations presented on Figures 8–10 further support this, and they are in agreement with other studies stating that drought stress often results in a larger root system (Nicolas et al., 1985), which can enhance the efficacy of water uptake and assist plants in resisting water stress at the reproductive stage (Lee et al., 2016). Previous studies have demonstrated that the frequency of irrigation not only fulfils tomatoes' water demand but also influences the efficacy of the root system. Specifically, an increase in irrigation frequency tends to slow down root system growth; thus, larger irrigation intervals stimulate development, improving secondary root branching, main root deepening, and water and nutrient uptake (Fara et al., 2019). According to our study, the irrigation rate affected root growth similarly to irrigation frequency, as lower irrigation rates resulted in the root system expanding to deeper layers. Root depth was also reported to increase upon exposure to water limitation in other studies, and this was confirmed by our results (Vadez et al., 2013). In addition to absorbing nutrients more efficiently, plants with deeper root systems can tolerate less frequent irrigation during subsequent growth stages (Koevoets et al., 2016; Marouelli & Silva, 2007; Wasaya et al., 2018), and as our results showed, tomato roots rapidly respond to low soil moisture conditions, promoting the deepening of the tomato root system. Previous research also demonstrated that plants adjusted their growth morphology and physiological indices to adapt to water stress over time (He & Dijkstra, 2014). Such structural changes could occur due to the reallocation of assimilates from the shoot to the root by plants under water stress, resulting in reduced shoot growth and enhanced root system traits such as root length and number under diminished irrigation, as water is the main driver of resource allocation (Khapte et al., 2019). Our records of rapid growth within a week reinforced this.

Root architecture may also play a significant role in water usage, as it can affect the timeline for utilizing water resources across different layers (Manschadi et al., 2006). Guida et al. (2017) mentioned that drought avoidance mechanisms, such as root deepening, allowed non-irrigated tomato plants to uptake water from soil layers much deeper than 40cm, which is a finding also supported by our results. This pattern aligns with previous studies (X. Li et al., 2017), suggesting that roots grow downwards, especially under the 50 cm soil layer, when the upper soil profile fails to meet the crop’s water requirements. However, Schneider et al. (2019) revealed that the plant’s ability to cope with water scarcity might decrease as the intensity and duration of the drought conditions increase, although other studies stated that this response can vary between different plant species and tissues (Lum et al., 2014; Mirzaee & Ghanati, 2013). In our case, processing tomato plants growing in dry soil during the vegetative growth stage developed deeper root systems than those growing in irrigated soil, which agrees with Marouelli & Silva, (2007). This was observed especially in 2021, but the roots under full irrigation were well developed in all three monitored layers, but the rate of development differed. This observation can be attributed to the increased piezometric head from the large amounts of water applied, causing moisture to move downward and promoting root system growth in the same direction, a phenomenon previously described under flood irrigation (Jha et al., 2017; Li et al., 2010). The results of the I50 and control treatment were more pronounced, with roots growing downward to explore available water sources, leading to more roots in the subsoil, as described by others as well (S. P. Sharma et al., 2014). A higher soil water content could enhance the growth of new roots as long as they remain aerated and below the hypoxia level (Patanè et al., 2011; Wang et al., 2019; Zotarelli et al., 2009).

Our findings from 2021 (Figure 10) showed that, for the full irrigation treatment, root development was more expressed in the upper soil layer and expanded to deeper layers over more time than for the control and I50 irrigation treatments. In another study, where the soil depth was monitored up to 2 m in wheat, it was found that the root length density in the 0–40 cm layer was the highest under well-irrigated conditions, followed by limited irrigation and no irrigation conditions (C. Xu et al., 2016). The highest total root length was reached under high water conditions, supplemented with a moderate nitrogen supply, in a study conducted with minirhizotron in similar depths on cotton plants, which is in contrast with our results acquired in 2020 (Wu et al., 2023).

The difference in root growth between the two years of tomato cultivation in the same soil with the same water supply level treatment could be attributed to climatic condition differences between the two growing seasons, causing the occurrence of continuous water deficit periods attached to different phases. Temperature variations could also impact root growth, as plants tend to adapt their growth patterns in response to optimizing water uptake (Luo et al., 2020). These findings align with other research in the field that has found that temperature and irrigation variations can significantly influence root growth patterns in various plant species (Chakraborty et al., 2022; Karlova et al., 2021).

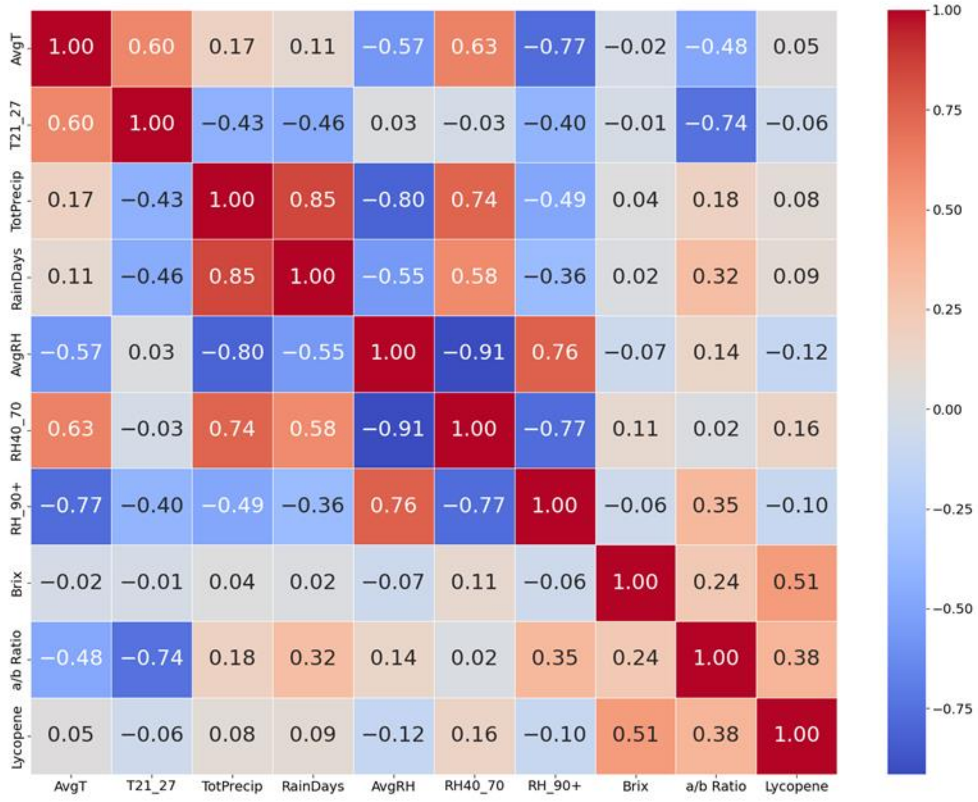
Higher SPAD values caused by water deficiency are consistent with the findings of other studies (Nemeskéri et al., 2019), where the highest SPAD values were recorded under non-irrigated conditions during flowering and fruit settings and early fruit development. The chlorophyll fluorescence data (Figure 13) from both years indicated fluctuations in Fv/Fm values between the different irrigation treatments. Previous research has indicated that heat stress can lead to decreased maximum photochemical efficiency in detached tomato leaves (Karlova et al., 2021). Furthermore, drought-tolerant transgenic tomato plants have been found to maintain anideal Fv/Fm value of 0.7, which indicates optimal photosynthetic performance (Mishra et al., 2012). In our study, while the Fv/Fm values varied, no significant differences were observed between treatments, suggesting that there are inherent mechanisms in tomato plants that maintain stable chlorophyll levels and photosynthetic efficiency, even under water stress conditions. These mechanisms could include adjustments in stomatal conductance, osmotic adjustment, or the activation of antioxidant defense systems that help mitigate water stress’s negative impacts on chlorophyll-related parameters (Nemeskéri & Helyes, 2019). In 2021, the Fv/Fm value was lower than the 2020 results, which could be due to differences in temperature between the two years. The higher Fv/Fm value in the control treatment in 2021 could indicate that plants without regular irrigation have to better adapt to environmental conditions, leading to increased photosynthetic efficiency.

While our non-destructive root monitoring method offers valuable insights, it is important to acknowledge certain inherent limitations. Due to the intricate nature of root systems, roots in close proximity might not always be distinctly analyzed. Additionally, image clarity can sometimes be compromised, potentially affecting the precision of the analysis. A notable constraint of this root scanning method was its ability to capture only the roots growing adjacent to the curved surface of the tube.

4.2 Part 2: A Comparative Analysis of XGBoost and Neural Network Models for Predicting Tomato Fruit Quality

4.2.1 Results

4.2.1.1 Correlation Heatmap



	AvgT	T21_27	TotPrecip	RainDays	AvgRH	RH40_70	RH_90+	Brix	a/b Ratio	Lycopene
AvgT	1.00	0.60	0.17	0.11	−0.57	0.63	−0.77	−0.02	−0.48	0.05
T21_27	0.60	1.00	−0.43	−0.46	0.03	−0.03	−0.40	−0.01	−0.74	−0.06
TotPrecip	0.17	−0.43	1.00	0.85	−0.80	0.74	−0.49	0.04	0.18	0.08
RainDays	0.11	−0.46	0.85	1.00	−0.55	0.58	−0.36	0.02	0.32	0.09
AvgRH	−0.57	0.03	−0.80	−0.55	1.00	−0.91	0.76	−0.07	0.14	−0.12
RH40_70	0.63	−0.03	0.74	0.58	−0.91	1.00	−0.77	0.11	0.02	0.16
RH_90+	−0.77	−0.40	−0.49	−0.36	0.76	−0.77	1.00	−0.06	0.35	−0.10
Brix	−0.02	−0.01	0.04	0.02	−0.07	0.11	−0.06	1.00	0.24	0.51
a/b Ratio	−0.48	−0.74	0.18	0.32	0.14	0.02	0.35	0.24	1.00	0.38
Lycopene	0.05	−0.06	0.08	0.09	−0.12	0.16	−0.10	0.51	0.38	1.00

Figure 14. Correlation heatmap of Brix, a/b ratio, Lycopene and key climatic variables. Hues of red and blue represent the strength of correlation as it is visualized on the legend (n=28,474).

The generated correlation heatmap, presented in Figure 14, offers a comprehensive insight into the linear relationships between the climatic variables, Brix, Lycopene and a/b ratio. The intensity and direction of relationships are visually represented through a spectrum ranging from cool blue for negative correlations to warm red for positive ones, a method validated by Waskom, (2021). Notably, the heatmap reveals a significant positive correlation between 'AvgT' (average temperature) and 'T21_27' (number of days with temperatures between 21°C and 27°C), suggesting that higher average temperatures during growing seasons often correlate with increased days in the optimal temperature range for growth. Moreover, 'TotPrecip' (total precipitation) and 'RainDays' (number of rainy days) show a strong alignment, underscoring the intuitive link between increased rainy days and higher total precipitation, a vital factor in agricultural water resource management and irrigation strategies. Conversely, an inverse relationship is observed between 'AvgRH' (average relative humidity) and 'RH40_70' (days with 40% to 70% humidity), indicating that seasons with higher overall humidity tend to have fewer days within the ideal humidity range for cultivation. Importantly, a pronounced correlation is detected between 'RH_90+' (days with over 90% humidity) and 'Brix', suggesting that elevated humidity levels might influence the sugar concentration in fruits. The 'a/b ratio' also demonstrates notable correlations with several climatic parameters. All eight meteorological variables were incorporated as independent factors in our predictive models, aiming to provide comprehensive insights into the influences on fruit quality and yield.

4.2.1.2 Model performance on Brix prediction

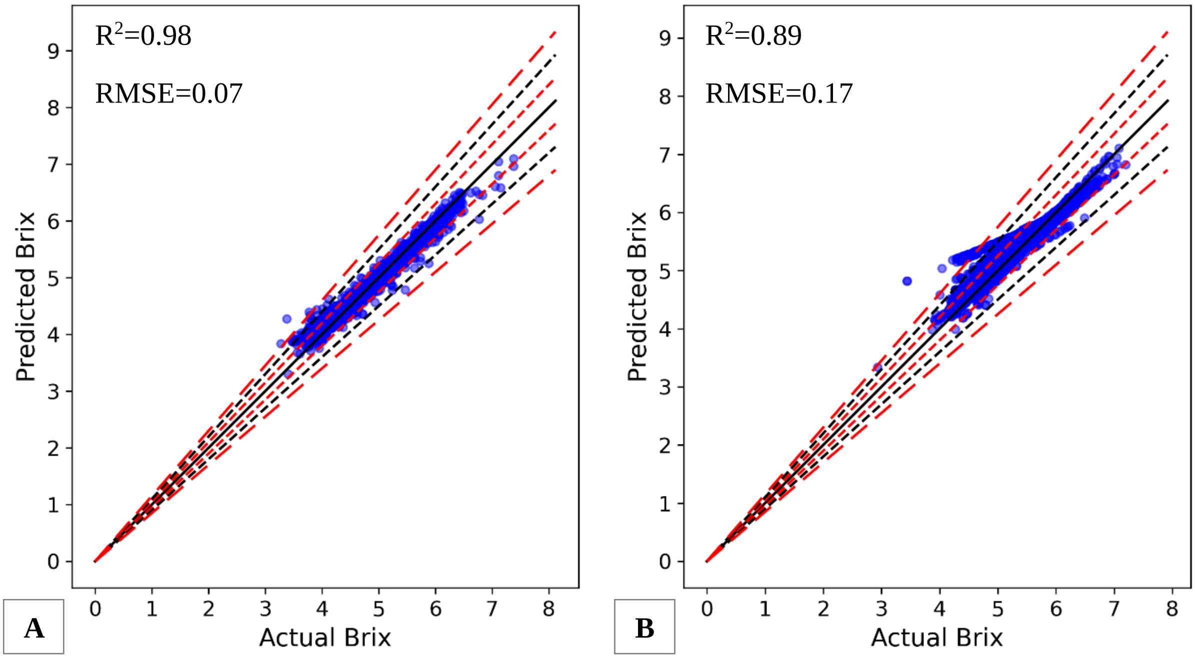


Figure 15. Actual vs Predicted Brix utilizing the XGBoost (A) and ANN (B) models. Black solid line indicates perfect prediction meaning that y = x, the red short-dashed lines, black dashed lines, and red long-dashed lines indicate ± 5, 10, and 15% deviation from the y = x line, respectively (n=4,746).

The developed algorithms exhibited a high degree of accuracy when estimating the Brix values (Figure 15). The XGBoost model yields an impressively a robust R^2 value of 0.98 and low RMSE of 0.07. Such results not only vouch for the XGBoost algorithm's capability but also highlight the significance of the chosen features in predicting Brix values from other climatic and quality variables. On the other hand, the ANN model resulted in an R^2 of 0.89 and RMSE of 0.17, marking its good performance in intricate predictive modelling scenarios. The presented scatter plots from the two distinct models provide insights into their performance efficacy in predicting Brix values. Both plots display a significant concentration of data points around the black line representing x=y, highlighting the commendable accuracy of both models. For the XGBoost model and the ANN respectively, the percentage of predictions deviating less than 5% are 97% and 89%; those deviating between 5% and 10% were 2.6% and 8.4%; those deviating between 10% and 15% were 0.4% and 1.4%; and those deviating more than 15% are 0.06% and 1.12%. It's noteworthy that a predominant cluster of data points for both models lied within the 5% error margin, signifying that the models' predictions are not only accurate but also consistent. These statistics underscore the models' competence in closely estimating the actual Lycopene content, despite some error margins which are to be expected in predictive modelling.

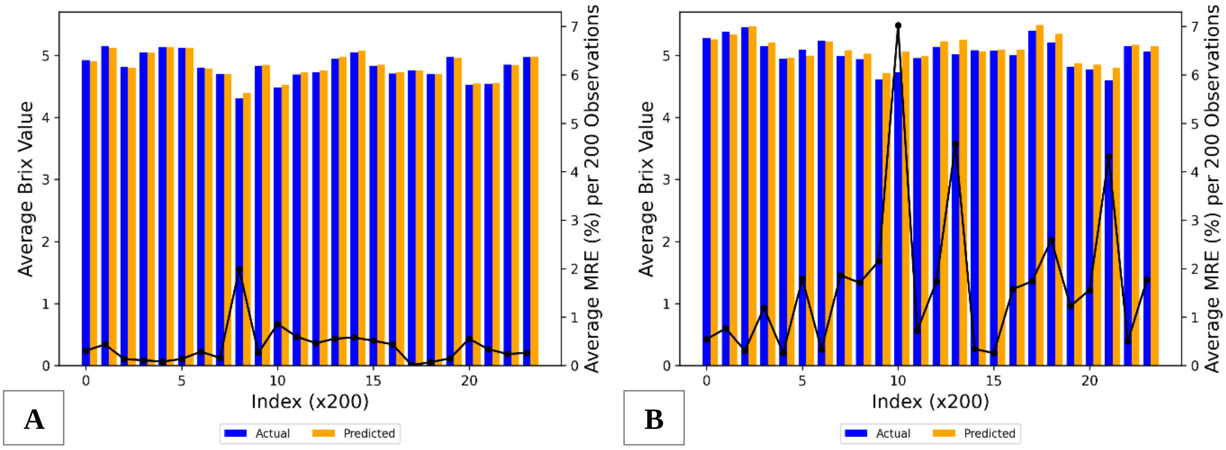


Figure 16. Comparison of MRE (black line) of actual (blue bars) vs predicted (orange bars) Brix values per 200 observations (Index x200) for XGBoost (A) and ANN (B) models.

The Mean Relative Error (MRE) graph in Figure 16 provided a visual assessment of the prediction errors made by the XGBoost and ANN models in estimating Brix values. According to Figure 16A, the MRE for the XGBoost model was as low as approximately 0.25% in some intervals, indicating high predictive accuracy, but it reached upwards of 2% in others, suggesting a reasonable predictive performance overall. On the other hand, the second graph demonstrated the MRE for the ANN model, which varied significantly, ranging from approximately 0.5% to nearly 7%. While both models showed areas of agreement between actual and predicted Brix values, the ANN model exhibited higher variability in prediction accuracy. This variability suggested that, in this specific application, the XGBoost model might have offered more consistent predictions compared to the ANN model.

4.2.1.3 Model performance on Lycopene prediction

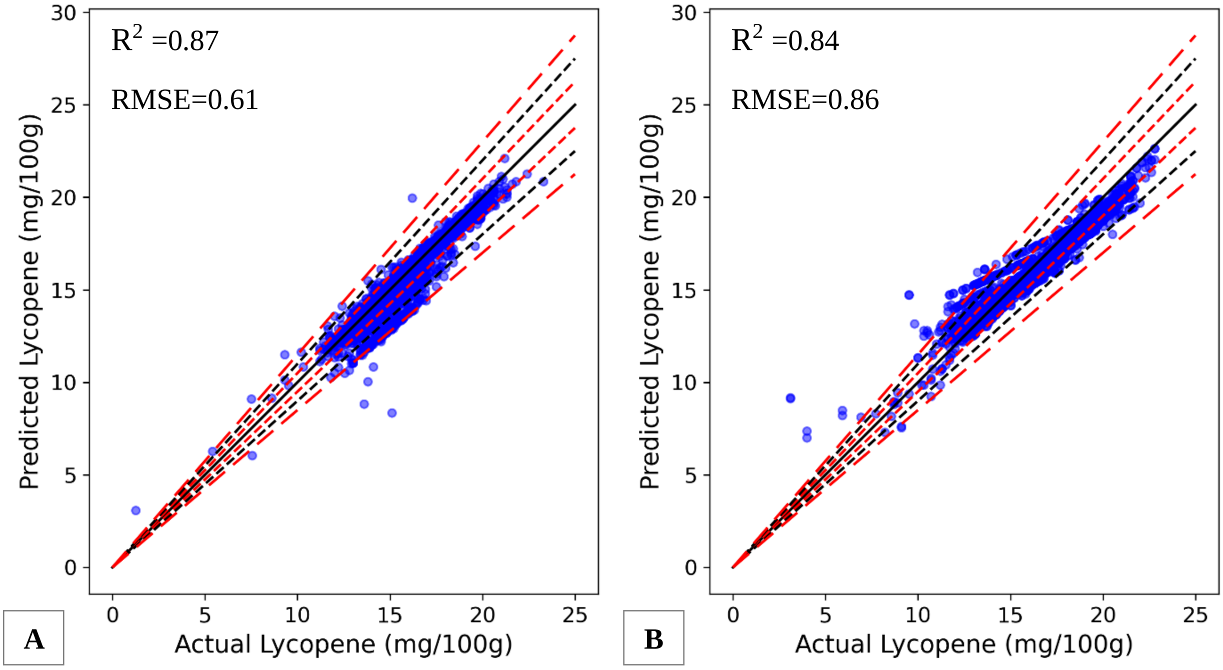


Figure 17. Actual vs Predicted Lycopene utilizing the XGBoost (A) and ANN (B) models. Black solid line indicates perfect prediction meaning that y = x, the red short-dashed lines, black dashed lines, and red long-dashed lines indicate ± 5, 10, and 15% deviation from the y = x line, respectively (n=4,746).

It is represented in the graphs, that a high degree of correlation was exhibited with predicted and actual Lycopene contents for both algorithms (Figure. 17). The XGBoost model yielded an R^2 value of 0.87 and an RMSE value of 0.61, accounting for 87% of the variance in observed Lycopene content. In contrast, the ANN model had an R^2 of 0.84 and an RMSE of 0.86, attesting to its substantial explanatory capability. While both models exhibited commendable accuracy in predicting Lycopene content, minor inconsistencies were observed. The line representing ideal prediction, where predicted values coincide with actual measurements, serves as a benchmark for accuracy. It was revealed, that a significant proportion of predictions from both models lied within the 10% deviation margin, underscoring their precision. More specifically, for the XGBoost model and the ANN respectively, the percentage of predictions deviating less than 5% were 84.55% and 86.45% and predictions that deviated between 5% and 10% were observed to be 10.31% and 10.28%. Those that fell between 10% and 15% deviation were 4.81% and 1.96%, and finally, predictions that deviated more than 15% were at 0.34% and 1.31% respectively.

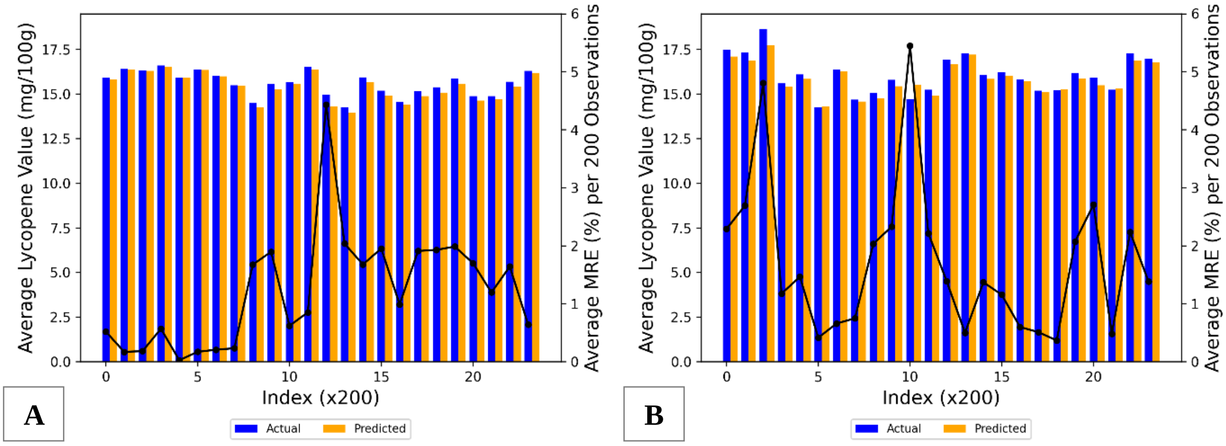


Figure 18. Comparison of MRE (black line) of actual (blue bars) vs predicted (orange bars) Lycopene values per 200 observations (Index x200) for XGBoost (A) and ANN (B) models.

The Mean Relative Error (MRE) graph in Figure 18 revealed fluctuations in prediction accuracy across the dataset. Comparatively, the XGBoost model demonstrated a more stable performance, with most data groups maintaining an MRE below 4%, suggesting generally robust predictive accuracy. On the other hand, the ANN model, as depicted in graph 5B, exhibited higher variability in its MRE, oscillating across different values and suggesting varying degrees of predictive accuracy. Notably, some segments exhibited a relatively high MRE, peaking just below 6%. While the bar representations of actual versus predicted Lycopene values in both graphs were closely aligned, indicating reasonable predictive capabilities, the XGBoost model presented slightly superior performance in terms of consistency and reduced error.

4.2.1.4 Model performance on a/b ratio

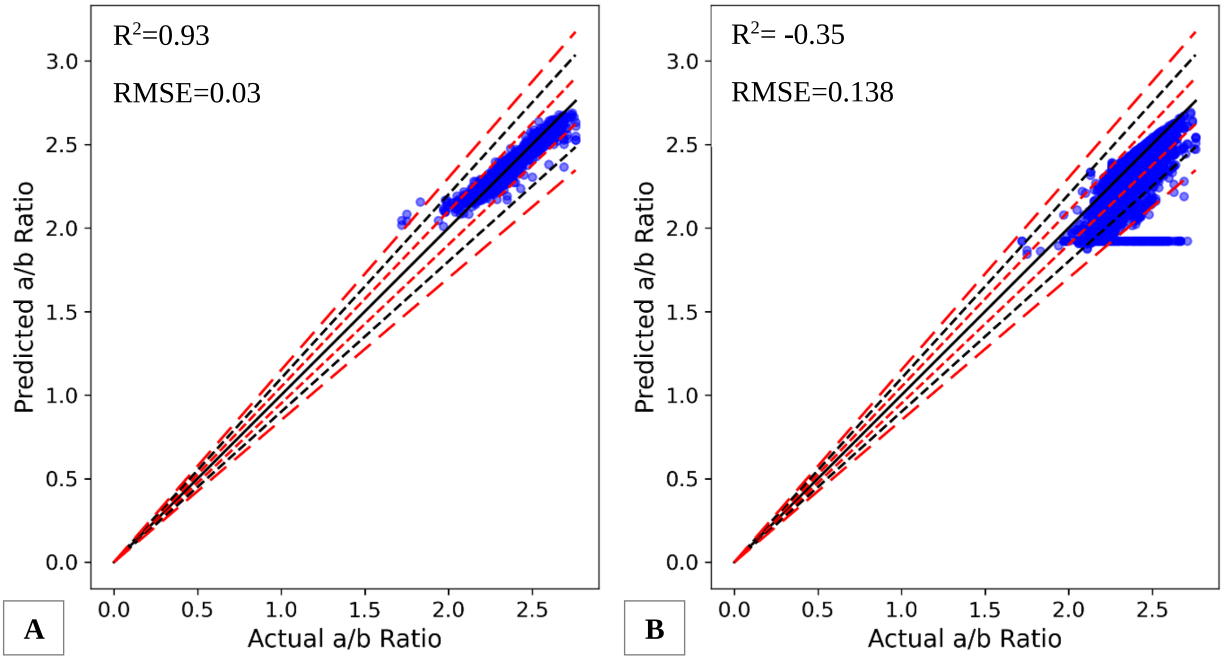


Figure 19. Actual vs Predicted a/b ratio utilizing the XGBoost (A) and ANN (B) models. Black solid line indicates perfect prediction meaning that y = x, the red short-dashed lines, black dashed lines, and red long-dashed lines indicate ± 5, 10, and 15% deviation from the y = x line, respectively (n=4,746).

The XGBoost model had demonstrated a high degree accuracy, achieving an R^2 value of 0.93 and an RMSE of 0.03, indicating a strong fit to the data (Figure 19A). In contrast, the ANN model had yielded a higher RMSE of 0.138. While this suggested a reasonable proximity of predictions to actual observations, the model's negative R^2 value of -0.35 indicated a poor fit to the dataset. This finding suggested that either the current ANN model was not optimal for this dataset, or there were underlying issues with the dataset or its processing. In terms of prediction deviation, for the XGBoost model, 99.45% of predictions had been within 5% of the actual values, while only 0.42%, 0.13%, and 0.00% had deviated by 5-10%, 10-15%, and over 15%, respectively. This indicated a high level of accuracy for most predictions. On the other hand, the ANN model had shown larger deviations: 81.29% of predictions had been within 5%, and 13.32%, 2.93%, and 2.47% had deviated by 5-10%, 10-15%, and over 15%, respectively. Notably, the ANN model had displayed significant deviations beyond the ±5% and ±10% margins (Figure 19B), suggesting areas of unreliability. It is worth noting that despite the moderate correlation observed in the ANN model, indicating a positive linear relationship between observed and predicted values, the negative R^2 value pointed to its failure in adequately fitting the variance in the data. This discrepancy underscored the importance of comprehensive evaluation metrics in model assessment. The RMSE of 0.138, while seemingly small, was significant if the dependent variable in the dataset exhibited low variability. This magnitude of RMSE reflected that the ANN model's predictions were, on average, 0.138 units away from the actual values, leading to consistent and notable inaccuracies. Thus, the practical utility of the ANN model in this context was limited, as evidenced by its negative R^2 value, despite a moderate correlation.

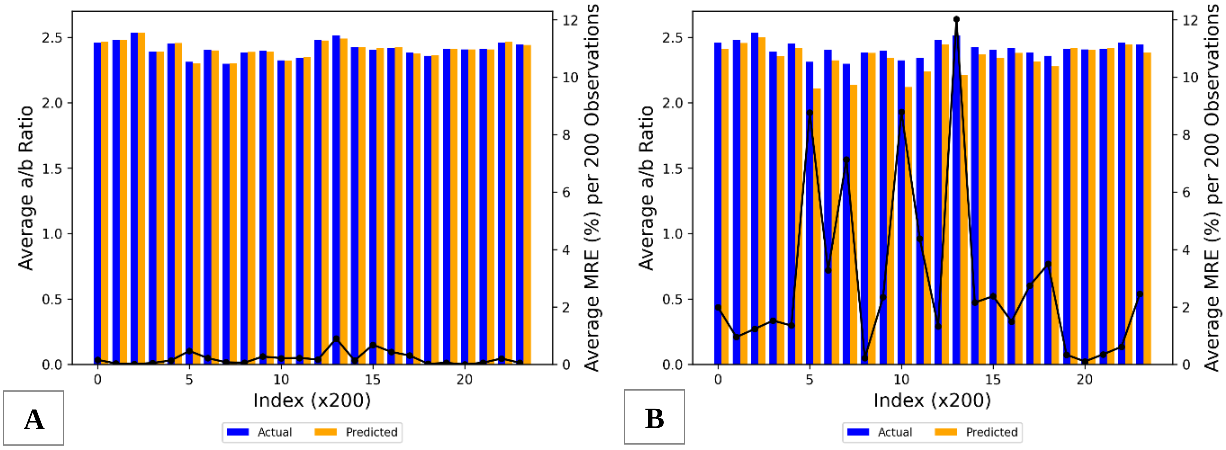


Figure 20. Comparison of MRE (black line) of actual (blue bars) vs predicted (orange bars) a/b ratio values per 200 observations (Index x200) for XGBoost (A) and ANN (B) models.

In our analysis, the XGBoost model demonstrated satisfactory predictive performance. Its MRE fluctuated but remained relatively low, peaking slightly above 0.8%, as illustrated in Figure 20A. In contrast, the ANN model exhibited significantly greater variability in its predictions. The MRE of the ANN model reached as high as approximately 12%, indicating that, on average, its predictions deviated by a maximum of 12% from the actual values. Although the bar representations of both actual and predicted a/b ratio values in the two graphs had suggested a decent level of predictive accuracy, the XGBoost model markedly outperformed the ANN model in terms of prediction fidelity and consistency.

4.2.1.5 SHAP

Brix

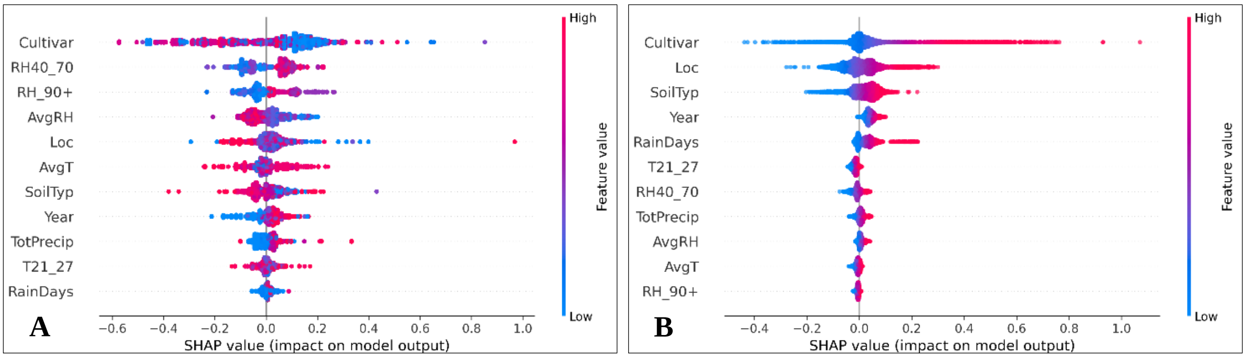


Figure 21. SHAP summary plot of Brix prediction for the XGBoost (A) and ANN (B) models.

Noticeable differences were observed in the importance of features and their effects on the models' predictions as a result of the conducted comparative analysis of the SHAP summary plots for the XGBoost and ANN models as depicted in Figure 21. The most important difference between the SHAP plots of the two-machine learning model was that positive feature values contributed to mainly positive SHAP values in the ANN model but they were sorted differently for the XGBoost. The 'Cultivar' feature was paramount in the XGBoost model, displaying a broad range of SHAP values that are both positive and negative values, indicating a robust association between certain cultivars and elevated Brix levels. This suggested the significance of genetic attributes in enhancing water soluble solids content. The features related to humidity, such as 'RH40_70' and 'AvgRH' showed a substantial spread of SHAP values across the x-axis, suggesting variable effects on Brix prediction, where both low and high relative humidity levels could either positively or negatively impact the accumulation of water-soluble solids in fruits, contingent upon other interacting variables. In contrast, in the ANN model the plot revealed a consistent pattern: higher feature values are invariably associated with positive SHAP values, while lower feature values correspond to negative SHAP values. This suggests a monotonic behavior where the magnitude of a feature's value is directly proportional to its impact on the model's output. The 'Cultivar' feature demonstrated a more uniform effect across the entire dataset, with a tendency toward positive contributions, reflecting its significant and consistent influence on the model's prediction of the Brix. Similarly, the SHAP values for 'Loc' and 'SoilTyp' indicate that geographical location and soil type are influential factors in predicting Brix levels, with higher and lower values of these features consistently impacting the model's output. The variable 'Year' also emerged as a significant temporal factor in the ANN model, potentially capturing the effects of varying climatic conditions across years, indicative of the model's capability to assimilate temporal dynamics into its predictive mechanism. The SHAP analysis showed that the XGBoost model attributed more importance to ‘AvgT’ than to ‘TotPrecip’, by contrast, the effect of ‘TotPrecip’ on the prediction of Brix was important in the ANN model. However, the ways in which these factors influenced Brix predictions in each model differed, possibly reflecting inherent differences in data assumptions and the models' strategies for integrating features.

Lycopene

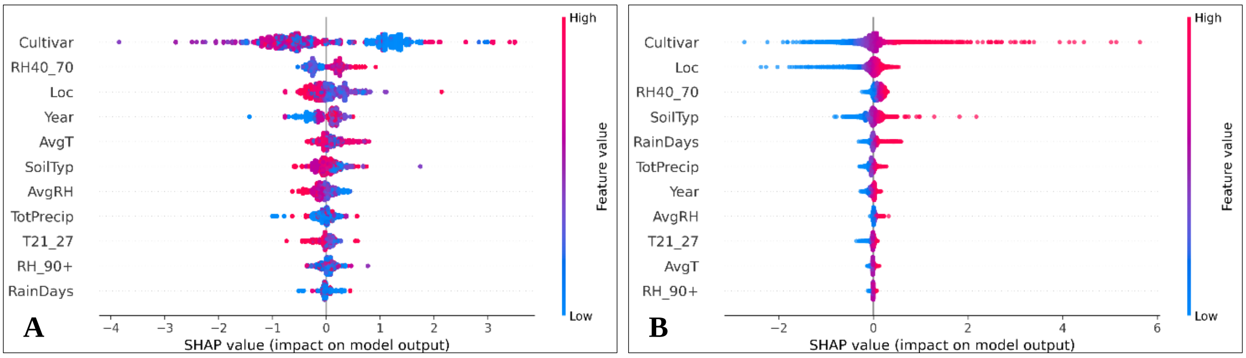


Figure 22. SHAP summary plot for Lycopene prediction for the XGBoost (A) and ANN (B) models.

The SHAP summary plots for the XGBoost and ANN models provided valuable insights into the determinants of lycopene content in tomato fruits (Figure 22). The analysis of the XGBoost model revealed that the 'Cultivar' and 'RH40_70' features had a significant impact on the model's predictions of Lycopene content. The 'Cultivar' feature, in particular, showed a wide spread of SHAP values, indicating that different cultivars had varying levels of influence on the Lycopene content prediction. This suggested a complex, potentially non-linear relationship with the target variable. 'RH40_70' showed a more concentrated range of SHAP values, suggesting a consistent but less influential effect on the model's predictions. Other features were represented with SHAP values clustered closer to the center, implying a more moderate impact on the Lycopene content prediction. For the ANN, the 'Cultivar' feature exhibited the most substantial influence on the model's output with a broad spread of dots, indicating that the influence was more positive than negative. This implied a complex interplay where certain cultivars could have had a substantial impact, either augmenting or diminishing the potential Lycopene content determined by genetic background. Although the general directionality of feature values and their impact on the model’s predictions might have suggested a monotonic pattern, the spread and distribution of the SHAP values did not necessarily imply a linear relationship but rather implied a consistent pattern recognized by the neural network where certain features were favorable for Lycopene production. The colour gradient added another layer of interpretability. For instance, the XGBoost plot showed that both high and low values of 'AvgT' did not exhibit simple linear relationships with Lycopene content. Instead, its impact was nuanced, with both high and low values influencing predictions in both positive and negative directions. This complexity may have mirrored how biological processes formed agricultural crops in response to environmental factors. Additionally, temporal trends reflected in the 'Year' feature's SHAP values could have pointed to evolving agricultural practices or climatic shifts over time, further highlighting the multifaceted nature of Lycopene biosynthesis.

a/b Ratio

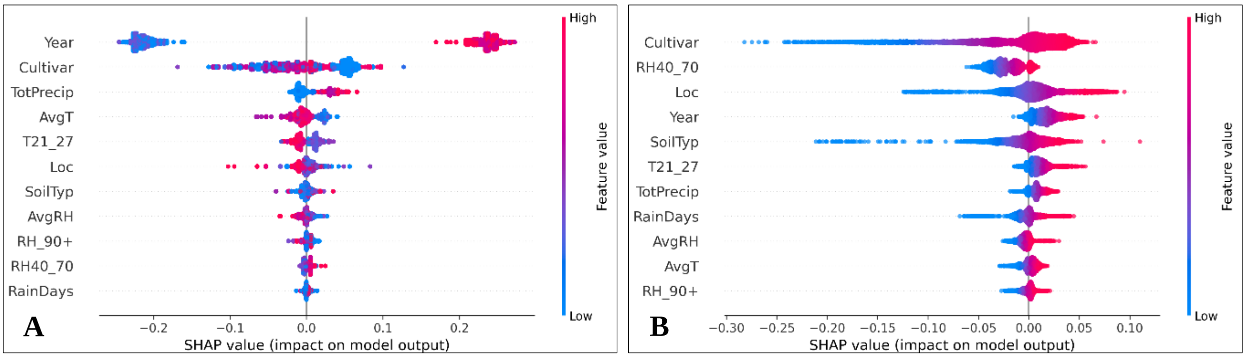


Figure 23. SHAP summary plot for a/b ratio prediction for the XGBoost (A) and ANN (B) models.

Examining the SHAP summary plots of the two machine learning models that had been designed to predict tomato fruit colour values, particularly the chromaticity ratio (a/b ratio), distinct patterns of feature influence had emerged (Figure 23). The 'Year' feature in the XGBoost model had exhibited a high distancing of SHAP values, with clusters on both the positive and negative sides of the zero line, indicating a variable influence on the model’s prediction, with some years contributing to an increase and others to a decrease in the predicted a/b ratio. The 'Cultivar' feature exhibited a unidirectional effect, with a pronounced aggregation of its SHAP values on the positive side, indicating a uniform contribution to the increase in the model’s predicted a/b ratio. Notably, this increase is predominantly associated with the lower encoded values of 'Cultivar,' as indicated by the abundance of blue points. Conversely, 'TotPrecip' was predominantly associated with decreases in the a/b ratio, suggesting a positive relationship. For the ANN model, interpreting the SHAP values became more challenging due to the negative R^2 score. The model had predominantly exhibited negative SHAP values for features such as 'Cultivar', 'SoilTyp', and 'RH40_70'. These consistently downward predictions indicated that these features often reduced the predicted value compared to the model's baseline. The dominance of negative SHAP values and the lack of variation in SHAP value direction, unlike the variability observed in the XGBoost model, raised concerns about potential overfitting, insufficient feature representation, or inadequate network architecture to capture the complexities of the dataset. Furthermore, the ANN’s poor performance metric, as highlighted by the negative R^2 score, had implied that the model was less informative than a simple average of the target variable, suggesting that the model's internal representations and learned weights did not generalize well to the data's underlying structure.

4.2.2 Discussion

4.2.2.1 Correlation Heatmap

The correlation heatmap provided an invaluable visual summary of the intricate interrelationships among climatic variables, Brix, Lycopene, and a/b ratio in tomato fruits. The strong positive association between 'AvgT' and 'T21_27' underscored the synchronicity of average seasonal temperatures with the frequency of days experiencing temperatures between 21 and 27°C. This relationship is pivotal, as temperatures within this range were known to be conducive for the optimal growth of tomato plants and could influence various biochemical processes, including the synthesis of sugars and pigments (Ayankojo & Morgan, 2020). Necessarily, close alignment was found between 'TotPrecip' and 'RainDays', affirming the notion that seasons with more accumulated rainfall were characterized by a higher number of rainy days. Excessive rainfall, especially during the fruit development stage, could influence fruit texture, water content, and even lead to conditions like fruit cracking (Bihon et al., 2022). As it was expected, inverse correlation was observed between 'AvgRH' and 'RH40_70' and could be indicative of specific climatic patterns affecting the impact of certain stresses. A season with consistently high humidity might have had fewer fluctuations, resulting in fewer days with humidity levels within the 40% to 70% range. Such patterns could influence plant transpiration rates, nutrient uptake, and susceptibility to certain diseases (Chowdhury et al., 2021). High humidity levels might reduce transpiration rates, leading to an accumulation of sugars in the fruit, thereby elevating the Brix values (Zheng et al., 2022), however, no correlation was found between Brix and RH_90+. It's well-established that external factors can modulate the synthesis of pigments and antioxidants in tomatoes (G. P. P. Lima et al., 2022; Vela-Hinojosa et al., 2019). By contrast, there was no significant correlation revealed between climatic factors and Brix or lycopene content. Merely the a/b ratio correlated significantly with T21_27, and moderate relationships were indicated with AvgT, RainDays, and RH_90+.

4.2.2.2 Brix Prediction

The prediction of water-soluble solids content, which is an important quality trait for the food and beverage sector, was effectively handled by our machine learning models (Jaywant et al., 2022). The XGBoost model demonstrated slightly superior performance, attributed to its gradient boosting mechanism which effectively handles linear and non-linear relationships, missing values, outliers, and diverse data types. Conversely, the ANN showcased robustness in capturing intricate patterns in multi-dimensional data, its performance in predicting Brix values, though substantial, suggested limitations in capturing certain complexities, unlike XGBoost. This research built upon previous findings, such as Silva et al. (2017) study, who used a global climate model and highlighted the significant impact of extreme climatic conditions, like increased heat and dry stress, on tomato quality. These conditions were crucial factors that could potentially enhance the accuracy of machine learning predictions. Complementing this, Zuo (2022) demonstrated the use of visual datasets in tomato quality grading using machine learning and image processing, and Égei et al. (2022) revealed the efficacy of Vis-NIR spectroscopy in determining soluble solids content applying partial least square regression (PLSR) model obtaining R^2 of 0.72 and 0.88 for calibration and validation, respectively. Notably, our models, derived from climatic and environmental data using more cost-effective methods, amplified their potential for broader, non-destructive applications. The significance of this approach was highlighted by comparing it with earlier works; for instance, Ecarnot et al. (2013) reported an R^2 of 0.86 using a portable VIS-NIR spectrometer for rapid assessment of tomato Brix, whereas our refined machine learning approaches demonstrated greater precision. Additionally, the non-destructive Brix prediction model by Gomes et al. (2014, 2017) showed an R^2 of 0.95 and RMSE of 1.34 using PLSR, and an R^2 of 0.91 and RMSE of 1.36 using principal component analysis (PCA), underscoring the enhanced efficacy of our machine learning methods especially regarding RMSE. Ultimately, the significant aggregation of predictions within the 5% error margin for both models highlighted their practical value for predicting tomato quality in relation to climatic conditions, demonstrating their potential for aiding in long-term agricultural planning and ensuring consistent product quality over time. The minimal inaccuracies observed, particularly for values scattered in brackets with higher error, further attested to the robustness and reliability of these models.

4.2.2.3 Lycopene Prediction

Lycopene content is a pivotal component in determining the nutritive and organoleptic qualities of tomatoes. In our analysis, both the XGBoost and ANN models demonstrated a significant positive correlation between the predicted and actual values of Lycopene content, with R^2 values of 0.87 for XGBoost and 0.84 for ANN. These figures indicated a strong correlation, aligning with previous studies that highlighted the effectiveness of machine learning in agricultural data analysis (Attri et al., 2023; Ayaz Mirani et al., 2021; Liakos et al., 2018). The XGBoost model, traditionally renowned for handling structured/tabular data (Chen & Guestrin, 2016), showed slightly better performance with an RMSE of 0.61, compared to the ANN model's RMSE of 0.86. This can be attributed to its scalability and capability to handle various types of prediction problems, including its resilience against overfitting and ability to implicitly handle missing values. Conversely, ANNs are known for their versatility in handling complex, non-linear data patterns (Almeida, 2002; Isaac Abiodun et al., 2018). Although the ANN model here showed a marginally lower precision than XGBoost, it is important to consider that its performance can be influenced by factors such as architecture design and the number of layers. The high percentage of predictions within 10% deviation from actual values (84.55% for XGBoost and 86.45% for ANN) underscored the practical applicability of these models in precision agriculture, particularly for quality control and breeding programs (Bdr et al., 2020). Liu et al. (2015) utilized methods including partial least squares (PLS), least squares-support vector machines (LS-SVM), and back propagation neural network (BPNN) to predict Lycopene content from spectral data, reporting R^2 values of 0.50, 0.91, and 0.93, respectively. Similarly, Sharma et al. (2023) used Linear Multivariate Regression (LMVR) to predict Lycopene content in tomatoes using physicochemical attributes, achieving an R^2 of 0.7. These findings highlighted the enhanced capabilities of modern XGB and ANN models in accurately predicting Lycopene content. Despite the impressive performance of our models, it is crucial to acknowledge that all predictive tools are subject to inherent limitations. Factors such as sample diversity, experimental conditions, and algorithmic assumptions can affect their precision. Ultimately, both the XGBoost and ANN models demonstrated significant potential for predicting Lycopene content. However, due to its simplicity and proven track record, the XGBoost model emerges as the more favourable choice in our study.

4.2.2.4 a/b Ratio Prediction

In our study, the comparison between the XGBoost and ANN models in predicting the a/b ratio in tomato cultivars offers significant insights. The XGBoost model, known for its gradient boosting framework and ability to manage varied datasets (Chen & Guestrin, 2016), demonstrated a substantial advantage. It not only showed higher accuracy, as evidenced by an impressive R^2 value of 0.93 and a minimal RMSE of 0.03, but also greater consistency in predictions. This indicates the model's robustness in capturing the complex interplay of climatic and soil parameters, benefiting from its adaptability and regularized boosting technique. Conversely, the ANN model's performance was not satisfactory. It exhibited a negative R^2 value of -0.35 and a higher RMSE of 0.138, suggesting significant issues in its fit to the dataset and potential problems such as overfitting, inadequate training, or a mismatch in model complexity (Draper & Smith, 1998; Goodfellow et al., 2016). The negative R^2 value suggests that the model's predictions were worse than a simple mean of the observed data, raising questions about its suitability for this application. Additionally, the discrepancy between RMSE and R^2 could be due to RMSE's sensitivity to outliers, while R^2 reflects the overall variance explained (James et al., 2013). Furthermore, the prediction deviation analysis underscored the XGBoost model's reliability, with 99.45% of its predictions within a 5% margin of the actual values, demonstrating its utility for precision-dependent applications (Chen & Guestrin, 2016). In contrast, the ANN model showed larger prediction deviations, with only 81.29% of predictions within the same margin, highlighting its limitations in high-precision applications (Goodfellow et al., 2016). The significance of RMSE in datasets with low variability becomes particularly noteworthy; even a seemingly small RMSE in the ANN model indicates consistent and notable inaccuracies (Hyndman & Koehler, 2006). Moreover, the ANN model's negative R^2 value underlines a fundamental inadequacy, suggesting its inefficiency compared to even basic mean-based prediction models (James et al., 2013).

4.2.2.5 SHAP

Recognizing the critical role of interpretability in agricultural applications, we extended our analysis to include SHAP value computations. The SHAP summary plots of the two machine learning models revealed the influence of different features on the prediction values. These plots served as interpretable visual aids that can elucidate the complex inner workings of these models, especially in a domain that requires a nuanced understanding of the interplay between multiple factors (Li, 2022).

Brix

Our analysis revealed distinguishing features between the XGBoost and ANN models in predicting Brix values in tomato fruits, aligning with previous research that highlights the sensitivity of machine learning models to feature selection and interaction (Suresh et al., 2022). The prominence of 'Cultivar' in the XGBoost model echoed the findings of Aldrich et al. (2010) and Rusu et al. (2023), who reported the genetic makeup of a cultivar as a decisive factor in fruit Soluble Solids Content. The positive SHAP values associated with 'Cultivar' suggest that certain genetic characteristics may be important drivers of Brix levels, potentially offering a pathway for targeted breeding programs (Beckles et al., 2012; Prinzenberg et al., 2021; Vallarino et al., 2020). The variable impacts of relative humidity observed in our study are consistent with the results published by Shin et al. (2008), which demonstrated the complex roles of relative humidity in tomato fruit development and ripening. Our findings suggest that not only the range but also the duration of specific humidity levels could be critical, warranting further investigation into their interactions with other environmental factors. In contrast, while ANNs are inherently equipped to model complex, non-linear interactions (Amiri et al., 2022; Aziz et al., 2018), the monotonic behavior observed in the SHAP plot suggests that, the model may be capturing more direct, additive relationships between features and the Lycopene content for the given dataset. Such an observation suggests that the neural network, has adapted to the dataset's structure by identifying and leveraging what appears to be a straightforward linear association of features with the target variable. The distributions of SHAP values for 'Loc' and 'SoilTyp' underscored the potential for ANN models to discern subtle influences of edaphic and geographical factors, aligning with the studies by Liu et al. (2021) and Xu et al. (2023), which posited that soil characteristics could profoundly affect fruit quality. For instance, certain soil types may be consistently beneficial or detrimental to the dissolved sugar content, depending on their nutrient profiles or water retention capacities. The role of the 'Year' variable in capturing annual climatic variations provided an intriguing insight into the temporal dynamics affecting Brix levels. As suggested by Montgomery & Biklé (2021) and Raza et al. (2019), shifts in agricultural practices, adoption of new technologies, or even changing climate patterns can manifest in fluctuations in the quality and nutritional content of crops. The distinct influences of meteorological factors observed in our study add to a growing body of evidence that suggested weather conditions played a pivotal role in water soluble solids content, a notion supported by the comprehensive analysis of climate impacts on fruit nutrition value by Stewart & Ahmed (2019). While our results provided valuable contributions to predictive modeling in agriculture, they also emphasize the importance of considering the specific model's interpretive framework. The differences in feature importance between the XGBoost and ANN models could reflect the indicate the fundamental differences in their data processing methodologies, as noted by Lima et al. (2021). This underlines the importance of interpretability and reliability in machine learning models, especially in domains where decision-making is closely tied to model outputs (Ahlquist et al., 2023).

Lycopene

The XGBoost model, except for the 'Cultivar' and 'RH40_70' features, demonstrated a balanced feature influence with tight SHAP value clustering, suggesting a nuanced consideration of feature contributions, akin to findings by Lundberg & Lee (2017) on interpretable machine learning models. Notably, the 'Cultivar' variable had stood out as a significant determinant with a complex and non-linear influence on Lycopene content, line with the research of Tsai et al. (2020), which reported the subtleties of genetic factors in crop quality predictions. Furthermore, were in agreement with the work Bineau et al. (2022), documenting the genetic diversity among tomato cultivars and its impact on the accumulation of secondary metabolites. However, the broad distribution of SHAP values for the 'Cultivar' feature within the ANN model likely signifies the model's ability to capture complex, non-linear interactions between this feature and the Lycopene, an aspect that mirrors the observations made by Wang et al. (2020) regarding the capabilities of deep learning in capturing intricate biological phenomena. The spread of SHAP values for environmental features like 'RH40_70' and 'Loc' underscored the multifactorial nature of Lycopene synthesis, as suggested by Panthee et al. (2012), emphasizing the critical roles of both genetic and environmental factors. This is further supported by research from Kuti & Konuru (2005), Guerra et al. (2021), and Srivastava & Srivastava (2015), who noted the influence of specific environmental conditions on Lycopene synthesis and preservation. The nuance influence of the 'AvgT' on the Lycopene content prediction potentially indicates adaptive physiological responses to environmental stresses, aligning with the work of Ahanger et al. (2017) on plant stress biology, where extreme temperature could be associated with either higher or lower Lycopene content. Temporal variability in Lycopene content, signified by the 'Year' SHAP values, could be indicative of the dynamic interplay between cultivation methods, environmental shifts, and plant genetics over time. This observation aligns with the longitudinal studies by Arah et al. (2016), highlighting the evolutionary trajectories in agricultural practices and post-harvest handling techniques. While these insights were compelling, a potential risk of overfitting with the ANN model, as indicated by the extensive spread of SHAP values, must be acknowledged. Further validation with independent datasets, as recommended by Kuhn & Johnson (2013), would be necessary to confirm the robustness of the findings. Additionally, integrating multi-omics data, as discussed by Kang et al. (2022), could enhance the interpretability of the predictive models, offering a more holistic view of the factors influencing the Lycopene content.

a/b ratio

The observed variability in the SHAP values for the 'Year' feature within the XGBoost model aligns with previous research that indicates temporal dynamics can significantly affect agricultural outcomes (Amankwah, 2023). The dispersion suggested that the impact of 'Year' on the a/b chromaticity ratio was not linear and may be influenced by other interacting factors, such as changing climate conditions or agricultural practices over time (Naeem et al., 2023; Pathak & Stoddard, 2018; Quinet et al., 2019). The 'Cultivar' feature's consistent influence on increasing the a/b ratio, particularly at lower feature values, confirmed the importance of genetic factors in determining tomato fruit colour (Zhao et al., 2021). The positive pronounced aggregation of the SHAP values reflected a potentially strong genotype-phenotype relationship, which has been widely documented in crop quality traits (Cobb et al., 2013). In contrast, the 'TotPrecip' feature's association with the a/b ratio may be indicative of the dilution effect of precipitation on fruit colour concentration, a finding that was supported by the work of Oh et al. (2010), who noted that water availability could lead to the dilution of phytochemicals in fruits. Additionally, precipitation can modulate physiological processes in plants, impacting the synthesis and accumulation of pigments responsible for colour, which in turn affected the a/b ratio as it was supported by the work of Kim et al. (2022). The ANN model's predominantly negative SHAP values and the accompanying negative R^2 score presented a stark contrast to the XGBoost model and raise questions about the ANN's suitability for this task. This finding is particularly concerning regarding the increasing reliance on ANN models in precision agriculture (Condran et al., 2022). The consistent underperformance, as indicated by the negative R² score, may be due to overfitting, which is a common challenge with ANN models (Salman & Liu, 2019), and suggested that the network architecture may have not been adequately optimized for the dataset. The lack of variation in the direction of SHAP values for the ANN model contrasted sharply with the XGBoost model and suggested that the former may not be capturing the true underlying data patterns. This discrepancy emphasizes the need for a thorough cross-validation and hyperparameter tuning process, which has been identified as a crucial step in model development (Bates et al., 2021; Jin, 2022). Furthermore, the negative R² score suggests that the ANN model's predictive power was worse than a naïve model that would simply predict the average a/b ratio for all observations (Wray et al., 2013).

4.2.3 Future Work and Recommendations

4.2.3.1 Potential Improvements to Models and Preprocessing Steps

To address limitations in our current models, a more comprehensive approach to data collection and diversity is could be implanted. By incorporating a broader spectrum of climatic variables, such as light intensity and quality and wind speed, we can provide a deeper understanding of environmental impacts on tomato quality (Xiao et al., 2022; Yu et al., 2022). Furthermore, by expanding the dataset to include a wider variety of tomato cultivars, like heirloom or more hybrid varieties, we could allow for a more robust analysis of genetic factors influencing Brix, Lycopene, and a/b ratio (Bai & Lindhout, 2007; Tripodi et al., 2023).

Using advanced data preprocessing methods, such as feature scaling normalization (Kuhn & Johnson, 2013), and non-linear transformations (Duraivel et al., 2023), can significantly improve our model’s accuracy. Additionally, the incorporation of anomaly detection methods (Chandola et al., 2009) can help in identifying and handling outliers, ensuring the reliability of the models.

For the ANN model, especially in predicting the a/b ratio, recalibration is needed. Investigating various neural network architectures, like deeper networks or recurrent neural networks, might help us better capture temporal and complex interactions (Goodfellow et al., 2016). Additionally, experimenting with different activation functions, such as leaky rectied linear function (LReL) (Maas et al., 2013), or optimization algorithms (Ruder, 2016), may also enhance the model's performance. In the same way, for the XGBoost model, optimizing hyperparameters like the learning rate, tree depth, and the number of trees can improve more its performance (Chen & Guestrin, 2016). Exploring feature interaction constraints (Oh, 2019) could be useful to understand complex data relationships better and improve the model's performance.

4.2.3.2 Suggestions for Other Algorithms or Methodologies

In agricultural research, particularly in predictive modeling, the exploration and implementation of diverse algorithms and methodologies hold significant potential. The idea of hybrid models, notably the combination of XGBoost and ANNs, could offer a promising research path. Such model could effectively integrate the feature interactions captured by tree-based algorithms with the complex pattern recognition abilities of neural networks. This approach aligns with ensemble techniques, as suggested by Shahhosseini et al. (2022), where combining multiple model predictions, such as a weighted ensemble of XGBoost and ANN, enhances both stability and accuracy. Moreover, the concept of model stacking, introduced by Wolpert (1992), involve using the outputs of XGBoost and ANN as inputs for a secondary model, possibly a simpler regression model, to enhance the accuracy of predictions further.

Deep learning, known its capability to handle large and complex datasets, stands as an effective strategy, for capturing nonlinear interactions between environmental, genetic, and temporal factors. Convolutional Neural Networks (CNNs) for example, could be used to analyse satellite or field imagery in order to assess crop health and predict quality traits (Kamilaris & Prenafeta-Boldú, 2018). Additionally, Recurrent Neural Networks (RNNs), especially LSTM (Long Short-Term Memory) networks, could be effective in modeling sequential data, such as time-series climatic data, to predict crop quality attributes (Nketiah et al., 2023).

Unsupervised learning algorithms are not only capable of analysis but also of evaluation. Clustering techniques, like K-means or hierarchical clustering, could give insights into sub-populations or environmental conditions within agricultural data, as indicated by Shahid (2023). Principal Component Analysis (PCA) can help reducing dataset complexity, highlighting key features, and boosts model efficiency and interpretability as described by Rahmat et al. (2023).

4.2.3.3 Recommendations for Practical Application Based on Results

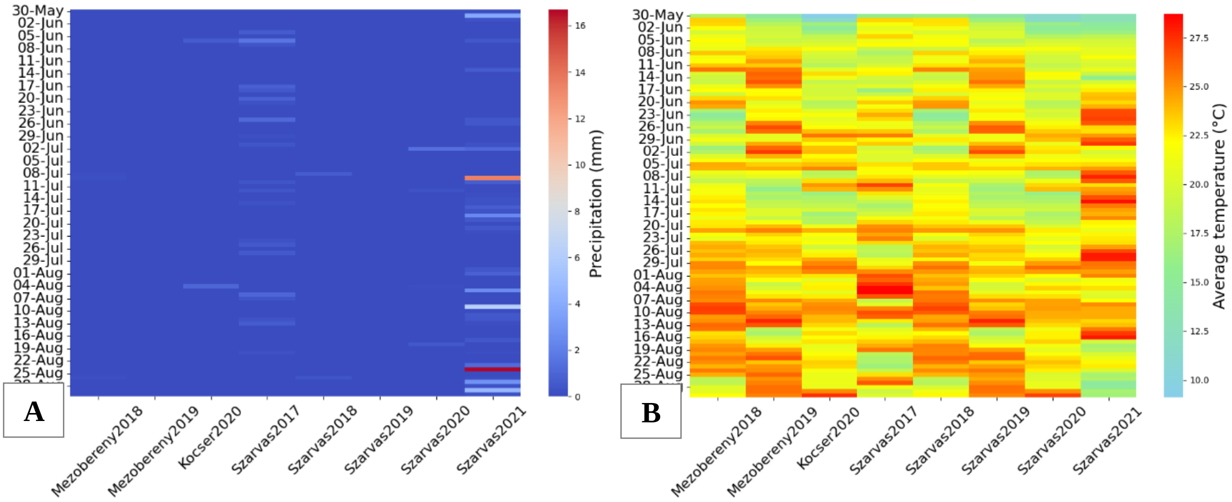
Machine learning in agriculture plays a pivotal as a decision-support tools, helping farmers in selecting appropriate cultivars and optimizing planting schedules by predicting important factors like Brix and Lycopene content. This aligns with the findings of Lobell & Gourdji (2012), who highlighted the importance of predictive models in crop selection and agricultural productivity. Moreover, considering the influence of climatic variables, these models can assist in adapting farming practices to changing weather patterns. Tools developed from these models can predict the impact of anticipated climatic changes on crop quality, thereby aiding in the development of proactive strategies, a concept reinforced by Ray et al. (2015) who emphasized the importance of climate-adaptive agricultural practices. Additionally, the substantial impact of climatic factors on tomato quality highlights these models' potential in studying the broader effects of climate change on agriculture. Researchers and policymakers can use these models to project future trends in crop quality under various climate scenarios, aiding in formulating mitigation strategies. This aspect is supported by Challinor et al. (2014), who emphasized the importance of modeling in understanding climate change impacts on agriculture.

These findings highly valuable in the food industry as they can serve the development of non-destructive quality assessment tools, especially for assessing Brix content, which is essential for ensuring taste and quality, as shown in Mendoza et al. (2011) studies. Additionally, predictive models also play a crucial role in maintaining product consistency, a piotal factor for consumer satisfaction and brand reputation, by adjusting processing parameters, a point highlighted by Akimov et al. (2021). Furthermore, the models from this study hold a promise potential application beyond tomatoes. They could advantage other crops, leading to a deeper understanding and optimization of quality parameters across various agricultural products, as supported by Liakos et al. (2018) in showcasing the diverse applications of machine learning in agriculture.

4.3 Part 3: Evaluation of Tomato Plant Genetic Resources for Brix and Lycopene in Different Environments

4.3.1 Results

4.3.1.1 Weather conditions



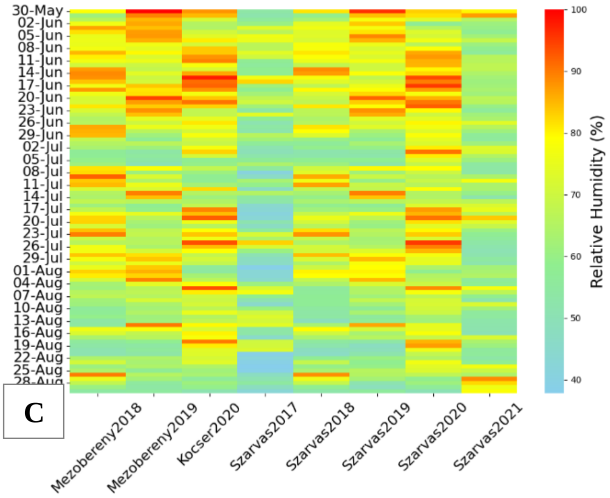


Figure 24. Heatmap Analysis of Meteorological Data from the 8 Selected Environments. (A: Precipitation, B: Average Temperature, and C: Relative Humidity).

Weather data are displayed in Figure 24. The precipitation heatmap revealed a pronounced dryness across all environments, with the exception of Szarvas2021, which received only a total of 63.5 mm throughout the growing season. In stark contrast, the remaining environments experienced severe aridity, receiving no more than 11 mm of precipitation at best, indicating exceptionally dry conditions that persisted throughout the growing seasons. Temperature trends further underscored the challenging climatic conditions for agriculture. Until the final third of July, the incidence of days with high temperatures (25°C≤) was remarkably low, signifying a predominance of cooler days during the initial phases of the growing season. However, a marked shift occurred post-July, with a majority of days registering high daily temperatures. Moreover, the analysis of daily relative humidity highlighted additional environmental stressors. Generally, days with the highest relative humidity were concentrated in the first third of the growing season. Notably, Szarvas2017 emerged as the environment with the lowest relative humidity, where many days recorded humidity levels below 40%, as depicted in the relative humidity heatmap. This observation points to considerable variations in moisture availability, with Szarvas2017 experiencing conditions that could exacerbate the effects of drought and high temperature on plant growth and development.

These meteorological insights reveal the extent of climatic adversities faced during the growing seasons, characterized by extreme dryness, fluctuating temperatures, and variable relative humidity levels, which collectively pose significant challenges to agricultural productivity in the studied environments.

4.3.1.2 ANOVA

Table 1. Analysis of variance for Brix and Lycopene for the 6 Hungarian tomato commercial varieties cultivated in 8 environments.

Source	**Df**	**Brix**		**Lycopene**	
		MS	**ESS (%)**	**MS**	**ESS (%)**
Genotype	5	15.13**	59.55	718.0**	43.42
Environment	7	59.26**	10.85	447.0**	37.84
Genotypes * environment	35	5.89**	29.59	44.3**	18.74
Residuals	5149	0.16		2.3	

DF=Degree of freedom. **MS=Mean squares.**

ESS=Explained Sum of Squares. **** Significant at P <0.01.**

The analysis of variance (ANOVA) conducted to investigate the effects of genotype and environment on Brix and Lycopene content (as shown in Table 1) revealed significant effects for both factors in influencing these traits (p <0.01). This indicates notable variability in Brix and Lycopene content attributable to different genotypes and environmental conditions. Additionally, the interaction between genotype and environment (Genotype * Environment) was also statistically significant for both Brix and Lycopene (p <0.01), suggesting that the effect of genotype on these parameters is influenced by environmental factors and vice versa. In terms of explained variance, the genotype had a more pronounced impact on Lycopene content, explaining 43.42% of the variance, compared to its influence on Brix. On the other hand, environmental factors were more influential for Brix, accounting for 59.55% of the explained variance. The interaction between genotype and environment contributed to 29.59% of the variance in Brix and 18.74% in Lycopene. These findings highlight the complexity of the relationships between genotype, environment, and these two important agricultural traits. The presence of significant genotype-environment interactions (GEI) highlights the need for further analysis to identify genotypes with stable and high Brix and Lycopene content across different environmental conditions.

4.3.1.3 GGE biplot analysis

The stability analysis of the 6 tomato genotypes using GGE biplot displayed the genotype main effect (G) and the GEI, which are the two most important sources of variation for genotypes evaluation in a multi environment trials (Yan et al., 2007).

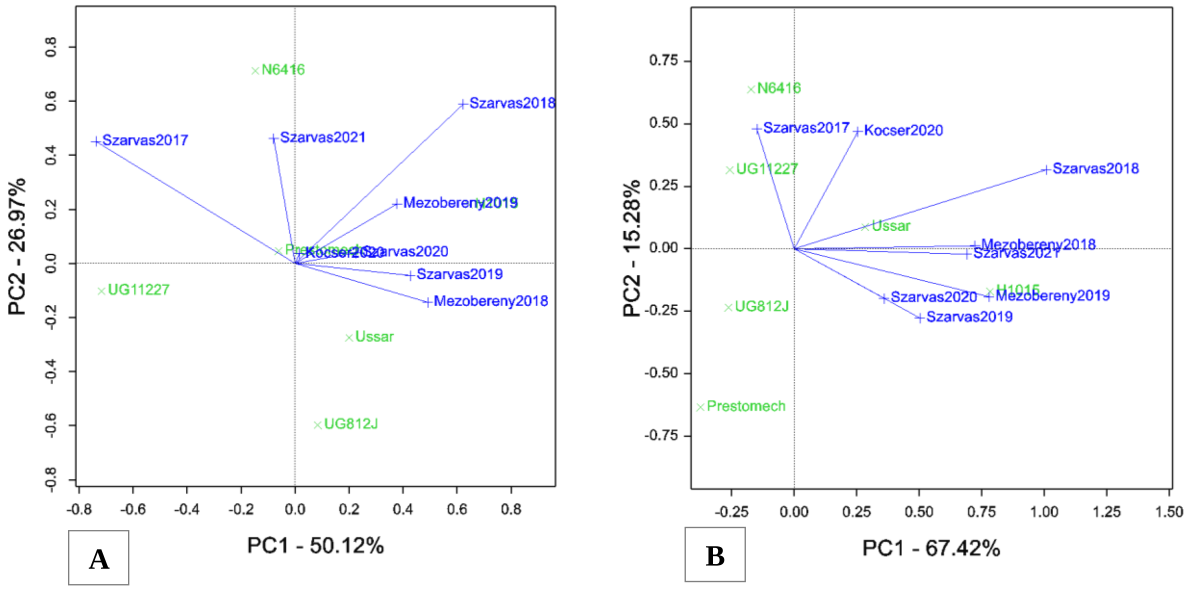


Figure 25. GGE biplot of Brix (A) and Lycopene (B) of 6 tomato genotypes planted in 8 environments.

Figure 25 shows that the first two principal PCs explain 77.10% (PC1 = 50.12%, PC2 = 26.97%) and 82.71% (PC1 = 67.42%, PC2 = 15.28%) of the total variation of the GGE model respectively for the Brix and Lycopene values. PC1 defines the mean performance of the genotype, while PC2 shows the GEI of each variety, which is a measure of variability (stability). The origin of the graph represents a virtual genotype that has an average performance in each environment.

Focusing on individual genotype performance, UG812J and Ussar, with PC1 values above zero, demonstrate high Brix and good adaptability. Prestomech, uniquely positioned near the biplot origin along PC2, demonstrates consistent stability across various conditions, although it exhibits average Brix content. Conversely, genotypes with PC1 values below zero show the opposite trend. The biplot's lack of clustering highlights significant environmental and genotype variation. Moreover, the acute angles observed between Szarvas2017 and Szarvas2021 indicate a positive correlation between each of these environments. This pattern repeats between the environments Szarvas2018, Szarvas2019, Szarvas2020, Mezobereny2018, Mezobereny2019. However, obtuse angles observed between Szarvas2017 and Szarvas2021 one side, and Szarvas2019 and Mezobereny2018 on the other side indicate a negative correlation between these environments. Same thing was observed when comparing Szarvas2020 with Szarvas2021 and Szarvas2019. A right angle between Szarvas2017 and Szarvas2018 indicates no correlation between these environments.

Regarding the Lycopene content, only H1015 and Ussar outperform the average and show good adaptability. The biplot also suggests environmental similarities, with Szarvas2017 and

Mezobereny2018 clustering together, indicating shared attributes. This pattern is repeated for Szarvas2019, Szarvas2020, and Mezobereny2019. All environments except for Szarvas2017 exhibit an acute angle between each pair of environments, indicating a positive correlation between them. However, obtuse angles between Szarvas2017 and the trio of Szarvas2019, Szarvas2020, and Mezobereny2019 suggest negative correlations between these environments.

4.3.1.4 The Which-Won-Where patterns

The Which-Won-Where patterns of the biplot help us to identify which genotype perform the best in each environment and each mega-environment. A polygon called A Convex Hull is constructed by connecting the genotypes furthest from the biplot origin by straight lines, and thus all other tested genotypes are contained in the polygon. Perpendicular lines starting from the biplot origin to each side of the polygon divide the biplot into environment sectors. Mega environments are ellipses around the environments within the same sector (Yan & Tinker, 2006).

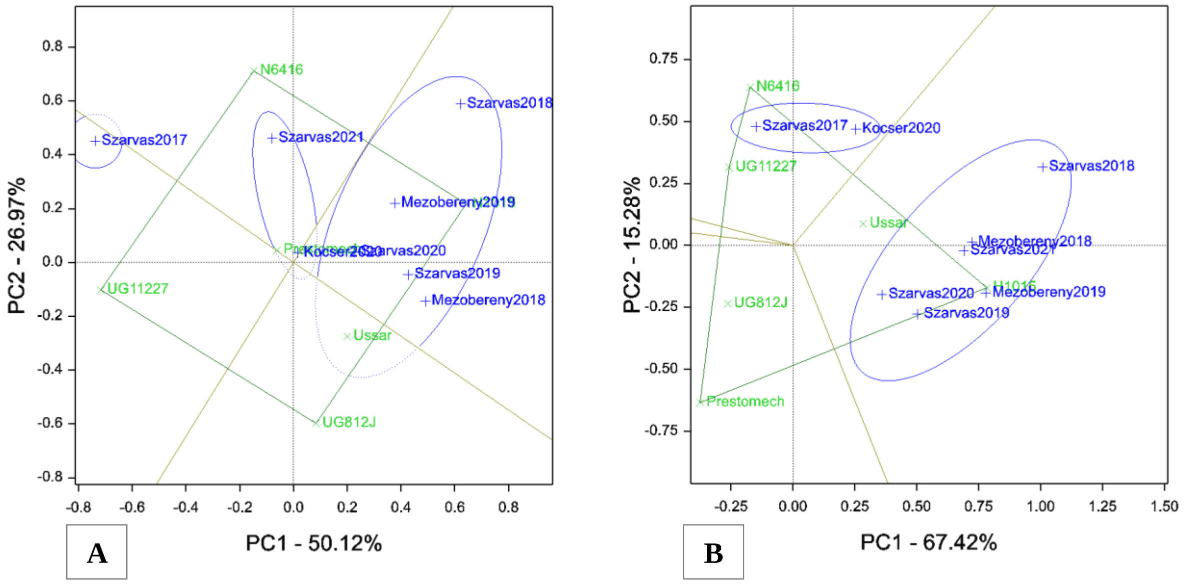


Figure 26. The which-won-where patterns of GGE biplot of Brix (A) and Lycopene (B) of 6 tomato genotypes planted in 8 environments.

Figure 26A showed the biplot evaluating the Brix values reveals the existence of three distinct mega-environments. The first mega-environment comprises Szarvas2017, whereas the second one includes Kocser2020 and Szarvas2021, and the third one all the rest of the environments. Notably, genotypes located in the same sector with a particular environment are the best performers in that environment. genotypes Prestomech is located in the same sector with environment as Kocser2020 and Szarvas2021, therefore, we would expect it to have the highest Brix values in these environments. Same thing goes for Ussar which is located in the same sector with Szarvas2018, Szarvas2019, Szarvas2020, Mezobereny2018 and Mezobereny2019.

Figure 26B showed the biplot evaluating the Lycopene values of identifies two mega-environments in the data, the first is consisting of Szarvas2017 and Kocser2020, the second of all the rest environments. Among these, genotype H1015 is expected to exhibit the best Lycopene in the environments Szarvas2018, Szarvas2019, Szarvas2020, Szarvas2021, Mezobereny2018 and Mezobereny2019 since it is located in the same sector.

Genotypes not associated with any particular mega-environment in a GGE biplot typically exhibit average performance, maintaining moderate Brix or Lycopene levels across diverse conditions without excelling in specific environments. Their stability and adaptability suggest they are less variable in response to environmental changes.

4.3.1.5 Ranking biplot

Figure 27 rank the performances of the 6 genotypes of tomato in all environments. The arrowed line passing through the biplot origin is called average environment coordinate (AEC), The average environment (represented by the small circle at the end of the arrow) has the average coordinates of all test environments and the direction of the arrow indicates increasing average performance in our case the Brix and the Lycopene content (Yan & Kang, 2003). The line which is perpendicular to AEC and passes through the origin represents stability of genotypes; so, the length of the orthogonal projection of each genotype on the AEC axis is a measure of the genotype stability, the longer the orthogonal projection line, the poorer the stability. An ideal genotype should have the highest mean performance and be absolutely stable (Yan & Kang, 2003). Thus Figure 27 could be interpreted by saying that.

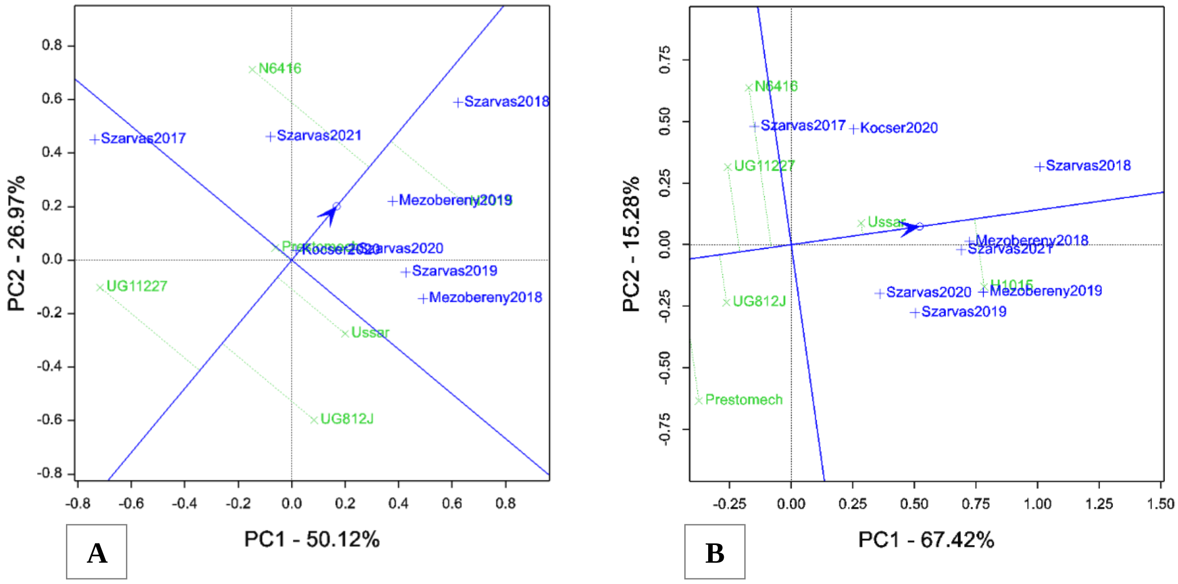


Figure 27. The mean vs. stability view of the GGE biplot of Brix (A) and Lycopene (B) of 6 tomato genotypes planted in 8 environments.

In the biplot analysis assessing Brix and Lycopene content, distinct patterns in genotype performance emerged. H1015 exhibited the highest mean Brix value, followed by N6416 and Prestomech, whereas UG11227 and UG812J recorded the lowest, falling below average. Notably, Prestomech demonstrated remarkable stability in Brix performance. In terms of Lycopene content, H1015 and Ussar showed higher mean values. Contrarily, Prestomech, alongside Ussar, UG812, UG1227, and N6416, exhibited lower mean Lycopene values. Ussar and UG812J were notably stable in their Lycopene performance. It is worth pointing out that according to Yan & Tinker (2006) if the biplot accounts for only a small fraction of the overall variation, it's possible that some genotypes which appear stable might not be genuinely stable. This is because their variability may not be completely captured in the biplot.

4.3.1.6 Comparison Biplot

The concept of an "ideal" genotype is pivotal in plant breeding, embodying the goal of achieving both high mean performance and robust stability across diverse environments. Figure 28 brings this concept to life, providing a comprehensive representation of how closely each of the tested genotypes approaches this ideal.

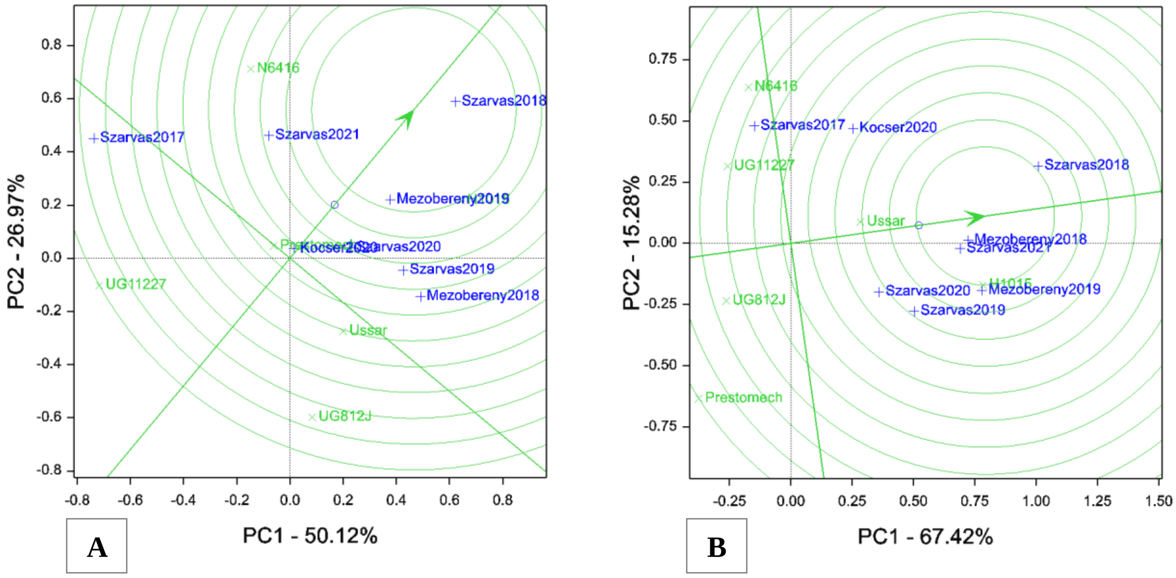


Figure 28. The genotypes comparison with an ideal genotype view of the GGE biplot of Brix (A) and Lycopene (B) of 6 tomato genotypes planted in 8 environments.

Figure 28 defines an “ideal” genotype at the center of concentric circles, characterized by a position on the AEA (Absolutely Stable Axis) in the positive direction and a vector length matching the longest vectors of genotypes on the AEA's positive side, indicative of the highest mean performance. Therefore, the smaller the circle containing a genotype the more attributes it shares with the “ideal genotype”, which makes it more desirable than others. In this context, N6416 surpasses Prestomech in terms of desirable high Brix values, while UG1127 ranks as the least desirable. Conversely, for Lycopene content, H1015 leads in desirability, followed by Ussar, with Prestomech exhibiting the poorest performance across all environments.

Szarvas2018 and Mezobereny2019 emerge as the most favorable environments for achieving high mean Brix values combined with genotype stability. Similarly, Szarvas2021 and Mezobereny2018 are identified as optimal for high mean Lycopene values and genotype stability.

Figure 28 depicts a crucial idea related to "stability". The term "high stability" is favorable only when linked with the mean performance (Yan & Tinker, 2006). This criterion reveals that while Prestomech and Ussar are deemed 'highly stable', they exhibit lower Brix values compared to the less stable genotype, N6416.

4.3.2 Discussion

The principle of plant phenotypic plasticity, as defined by Pigliucci et al. (2006) as "the ability of individual genotypes to produce different phenotypes when exposed to different environmental conditions", is deeply intertwined with the goals of the tomato cultivation industry. This plasticity was clearly illustrated with plant morphology and phenology adjusting to environmental conditions (Dingkuhn et al., 2005). Producers constantly aim to identify genotypes that exhibit high yield and desirable phenotypic traits, as well as production stability and adaptability to different environments. To achieve this goal, it is crucial to investigate the impact of different genotypes and their interactions with the environment. Such investigations are necessary to evaluate and select superior genotypes in multi-environment trials (Yan et al., 2007).

In support of this principle, our GGE biplot analysis of 6 tomato genotypes across 8 distinct environments highlight the importance of understanding genotype by environment interactions (GEI). This study provided crucial insights into GEI for both Brix value and Lycopene. The high proportion of the variation explained by the first two principal components (77.10% for Brix and 82.71% for Lycopene) supports the complexity of these traits and aligns with previous studies investigating the multi-dimensional nature of such traits (Kim et al., 2021). Notably, the genotypes UG812J and Ussar exhibited high Brix values and adaptability across the diversified environments. This finding corroborates with previous studies that reported a strong genetic influence on Brix values, a crucial factor in determining tomato quality (Prinzenberg et al., 2021). Particularly, the stability of Prestomech across different environments, as indicated by its near-zero PC2 value, highlights its potential as a high-performing, adaptable genotype for various environments. The stability of UG11227 for Brix, despite its below-average values, provides a foundation for further selective breeding efforts to improve this trait (Fernandez-Moreno et al., 2017). For Lycopene content, the superior performance of genotypes such as H1015 and Ussar highlights their potential in breeding programs aimed at nutritional quality. In contrast, the remaining genotypes, which exhibited below-average Brix values, demonstrated stability as indicated by higher PC2 values, meriting further investigation. Future research should focus on uncovering the genetic basis of this stability to potentially enhance Lycopene content without sacrificing environmental adaptability.

The environmental clustering observed in this study indicates common factors influencing Lycopene content, an essential nutritional trait (Kuti & Konuru, 2005). This finding aligns with Gauch et al. (2008), who emphasized that environmental effects are distinct from genotypic ones and can significantly alter genetic trait expression. Furthermore, the observed variability in correlations among different environments supports Yan et al. (2000)'s argument that genotype by environment interactions can significantly vary across locations and years, underscoring the complexity of environmental dependencies in trait development.

The results of the 'Which-Won-Where' patterns illustrates the nuanced effects of environment on genotype performance by the demarcation of three distinct mega-environments and their corresponding top-performing genotypes. For instance, genotypes H1015 and Ussar consistently exhibited the highest Brix values across multiple environments, whereas Prestomech was particularly suited to Kocser2020 and Szarvas2021. This observation aligns with the principle that genotypes located in the same sector with an environment on a GGE biplot are best adapted to that environment, as outlined in previous studies (Yan & Rajcan, 2002). Furthermore, these results corroborate the long-established principle, highlighted in Makumbi et al.'s 2011 research, that certain genotypes are particularly well-suited to specific environments. Identifying these mega-environments and the corresponding best-suited genotypes provides invaluable insights for strategic cultivation plans aimed at maximizing Brix values, offering critical insights for targeted breeding programs by indicating genotypes likely to excel in particular conditions.

The analysis of Lycopene levels has identified two distinct mega-environments. The first includes Szarvas2017 and Kocser2020, while the second comprises all other studied environments. In this second mega-environment, the H1015 variety is expected to exhibit superior performance. These findings emphasize the complex interplay between genotype and environment that impacts traits like Lycopene, contributing to the expanding studies highlighting these interactions (Fridman et al., 2004; Zorrilla-Fontanesi et al., 2012).

Genotypes not associated with any particular mega-environment, exhibiting average performance across conditions, are indicative of a broad adaptability and stability. This characteristic is often desirable in breeding programs aimed at developing genotypes with consistent performance across a range of environments, a strategy that has been emphasized in several crop improvement studies (Malosetti et al., 2013).

The ranking biplot offers an invaluable representation of the genotypic performance and stability across multiple environments, helping to explain the inherent complexity of the genotype by environment interactions (Yan & Kang, 2003). Results reveals that H1015, the genotype with the highest mean Brix value, also excels in mean Lycopene content. This dual excellence underscores H1015's potential for maintaining fruit quality consistency in diverse environments.

The considerable variation in Brix and Lycopene values depending on the environment, as illustrated by N6416 and Ussar, aligns with previous studies, underscoring the complex interaction of genetics and environmental factors influencing these traits (Zhang et al., 2005). Furthermore, the fact that genotypes such as N 6438 demonstrate high Brix values but lower stability emphasizes the tricky balancing act between productivity and adaptability that often poses challenges for breeders and farmers (Jat et al., 2017).

The lower performance of genotypes like UG11227 and UG812J in Brix, and Prestomech, alongside others, in Lycopene content, indicates a genetic divergence in the expression of these quality traits. Such variability underscores the complexity of breeding for multiple quality traits, as genotypes often exhibit trade-offs between different attributes (Yan, 2019). This aspect of genetic variability is crucial for breeders, as it guides the selection of parental lines for developing new varieties with a balance of desired traits. Moreover, the significant variations in Lycopene content among genotypes in different environments underscore the impact of genotype by environment interactions on visual fruit quality. For instance, N6416 shows superior performance in the Szarvas2017 environment, aligning with (Beckles, 2012)'s suggestion that selective breeding for specific environments can enhance particular traits.

Notably, the study points out important limitations about interpreting biplot analysis results. If the biplot accounts for a small fraction of the total variation, perceived stability may be deceptive, highlighting the need for comprehensive datasets that cover the complete range of variability (Yan & Tinker, 2006).

The comparison biplot analysis results provide unique insights into the dynamics of genotype performance. N6416 stands out for its high Brix values but shows a notable deficiency in Lycopene content. This contrast illustrates the challenge of optimizing multiple traits within a single genotype, a common issue in plant breeding programs (Cobb et al., 2013). Similarly, the significant difference in Lycopene between Ussar and H1015 underscores the complex trade-off between stability and performance that breeders must navigate. Although Prestomech demonstrates lower Brix values, its high level of stability presents a promising trait for conditions that are more variable, highlighting the importance of maintaining a diversity of genotypes for different environments. This concept aligns the findings of previous research (Swarup et al., 2021), emphasizing the value of genetic diversity in agriculture. N6416 also demonstrates the potential for balancing high Lycopene value and stability, yet its below-average Brix values remind us of the inherent challenges in optimizing multiple traits simultaneously.

According to Yan & Tinker (2006), stability is a desirable trait only when linked with satisfactory performance. Thus, while UG812J demonstrate high stability, they exhibit lower Lycopene, highlighting the importance of not prioritizing stability over productivity.

These findings reiterate the nuances of genotype selection and the necessity of considering a variety of performance metrics, including both trait-specific and overall yield performance, as well as adaptability across varying environmental conditions.

4.4 Integrated Interpretation

The integrated interpretation of our findings reveals a compelling convergence between root development, the predictive accuracy of machine learning models, and the impact of genetic diversity on fruit quality in processing tomato plants. This section is crucial as it synthesizes the core insights from our research, providing a holistic understanding of the interplay between environmental factors, advanced predictive techniques, and genetic resources.

4.4.1 Connection between Root Development and Fruit Quality

Our investigation into root development under varying water supply levels underscores how varying water supply levels influence tomato plant root architecture, with plants under water stress developing deeper, more extensive root systems. This adaptation strategy is crucial for maximizing water uptake in suboptimal conditions, reflecting an innate survival mechanism that directly impacts plant health and productivity. The observed year-to-year variations in root development suggest that environmental factors, alongside irrigation levels, play a significant role in shaping root architecture. This finding lays the groundwork for understanding how root system adaptations contribute to the physiological resilience of the plant, potentially affecting fruit quality traits such as Brix and Lycopene content. Thus, the health and development of the root system underpin the physiological capacity of the tomato plant to produce fruits of high nutritional and commercial quality.

4.4.2 Predictive Modeling of Fruit Quality

The comparative analysis of XGBoost and ANN models in predicting important quality metrics of tomato fruits further enriches our understanding of the interplay between environmental factors, genetic diversity, and fruit quality. The superior performance of XGBoost over ANN in predicting Brix, Lycopene content, and a/b ratio underscores the model's ability to accurately capture the complex relationships between the physicochemical characteristics of tomato fruits and their growing conditions. The SHAP analysis highlights the significant influence of cultivar and climatic conditions on quality predictions, suggesting that both genetic makeup and environmental conditions are pivotal in determining fruit quality.

4.4.3 Genetic Diversity and Environmental Interactions

The evaluation of genetic resources through GGE biplot analysis reveals the profound impact of genotype-by-environment interactions on Brix value and Lycopene content. Genotypes demonstrating high adaptability and performance in diverse environments suggest that genetic factors significantly contribute to the quality traits of tomato fruits. This part of the research underscores the importance of selecting genotypes with broad adaptability for breeding programs aimed at enhancing fruit quality.

4.4.4 Integrating the Findings

Integrating these findings, it becomes evident that root development under different water supply levels, the predictive accuracy of machine learning models, and the genetic diversity of tomato plants are profoundly interconnected with the overall quality of tomato fruits. The adaptation of root systems to water stress conditions not only reflects a fundamental physiological response but also affects the plant's ability to synthesize and accumulate quality determinants such as Brix and Lycopene. The efficacy of machine learning models in predicting these quality traits, based on a range of environmental and genetic factors, offers a powerful tool for precision agriculture, enabling better crop management and breeding decisions. Furthermore, the identification of genotypes with high adaptability and quality performance across various environments highlights the potential for targeted breeding strategies to improve fruit quality.

This study offers a comprehensive perspective on the factors influencing tomato fruit quality, emphasizing the need for an integrated approach that considers physiological traits, genetic diversity, and advanced analytical tools. By understanding the complex interactions between plant physiology, genetics, and the environment, we can develop more sustainable agricultural practices and breeding programs that enhance the quality and resilience of processing tomato plants.

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

5.1.1 Root Development Monitoring under Different Water Supply Levels

Our study highlights the adaptability of tomato plants in response to varying water supply levels. The development of deeper roots under water stress, as observed in our findings, emphasizes plants' inherent strategies to counter water deficits and optimize water uptake. According to our results, the root system expansion to layers with higher soil moisture levels can happen quickly (<one week). The data suggested that root length could triple in 8 days. However, tomato plants that are irrigated regularly with sufficient water quantities develop shorter roots during the intensive root development phase.

Our findings also shed light on the impact of water supply on root system efficacy, with lower irrigation rates and water quantity levels stimulating more intensive root development. The observed variances in root growth over the two consecutive years, influenced by factors such as irrigation water levels and temperature variations, underscore the multifaceted nature of plant responses to environmental conditions.

The relationship between relative chlorophyll content and root development is stronger during the intensive root development period. The consistency in chlorophyll fluorescence across treatments, despite varying water conditions, suggests robust plant mechanisms that maintain photosynthetic efficiency under stress, even if the relative chlorophyll content is affected.

Our research contributes valuable insights into the adaptive strategies of plants under drought stress. This knowledge could inform plant breeding efforts aimed at developing cultivars that are more effectively adapted to water-deficient conditions. It is also pertinent to irrigation professionals seeking to enhance the use of soil layers and improve the effectiveness of root zones.

5.1.2 Comparative Analysis of XGBoost and Neural Network Models for Predicting Tomato Fruit Quality

Our study offers a detailed analysis of Brix, Lycopene, and a/b ratio predictions using XGBoost and ANN models. For Brix prediction, the XGBoost model proved to be highly effective, explaining approximately 98% of the variance in actual Brix values, compared to about 89% by the ANN model. In Lycopene content prediction, the XGBoost model demonstrated high efficacy with an 87% variance explanation, marginally outperforming the ANN model, which accounted for 84%. However, in predicting the a/b ratio, the XGBoost model maintained strong performance with 93% of the variance explained, while the ANN model was notably less effective, indicated by a negative R^2 value of -0.35.

These findings underscore the superior predictive capabilities of the XGBoost model in these scenarios and reveal limitations of the ANN model, especially in predicting the a/b ratio. The SHAP summary plot analysis shows that both models effectively predict Brix values and Lycopene content in tomatoes, but with different focal points. XGBoost emphasized the genetic makeup of cultivars and their interaction with environmental factors, whereas the ANN model captures complex genetic interactions and direct feature relationships. Additionally, our results highlighted the significant influence of temporal factors, particularly 'Year', on the a/b chromaticity ratio, suggesting a complex interplay with climatic conditions and agricultural practices. The limitations of the ANN model in this aspect, as evidenced by its negative SHAP values and R^2 score, underline the necessity of meticulous model selection, optimization, and validation in precision agriculture.

5.1.3 Evaluation of Genetic Resources for Brix and Lycopene in Different Environments

The comprehensive analysis utilizing GGE biplot methodology has revealed distinct patterns of adaptability and performance across different tomato genotypes in varied environmental contexts, demonstrating the complexity and multidimensional nature of these interactions. The identification of mega-environments and their corresponding well-suited genotypes for particular quality traits provides a strategic framework for targeted breeding programs and cultivation practices aimed at enhancing tomato quality. The genotype H1015, for instance, emerged as a notable performer with high mean values in both Brix and Lycopene, suggesting its potential as a cornerstone in breeding programs focused on improving nutritional quality and taste.

Moreover, the variability in performance among genotypes across different environments underscores the essential role of GEI in determining the phenotypic expression of principal quality traits. This variability presents both challenges and opportunities for breeders in selecting and developing genotypes that can deliver consistent performance across diverse environmental conditions. The study also highlights the importance of maintaining genetic diversity in breeding programs, as different genotypes exhibit varied responses to environmental factors, thus enabling the cultivation of tomatoes that meet specific quality standards in different environments.

The limitations inherent in biplot analysis, such as the potential for deceptive stability in cases where a small fraction of the total variation is accounted for, emphasize the need for further research. Comprehensive datasets encompassing a wider range of environmental variability are essential for developing a more nuanced understanding of GEI and its implications for tomato quality trait improvement.

Together, these studies illustrate a comprehensive picture of the factors influencing tomato plant production and fruit quality. The adaptive strategies of plants to water stress, the predictive power of machine learning models, and the critical role of genetic diversity in shaping quality traits are all interconnected facets of a larger agricultural ecosystem. This research underscores the potential for a holistic approach to crop management, one that leverages advanced technologies and genetic insights to foster sustainable and efficient agricultural practices.

5.2 Recommendations

Water Management Practices: It is crucial to develop and implement adaptive water management strategies that are finely tuned to the adaptability of root architecture. Such strategies are designed to maximize water utilization efficiency and sustainability, taking into account the particular requirements and reactions of plants to the availability of water.

Adoption of Machine Learning Models: It is recommended to incorporate sophisticated machine learning techniques, especially XGBoost, to transform the decision-making framework in agriculture. These models offer remarkable capabilities in predicting and optimizing fruit quality, thereby enabling farmers and agronomists to make informed decisions based on precise data analysis. The integration of such machine learning techniques can lead to the development of more resilient agricultural systems that are both more adaptive to fluctuating environments and capable of sustaining elevated productivity levels.

Breeding Programs: There is a pressing need to focus breeding programs on the selection and development of genotypes that exhibit both high-quality traits and adaptability to a range of environmental conditions. This involves leveraging insights from genotype-environment interaction studies to identify plant varieties that can thrive under diverse climatic adversities. By prioritizing genotypes with superior adaptability, breeding programs can contribute to the sustainability and resilience of agricultural practices in the face of global climate change.

Further Research: It is crucial to delve deeper into the integration of environmental, genetic, and technological data to enhance the accuracy of predictive algorithms and breeding methodologies. Further research should explore the impact of emerging climatic challenges on agricultural systems, aiming to develop innovative solutions that can mitigate the adverse effects of climate change on crop production. Furthermore, investigating the synergies between genetic potential and environmental conditions will enable the development of more accurate models for predicting plant performance, thus guiding the creation of more effective and sustainable agricultural practices.

6 NEW SCIENTIFIC RESULTS

- Enhanced Root Development under Water Stress

Water-stressed tomato plants (I50 and control) exhibited significantly more and longer roots compared to fully irrigated plants (I100), particularly in deeper soil layers (30-70 cm). While full irrigation promotes initial root development primarily in the upper soil layers with eventual expansion to deeper layers, limited irrigation (I50 treatment) encourages deeper root growth from the outset as plants seek available water sources in the subsoil.

- Yearly Variations in Root Growth

Tomato plants exhibited significantly greater root growth and longer roots in 2020 compared to 2021, with the I50 treatment showing the most substantial root development (128 roots, 4313 mm in 2020 vs. 45 roots, 2058 mm in 2021); these variations were attributed to temperature and precipitation patterns, which impacted the development and distribution of root systems across different soil depths.

- Superior Performance of XGBoost Model

XGBoost consistently outperformed ANN, achieving high accuracy in predicting Brix ($R^2 = 0.98$, RMSE = 0.07) and lycopene content ($R^2 = 0.87$, RMSE = 0.61), and excelling in colour prediction (a/b ratio) with a R^2 of 0.93 and RMSE of 0.03. ANN lagged behind particularly in colour prediction, showing a negative R^2 value of -0.35.

- Importance of Specific Features in Prediction Models

SHAP value analysis further highlights the critical role of specific features such as 'Cultivar', relative humidity, and soil type, underscoring the complex interplay between genetic makeup, environmental conditions, and tomato quality.

- Differences in Model Interpretability

SHAP analysis reveals distinct differences in feature importance and model interpretability between XGBoost and ANN models, offering nuanced understanding of how genetic factors like 'Cultivar', environmental variables, and even temporal dynamics influence tomato quality traits such as Brix values, Lycopene content, and a/b ratio.

- Genotype and Environment Interaction

The interaction between genotype and environment is significant, contributing to 29.59% of the variance in Brix and 18.74% in lycopene. This indicates that the effect of genotype on these parameters is influenced by environmental factors and vice versa.

- Identification of Superior Genotypes

Genotypes such as UG812J and Ussar exhibit high Brix values and adaptability across diverse environments, while genotypes like H1015 and Ussar are highlighted for their superior Lycopene content, making them ideal candidates for breeding programs focused on nutritional quality enhancement.

7 SUMMARY

Introduction

The tomato, a staple in global diets and a key player in a multi-billion-dollar industry, is celebrated for its nutritional value, economic significance, and health benefits. This cultivation benefits from advanced breeding techniques and sustainable farming practices, enhancing yield and quality. Tomatoes are integral to various cuisines and the processed food sector, with their rich nutrient and bioactive compound content contributing to their dietary value. They possess a unique sweet-sour taste, vibrant colour, and aroma, primarily due to their sugar and acid content.

However, the cultivation of processing tomatoes faces challenges, including environmental stressors like water availability, which impact the plant's growth and fruit quality. Adhering to fruit quality standards under diverse climatic conditions and managing the ripening process, are critical for meeting consumer expectations. Genetic diversity among tomato genotypes presents both challenges and opportunities for breeding resilient varieties.

The approach of the current research aims to enhance the cultivation of processing tomatoes by integrating machine learning methodologies to predict quality attributes and examine the interplay between genetic makeup and environmental conditions.

Materials and Methods

Part 1: We evaluated how tomato plants adapt to varying water levels (100%, 50% of crop evapotranspiration, and non-irrigated control) over two growing seasons in 2020 and 2021. Root images were captured weekly during an 8-week monitoring period in 2020 and 6 weeks in 2021 using a non-destructive CI-600 in-situ root imager at depths between 10 and 70 cm. Also, the SPAD index was measured using a SPAD 502 chlorophyll meter and chlorophyll fluorescence with a PAM 2500 fluorometer device.

Part 2: We utilized a large dataset of physicochemical characteristics and environmental factors for 20 tomato cultivars cultivated across 20 different locations over five growing seasons in Hungary. Two models, eXtreme Gradient Boosting (XGBoost) and Artificial Neural Network (ANN) were trained to predict Brix, Lycopene content, and chromaticity ratio (a/b ratio). Then SHAP values were introduced to measure each feature's contribution to the prediction.

Part 3: We assessed six commercial tomato varieties across three distinct locations in Hungary, with experiments conducted at various intervals within a five-year period from 2017 to 2021. Utilizing advanced instrumental measurements for quality attributes, we conducted GGE biplot analysis to understand genotype-environment interactions. This approach highlighted the adaptability and performance of each variety under diverse conditions, providing insights into optimal cultivation practices for enhanced fruit quality.

Results and Discussion

Part 1: Under water stress, plants developed deeper, more extensive root systems to maximize water uptake, consistent with prior research. Root depth and architecture varied with soil depth and the severity of water stress. Year-to-year variations were also found, likely due to changes in irrigation levels and environmental conditions such as temperature. SPAD values were higher under control conditions, especially in the 2021 growing season, suggesting reduced chlorophyll degradation, while no significant differences were observed in chlorophyll fluorescence (Fv/Fm) between treatments, suggesting stable photosynthetic efficiency under varied water stress conditions.

Part 2: The results revealed that XGBoost outperformed ANN in predicting Brix and Lycopene content. For Brix prediction, XGBoost achieved an R^2 value of 0.98 and an RMSE of 0.07, outperforming ANN which scored an R^2 of 0.89 and an RMSE of 0.17. In prediction of Lycopene content, XGBoost yielded an R^2 of 0.87 and an RMSE of 0.61, compared to the ANN's R^2 of 0.84 and RMSE of 0.86. Additionally, XGBoost excelled in predicting the a/b ratio with an R^2 of 0.93 and RMSE of 0.03, while the ANN model displayed its limitations, with a negative R^2 value of -0.35. SHAP summary plot analysis indicated that both models are effective in predicting Brix and Lycopene content in tomatoes, highlighting different aspects of the data. The SHAP analysis also revealed that cultivar affected the Brix and Lycopene predictions to a great extent, while climatic conditions related to relative humidity or temperature played a great role as well as environmental factors such as soil type or the place of growing.

Part 3: The GGE biplot analysis of tomato genotypes revealed crucial insights into genotype performance and stability across various environments, focusing on Brix and fruit color values. The analysis showed significant variation in genotype performance, with UG812J and Ussar displaying high Brix and good adaptability, while Prestomech showed consistent stability. The Which-Won-Where patterns identified genotypes excelling in specific mega-environments, indicating environmental and genotype interactions. Ranking biplots highlighted genotypes with superior mean performance and stability, with H1015 excelling in Brix value and adaptability. The comparison biplot illustrated the concept of an "ideal" genotype, emphasizing the balance between high performance and stability across environments, with N6416 and H1015 being most desirable for Brix and Lycopene content, respectively.

To enhance agricultural sustainability, it's essential to implement adaptive water management strategies tailored to plant root adaptability, maximizing water efficiency. Incorporating advanced machine learning models, like XGBoost, can revolutionize decision-making by predicting and optimizing fruit quality. Breeding programs should focus on developing genotypes with high-quality traits and environmental adaptability, leveraging genotype-environment interactions. Further research is vital for integrating data across fields, improving predictive algorithms, and addressing climate change impacts, aiming for resilient and sustainable agricultural practices.

8 SCIENTIFIC PUBLICATIONS

Published in impact factor journals

M’hamdi, O., Égei, M., Pék, Z., Ilahy, R., Nemeskéri, E., Helyes, L., & Takács, S. (2023). Root Development Monitoring under Different Water Supply Levels in Processing Tomato Plants. Plants, 12(20), 3517. https://doi.org/10.3390/plants12203517

M’hamdi, O., Takács, S., Palotás, G., Ilahy, R., Helyes, L., & Pék, Z. (2024). A Comparative Analysis of XGBoost and Neural Network Models for Predicting Some Tomato Fruit Quality Traits from Environmental and Meteorological Data. Plants, 13(5), 746. https://doi.org/10.3390/plants13050746

Publications published in non-Impact factor journals

M'HAMDI OUSSAMA, MÁRTON ÉGEI, ZOLTÁN PÉK, SÁNDOR TAKÁCS (2023). Effect of different water supply levels on the root system of industrial tomatoes. KERTGAZDASÁG, 55(1), 50-64. https://budaicampus.uni-mate.hu/documents/54944/7554750/50-64__zoldseg_Mhamdi_et+al_.pdf/129f46a5-a145-abea-552e-7b9550b2ea84?t=1678280426867

Submitted and accepted to Acta Horticulturae

M’hamdi, O., Égei, M., Pék, Z., Ilahy, R., Nemeskéri, E., Helyes, L., & Takács, S. (2024). Adaptive Responses of Tomato Plants to Varying Irrigation Levels: Insights into Root Development Efficiency

M’hamdi, O., Takács, S., Palotás, G., Ilahy, R., Helyes, L., & Pék, Z. (2024). Prediction of Tomato Quality Traits Utilizing Machine Learning Models

Posters

- Adaptive Responses of Tomato Plants to Varying Irrigation Levels: Insights into Root Development Efficiency
- Prediction of Tomato Quality Traits Utilizing Machine Learning Models

Conferences

Participation: 16th World Processing Tomato Congress and 18th ISHS Symposium on Processing Tomatoes (07/06/2026 - 10/06/2026).

Award: Winner of the ISHS Young Minds Award for Best Poster Presentation at the XVII International Symposium on Processing Tomatoes, Budapest, June 2024, for the paper titled “Prediction of Tomato Quality Traits Utilizing Machine Learning Models.”

9 REFERENCES

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International*, *11*(7), 36–42.

Adalid, A. M., Roselló, S., & Nuez, F. (2010). Evaluation and selection of tomato accessions (Solanum section Lycopersicon) for content of lycopene, β-carotene and ascorbic acid. *Journal of Food Composition and Analysis*, *23*(6), 613–618. https://doi.org/10.1016/j.jfca.2010.03.001

Aflitos, S., Schijlen, E., De Jong, H., De Ridder, D., Smit, S., Finkers, R., Wang, J., Zhang, G., Li, N., Mao, L., Bakker, F., Dirks, R., Breit, T., Gravendeel, B., Huits, H., Struss, D., Swanson-Wagner, R., Van Leeuwen, H., Van Ham, R. C. H. J., … Peters, S. (2014). Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. *Plant Journal*, *80*(1), 136–148. https://doi.org/10.1111/tpj.12616

Agathokleous, E., Belz, R. G., Kitao, M., Koike, T., & Calabrese, E. J. (2019). Does the root to shoot ratio show a hormetic response to stress? An ecological and environmental perspective. In *Journal of Forestry Research* (Vol. 30, Issue 5, pp. 1569–1580). Northeast Forestry University. https://doi.org/10.1007/s11676-018-0863-7

Agbemafle, R., Danso Owusu-Sekyere, J., Bart-Plange, A., Otchere, J. K., Owusu-Sekyere, J., Bart-Plange, A., & Otchere, J. (2014). Effect of Deficit Irrigation and Storage on Physicochemical Quality of Tomato (Lycopersicon esculentum Mill. Var. Pechtomech). *Food Science and Quality Management*, *34*. www.iiste.org

Agius, C., von Tucher, S., Poppenberger, B., & Rozhon, W. (2018). Quantification of sugars and organic acids in tomato fruits. *MethodsX*, *5*, 537–550. https://doi.org/10.1016/j.mex.2018.05.014

Ahanger, M. A., Akram, N. A., Ashraf, M., Alyemeni, M. N., Wijaya, L., & Ahmad, P. (2017). Plant responses to environmental stresses-From gene to biotechnology. *AoB PLANTS*, *9*(4). https://doi.org/10.1093/aobpla/plx025

Ahlquist, K. D., Sugden, L. A., & Ramachandran, S. (2023). Enabling interpretable machine learning for biological data with reliability scores. *PLoS Computational Biology*, *19*(5). https://doi.org/10.1371/journal.pcbi.1011175

Akimov, S. S., Lebedev, S. V., Grechkina, V. V., Miroshnikova, M. S., & Topuria, G. M. (2021). The effectiveness of using mathematical modeling in assessing the quality of food products. *IOP Conference Series: Earth and Environmental Science*, *624*(1). https://doi.org/10.1088/1755-1315/624/1/012158

Aldrich, H. T., Salandanan, K., Kendall, P., Bunning, M., Stonaker, F., Külen, O., & Stushnoff, C. (2010). Cultivar choice provides options for local production of organic and conventionally produced tomatoes with higher quality and antioxidant content. *Journal of the Science of Food and Agriculture*, *90*(15), 2548–2555. https://doi.org/10.1002/jsfa.4116

Ali, I., Cawkwell, F., Dwyer, E., & Green, S. (2017). Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data-A Machine Learning Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(7), 3254–3264. https://doi.org/10.1109/JSTARS.2016.2561618

Ali, M. Y., Sina, A. A. I., Khandker, S. S., Neesa, L., Tanvir, E. M., Kabir, A., Khalil, M. I., & Gan, S. H. (2021). Nutritional composition and bioactive compounds in tomatoes and their impact on human health and disease: A review. *Foods*, *10*(1). https://doi.org/10.3390/foods10010045

Allard, R. W., & Bradshaw, A. D. (1964). *Implications of Genotype-Environmental Interactions in Applied Plant Breeding*.

Almeida, J. S. (2002). Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology*, *13*(1), 72–76. http://mb01.cbs.dtu.dk/index.html

Amankwah, A. (2023). Climate variability, agricultural technologies adoption, and productivity in rural Nigeria: a plot-level analysis. *Agriculture and Food Security*, *12*(1). https://doi.org/10.1186/s40066-023-00411-x

Amatya, S., Karkee, M., Gongal, A., Zhang, Q., & Whiting, M. D. (2016). Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. *Biosystems Engineering*, *146*, 3–15. https://doi.org/10.1016/j.biosystemseng.2015.10.003

Amira, J. O., Ojo, D. K., Ariyo, O. J., Oduwaye, O. A., & Ayo-Vaughan, M. A. (2013). RELATIVE DISCRIMINATING POWERS OF GGE AND AMMI MODELS IN THE SELECTION OF TROPICAL SOYBEAN GENOTYPES. *African Crop Science Journal*, *21*(1), 67–73.

Amiri, M., Jafari, A. H., Makkiabadi, B., & Nazari, S. (2022). Recognizing intertwined patterns using a network of spiking pattern recognition platforms. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-23320-8

Arah, I. K., Ahorbo, G. K., Anku, E. K., Kumah, E. K., & Amaglo, H. (2016). Postharvest Handling Practices and Treatment Methods for Tomato Handlers in Developing Countries: A Mini Review. *Advances in Agriculture*, *2016*. https://doi.org/10.1155/2016/6436945

Arbex de Castro Vilas Boas, A., Page, D., Giovinazzo, R., Bertin, N., & Fanciullino, A. L. (2017). Combined effects of irrigation regime, genotype, and harvest stage determine tomato fruit quality and aptitude for processing into puree. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01725

Arias, R., Lee, T. C., Logendra, L., & Janes, H. (2000). Correlation of lycopene measured by HPLC with the L*, a*, b* color readings of a hydroponic tomato and the relationship of maturity with color and lycopene content. *Journal of Agricultural and Food Chemistry*, *48*(5), 1697–1702. https://doi.org/10.1021/jf990974e

Atnaf, M. (2013). GGE biplots to analyze soybean multi-environment yield trial data in north Western Ethiopia. *Journal of Plant Breeding and Crop Science*, *5*(12), 245–254. https://doi.org/10.5897/jpbcs13.0403

Attri, I., Awasthi, L. K., & Sharma, T. P. (2023). Machine learning in agriculture: a review of crop management applications. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-16105-2

Ayankojo, I. T., & Morgan, K. T. (2020). Increasing air temperatures and its effects on growth and productivity of tomato in south florida. *Plants*, *9*(9), 1–16. https://doi.org/10.3390/plants9091245

Ayaz Mirani, A., Muhammad Suleman Memon, E., Qabulio, M., Suleman Memon, M., Chohan, R., & Ali Wagan, A. (2021). Machine Learning In Agriculture: A Review. *LUME*, *10*(5). www.ijstr.org

Aziz, R., Verma, C. K., & Srivastava, N. (2018). Artificial Neural Network Classification of High Dimensional Data with Novel Optimization Approach of Dimension Reduction. *Annals of Data Science*, *5*(4), 615–635. https://doi.org/10.1007/s40745-018-0155-2

Bai, Y., & Lindhout, P. (2007). Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Annals of Botany*, *100*(5), 1085–1094. https://doi.org/10.1093/aob/mcm150

Baldwin, E. A., Goodner, K., & Plotto, A. (2008). Interaction of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors. *Journal of Food Science*, *73*(6). https://doi.org/10.1111/j.1750-3841.2008.00825.x

Baltazar, A., Aranda, J. I., & González-Aguilar, G. (2008). Bayesian classification of ripening stages of tomato fruit using acoustic impact and colorimeter sensor data. *Computers and Electronics in Agriculture*, *60*(2), 113–121. https://doi.org/10.1016/j.compag.2007.07.005

Barickman, T. C., Kopsell, D. A., & Sams, C. E. (2016). Abscisic Acid Impacts Tomato Carotenoids, Soluble Sugars, and Organic Acids. *HORTSCIENCE*, *51*(4), 370–376. https://doi.org/org/10.21273/HORTSCI.51.4.370

Basu, A., & Imrhan, V. (2007). Tomatoes versus lycopene in oxidative stress and carcinogenesis: Conclusions from clinical trials. In *European Journal of Clinical Nutrition* (Vol. 61, Issue 3, pp. 295–303). https://doi.org/10.1038/sj.ejcn.1602510

Bates, S., Hastie, T., & Tibshirani, R. (2021). *Cross-validation: what does it estimate and how well does it do it?* http://arxiv.org/abs/2104.00673

Bdr, M. F., Anshori, M. F., Emanuella, G., Pratiwi, N., Ermiyanti, I., Yovita, V., Musdalifa, M., & Nasaruddin, N. (2020). High Lycopene Tomato Breeding Through Diallel Crossing. *Agrotech Journal*, *5*(2), 63–72. https://doi.org/10.31327/atj.v5i2.1347

Beckles, D. M. (2012). Factors affecting the postharvest soluble solids and sugar content of tomato (Solanum lycopersicum L.) fruit. In *Postharvest Biology and Technology* (Vol. 63, Issue 1, pp. 129–140). https://doi.org/10.1016/j.postharvbio.2011.05.016

Beckles, D. M., Hong, N., Stamova, L., & Luengwilai, K. (2012). Biochemical factors contributing to tomato fruit sugar content: A review. In *Fruits* (Vol. 67, Issue 1, pp. 49–64). EDP Sciences. https://doi.org/10.1051/fruits/2011066

Behmann, J., Mahlein, A. K., Rumpf, T., Römer, C., & Plümer, L. (2015). A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. In *Precision Agriculture* (Vol. 16, Issue 3, pp. 239–260). Kluwer Academic Publishers. https://doi.org/10.1007/s11119-014-9372-7

Bellini, C., Pacurar, D. I., & Perrone, I. (2014). Adventitious roots and lateral roots: Similarities and differences. In *Annual Review of Plant Biology* (Vol. 65, pp. 639–666). Annual Reviews Inc. https://doi.org/10.1146/annurev-arplant-050213-035645

Bergougnoux, V. (2014). The history of tomato: From domestication to biopharming. In *Biotechnology Advances* (Vol. 32, Issue 1, pp. 170–189). https://doi.org/10.1016/j.biotechadv.2013.11.003

Bergstra, J., Ca, J. B., & Ca, Y. B. (2012). Random Search for Hyper-Parameter Optimization Yoshua Bengio. In *Journal of Machine Learning Research* (Vol. 13). http://scikit-learn.sourceforge.net.

Bhandari, H. R., Srivastava, K., Tripathi, M. K., Chaudhary, B., & Biswas, S. (2022). Genotype-Environment Interaction for Quality Traits in Tomato Hybrids. *Agricultural Research*, *11*(3), 382–389. https://doi.org/10.1007/s40003-021-00579-3

Bihon, W., Ognakossan, K. E., Tignegre, J. B., Hanson, P., Ndiaye, K., & Srinivasan, R. (2022). Evaluation of Different Tomato (Solanum lycopersicum L.) Entries and Varieties for Performance and Adaptation in Mali, West Africa. *Horticulturae*, *8*(7). https://doi.org/10.3390/horticulturae8070579

Bineau, E., Rambla, J. L., Duboscq, R., Corre, M. N., Bitton, F., Lugan, R., Granell, A., Plissonneau, C., & Causse, M. (2022). Inheritance of Secondary Metabolites and Gene Expression Related to Tomato Fruit Quality. *International Journal of Molecular Sciences*, *23*(11). https://doi.org/10.3390/ijms23116163

Bishop, C. M. (1995). Neural Networks for Pattern Recognition. *Oxford University Press*.

Bláha, L. (2019). Importance Of Root-Shoot Ratio For Crops Production. *Agronomy & Agricultural Science*, *2*(2), 1–7. https://doi.org/10.24966/AAS-8292/100012

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Bucksch, A., Burridge, J., York, L. M., Das, A., Nord, E., Weitz, J. S., & Lynch, J. P. (2014). Image-based high-throughput field phenotyping of crop roots. *Plant Physiology*, *166*(2), 470–486. https://doi.org/10.1104/pp.114.243519

Burton-Freeman, B. M., & Sesso, H. D. (2014). Whole food versus supplement: Comparing the clinical evidence of tomato intake and lycopene supplementation on cardiovascular risk factors. In *Advances in Nutrition* (Vol. 5, Issue 5, pp. 457–485). American Society for Nutrition. https://doi.org/10.3945/an.114.005231

Burton-Freeman, B., & Reimers, K. (2011). Tomato Consumption and Health: Emerging Benefits. In *American Journal of Lifestyle Medicine* (Vol. 5, Issue 2, pp. 182–191). https://doi.org/10.1177/1559827610387488

Cassab, G. I., Eapen, D., & Campos, M. E. (2013). ROOT HYDROTROPISM: AN UPDATE. *American Journal of Botany*, *100(1)*, 14–24.

Chakraborty, P. K., Banerjee, S., Nath, R., & Samanta, S. (2022). Assessing congenial soil temperature and its impact on root growth, grain yield of summer rice under varying water stress condition in Lower Gangetic Plain of India. *Journal of the Saudi Society of Agricultural Sciences*, *21*(2), 98–107. https://doi.org/10.1016/j.jssas.2021.07.001

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., & Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, *4*(4), 287–291. https://doi.org/10.1038/nclimate2153

Chandola, V., Varun, A., & Kumar, V. (2009). Anomaly Detection : A Survey. *ACM Computing Surveys (CSUR)*, *41*(3), 1–58. https://doi.org/10.1145/1541880.1541882

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, X., Yao, Q., Gao, X., Jiang, C., Harberd, N. P., & Fu, X. (2016). Shoot-to-Root Mobile Transcription Factor HY5 Coordinates Plant Carbon and Nitrogen Acquisition. *Current Biology*, *26*(5), 640–646. https://doi.org/10.1016/j.cub.2015.12.066

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. In *Computers and Electronics in Agriculture* (Vol. 151, pp. 61–69). Elsevier B.V. https://doi.org/10.1016/j.compag.2018.05.012

Chowdhury, M., Kiraga, S., Islam, M. N., Ali, M., Reza, M. N., Lee, W. H., & Chung, S. O. (2021). Effects of temperature, relative humidity, and carbon dioxide concentration on growth and glucosinolate content of kale grown in a plant factory. *Foods*, *10*(7). https://doi.org/10.3390/foods10071524

Clark, R. T., Famoso, A. N., Zhao, K., Shaff, J. E., Craft, E. J., Bustamante, C. D., Mccouch, S. R., Aneshansley, D. J., & Kochian, L. V. (2013). High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. *Plant, Cell and Environment*, *36*(2), 454–466. https://doi.org/10.1111/j.1365-3040.2012.02587.x

Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics*, *126*(4), 867–887. https://doi.org/10.1007/s00122-013-2066-0

Collins, E. J., Bowyer, C., Tsouza, A., & Chopra, M. (2022). Tomatoes: An Extensive Review of the Associated Health Impacts of Tomatoes and Factors That Can Affect Their Cultivation. In *Biology* (Vol. 11, Issue 2). MDPI. https://doi.org/10.3390/biology11020239

Condran, S., Bewong, M., Islam, M. Z., Maphosa, L., & Zheng, L. (2022). Machine Learning in Precision Agriculture: A Survey on Trends, Applications and Evaluations over Two Decades. *IEEE Access*, *10*, 73786–73803. https://doi.org/10.1109/ACCESS.2022.3188649

Corbin, K. R., Bolt, B., & Rodríguez López, C. M. (2020). Breeding for Beneficial Microbial Communities Using Epigenomics. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00937

Costa, J. M., Grant, O. M., & Chaves, M. M. (2013). Thermography to explore plant-environment interactions. In *Journal of Experimental Botany* (Vol. 64, Issue 13, pp. 3937–3949). https://doi.org/10.1093/jxb/ert029

Crossa, J. (1990). STATISTICAL ANALYSES OF MU LTI LOCATION TRIALS. In *ADVANCES IN AGRONOMY* (Vol. 44).

Csambalik, L., Gál, I., Sipos, L., Gere, A., Koren, D., Bíró, B., & Divéky-Ertsey, A. (2019). Evaluation of processing type tomato plant genetic resources (Solanum lycopersicum L.) for their nutritional properties in different environments. *Plant Genetic Resources: Characterisation and Utilisation*, *17*(6), 488–498. https://doi.org/10.1017/S1479262119000297

Daood, H. G., Bencze, G., Palotás, G., Pék, Z., Sidikov, A., & Helyes, L. (2014). HPLC analysis of carotenoids from tomatoes using cross-linked C18 column and MS detection. *Journal of Chromatographic Science*, *52*(9), 985–991. https://doi.org/10.1093/chromsci/bmt139

Darwin, B., Dharmaraj, P., Prince, S., Popescu, D. E., & Hemanth, D. J. (2021). Recognition of bloom/yield in crop images using deep learning models for smart agriculture: A review. In *Agronomy* (Vol. 11, Issue 4). MDPI AG. https://doi.org/10.3390/agronomy11040646

De Bauw, P., Ramarolahy, J. A., Senthilkumar, K., Rakotoson, T., Merckx, R., Smolders, E., Houtvinck, R. Van, & Vandamme, E. (2020). Phenotyping Root Architecture of Soil-Grown Rice: A Robust Protocol Combining Manual Practices with Image-based Analyses. *BioRxiv*, 2020–05. https://doi.org/10.1101/2020.05.13.088369

Deák, K., Szigedi, T., Pék, Z., Baranowski, P., & Helyes, L. (2015). Carotenoid determination in tomato juice using near infrared spectroscopy. *International Agrophysics*, *29*(3), 275–282. https://doi.org/10.1515/intag-2015-0032

D'Esposito, D., Ferriello, F., Molin, A. D., Diretto, G., Sacco, A., Minio, A., Barone, A., Di Monaco, R., Cavella, S., Tardella, L., Giuliano, G., Delledonne, M., Frusciante, L., & Ercolano, M. R. (2017). Unraveling the complexity of transcriptomic, metabolomic and quality environmental response of tomato fruit. *BMC Plant Biology*, *17*(1). https://doi.org/10.1186/s12870-017-1008-4

Dietrich, R. C., Bengough, A. G., Jones, H. G., & White, P. J. (2012). A new physical interpretation of plant root capacitance. *Journal of Experimental Botany*, *63*(17), 6149–6159. https://doi.org/10.1093/jxb/ers264

Dietrich, R. C., Bengough, A. G., Jones, H. G., & White, P. J. (2013). Can root electrical capacitance be used to predict root mass in soil? *Annals of Botany*, *112*(2), 457–464. https://doi.org/10.1093/aob/mct044

Dingkuhn, M., Luquet, D., Quilot, B., & De Reffye, P. (2005). Environmental and genetic control of morphogenesis in crops: Towards models simulating phenotypic plasticity. *Australian Journal of Agricultural Research*, *56*(11), 1289–1302. https://doi.org/10.1071/AR05063

Divéky-Ertsey, A., Ladányi, M., Biró, B., Máté, M., Drexler, D., Tóth, F., Boziné Pullai, K., Gere, A., Pusztai, P., & Csambalik, L. (2022). Tomato Landraces May Benefit from Protected Production—Evaluation on Phytochemicals. *Horticulturae*, *8*(10). https://doi.org/10.3390/horticulturae8100937

Dixit, S., & Pandey, V. R. (2017). Genetic variability, heritability and genetic advance in tomato [Solanum lycopersicon (Mill.) Wettsd]. *Asian Journal of Horticulture*, *12*(1), 75–78. https://doi.org/10.15740/has/tajh/12.1/75-78

Dong, F., Wang, C., Sun, X., Bao, Z., Dong, C., Sun, C., Ren, Y., & Liu, S. (2019). Sugar metabolic changes in protein expression associated with different light quality combinations in tomato fruit. *Plant Growth Regulation*. https://doi.org/10.1007/s10725-019-00506-1

Dorais, M., Ehret, D. L., & Papadopoulos, A. P. (2008). Tomato (Solanum lycopersicum) health components: From the seed to the consumer. In *Phytochemistry Reviews* (Vol. 7, Issue 2, pp. 231–250). https://doi.org/10.1007/s11101-007-9085-x

Downie, H., Holden, N., Otten, W., Spiers, A. J., Valentine, T. A., & Dupuy, L. X. (2012). Transparent Soil for Imaging the Rhizosphere. *PLoS ONE*, *7*(9). https://doi.org/10.1371/journal.pone.0044276

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Vol. 326). John Wiley & Sons. https://doi.org/10.1002/9781118625590.fmatter

Duraivel, S., Rahimpour, S., Chiang, C. H., Trumpis, M., Wang, C., Barth, K., Harward, S. C., Lad, S. P., Friedman, A. H., Southwell, D. G., Sinha, S. R., Viventi, J., & Cogan, G. B. (2023). High-resolution neural recordings improve the accuracy of speech decoding. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-42555-1

Ecarnot, M., Baogonekczyk, P., Tessarotto, L., & Chervin, C. (2013). Rapid phenotyping of the tomato fruit model, Micro-Tom, withaportable VIS-NIR spectrometer. *Plant Physiology and Biochemistry*, *70*, 159–163. https://doi.org/10.1016/j.plaphy.2013.05.019

Égei, M., Takács, S., Palotás, G., Palotás, G., Szuvandzsiev, P., Daood, H. G., Helyes, L., & Pék, Z. (2022). Prediction of Soluble Solids and Lycopene Content of Processing Tomato Cultivars by Vis-NIR Spectroscopy. *Frontiers in Nutrition*, *9*. https://doi.org/10.3389/fnut.2022.845317

FAO. (2022). *Food and Agriculture Organization of the United Nations*. FAOSTAT. https://www.fao.org/faostat/en/#data/QCL

Fara, S. J., Teixeira Delazari, F., Silva Gomes, R., Araújo, W. L., & da Silva, D. J. H. (2019). Stomata opening and productiveness response of fresh market tomato under different irrigation intervals. *Scientia Horticulturae*, *255*, 86–95. https://doi.org/10.1016/j.scienta.2019.05.025

Farifteh, J., Van der Meer, F., Atzberger, C., & Carranza, E. J. M. (2007). Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sensing of Environment*, *110*(1), 59–78. https://doi.org/10.1016/j.rse.2007.02.005

Fayeun, L. S., Alake, G. C., & Akinlolu, A. O. (2018). GGE biplot analysis of fluted pumpkin (Telfairia occidentalis) landraces evaluated for marketable leaf yield in Southwest Nigeria. *Journal of the Saudi Society of Agricultural Sciences*, *17*(4), 416–423. https://doi.org/10.1016/j.jssas.2016.10.001

Fernandez-Moreno, J. P., Levy-Samoha, D., Malitsky, S., Monforte, A. J., Orzaez, D., Aharoni, A., & Granell, A. (2017). Uncovering tomato quantitative trait loci and candidate genes for fruit cuticular lipid composition using the Solanum pennellii introgression line population. *Journal of Experimental Botany*, *68*(11), 2703–2716. https://doi.org/10.1093/jxb/erx134

Fischer, I., Camus-Kulandaivelu, L., Allal, F., & Stephan, W. (2011). Adaptation to drought in two wild tomato species: The evolution of the Asr gene family. *New Phytologist*, *190*(4), 1032–1044. https://doi.org/10.1111/j.1469-8137.2011.03648.x

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A. R., & Zamir, D. (2004). Zooming In on a Quantitative Trait for Tomato Yield Using Interspecific Introgressions. *Science*, *305*(5691), 1786–1786. www.sciencemag.org

Friedman, J. H. (2001). 999 REITZ LECTURE GREEDY FUNCTION APPROXIMATION: A GRADIENT BOOSTING MACHINE 1. In *The Annals of Statistics* (Vol. 29, Issue 5).

Frutos, E., Galindo, M. P., & Leiva, V. (2014). An interactive biplot implementation in R for modeling genotype-by-environment interaction. In *Stochastic Environmental Research and Risk Assessment* (Vol. 28, Issue 7, pp. 1629–1641). Springer Science and Business Media, LLC. https://doi.org/10.1007/s00477-013-0821-z

Fry, E. L., Evans, A. L., Sturrock, C. J., Bullock, J. M., & Bardgett, R. D. (2018). Root architecture governs plasticity in response to drought. *Plant and Soil*, *433*(1–2), 189–200. https://doi.org/10.1007/s11104-018-3824-1

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., Burzynski-Chang, E. A., Fish, T. L., Stromberg, K. A., Sacks, G. L., Thannhauser, T. W., Foolad, M. R., Diez, M. J., Blanca, J., Canizares, J., Xu, Y., van der Knaap, E., Huang, S., Klee, H. J., … Fei, Z. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nature Genetics*, *51*(6), 1044–1051. https://doi.org/10.1038/s41588-019-0410-2

Gascuel, Q., Diretto, G., Monforte, A. J., Fortes, A. M., & Granell, A. (2017). Use of natural diversity and biotechnology to increase the quality and nutritional content of tomato and grape. In *Frontiers in Plant Science* (Vol. 8). Frontiers Media S.A. https://doi.org/10.3389/fpls.2017.00652

Gauch, H. G., Piepho, H. P., & Annicchiarico, P. (2008). Statistical analysis of yield trials by AMMI and GGE: Further considerations. In *Crop Science* (Vol. 48, Issue 3, pp. 866–889). https://doi.org/10.2135/cropsci2007.09.0513

Ge, J., Zhao, L., Yu, Z., Liu, H., Zhang, L., Gong, X., & Sun, H. (2022). Prediction of Greenhouse Tomato Crop Evapotranspiration Using XGBoost Machine Learning Model. *Plants*, *11*(15). https://doi.org/10.3390/plants11151923

Gharezi, M., Joshi, N., & Sadeghian, E. (2012). Effect of Post Harvest Treatment on Stored Cherry Tomatoes. *Journal of Nutrition & Food Sciences*, *02*(08). https://doi.org/10.4172/2155-9600.1000157

Giovannucci, E. (1999). Tomatoes, Tomato-Based Products, Lycopene, and Cancer: Review of the Epidemiologic Literature. *JNCI: Journal of the National Cancer Institute*, *91*(4), 317–331. https://doi.org/https://doi.org/10.1093/jnci/91.4.317

Giuliani, M. M., Gatta, G., Cappelli, G., Gagliardi, A., Donatelli, M., Fanchini, D., De Nart, D., Mongiano, G., & Bregaglio, S. (2019). Identifying the most promising agronomic adaptation strategies for the tomato growing systems in Southern Italy via simulation modeling. *European Journal of Agronomy*, *111*. https://doi.org/10.1016/j.eja.2019.125937

Givnish, T. J. (2002). Ecological constraints on the evolution of plasticity in plants. *Evolutionary Ecology*, *16*, 213–242.

Gliessman, S. (2022). Preserving the wildness of crop wild relatives. In *Agroecology and Sustainable Food Systems* (Vol. 46, Issue 1, pp. 1–2). Taylor and Francis Ltd. https://doi.org/10.1080/21683565.2021.1998981

Gliński, Jan., & Lipiec, J. (1990). *Soil physical conditions and plant roots*. CRC Press.

Goisser, S., Krause, J., Fernandes, M., Mempel, H., Goisser, S., Krause, J., Fernandes, M., & Mempel, H. (2019). Determination of tomato quality attributes using portable NIR-sensors. *KIT Scientific Publishing.* https://doi.org/10.5445/IR/1000092314

Goisser, S., Wittmann, S., Fernandes, M., Mempel, H., & Ulrichs, C. (2020). Comparison of colorimeter and different portable food-scanners for non-destructive prediction of lycopene content in tomato fruit. *Postharvest Biology and Technology, 167*. https://doi.org/10.1016/j.postharvbio.2020.111232

Gomes, V. M., Fernandes, A. M., Faia, A., & Melo-Pinto, P. (2014). Determination of sugar content in whole Port Wine grape berries combining hyperspectral imaging with neural networks methodologies. *IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*, 188–193.

Gomes, V. M., Fernandes, A. M., Faia, A., & Melo-Pinto, P. (2017). Comparison of different approaches for the prediction of sugar content in new vintages of whole Port wine grape berries using hyperspectral imaging. *Computers and Electronics in Agriculture, 140,* 244–254. https://doi.org/10.1016/j.compag.2017.06.009

Gonin, M., Bergougnoux, V., Nguyen, T. D., Gantet, P., & Champion, A. (2019). What makes adventitious roots? In *Plants* (Vol. 8, Issue 7). MDPI AG. https://doi.org/10.3390/plants8070240

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT press.

Grift, T. E., Novais, J., & Bohn, M. (2011). High-throughput phenotyping technology for maize roots. *Biosystems Engineering*, *110*(1), 40–48. https://doi.org/10.1016/j.biosystemseng.2011.06.004

Guerra, A. S., Hoyos, C. G., Molina-Ramírez, C., Velásquez-Cock, J., Vélez, L., Gañán, P., Eceiza, A., Goff, H. D., & Zuluaga, R. (2021). Extraction and preservation of lycopene: A review of the advancements offered by the value chain of nanotechnology. *Trends in Food Science and Technology*, *116*, 1120–1140. https://doi.org/10.1016/j.tifs.2021.09.009

Guerrero-Campo, J., & Fitter, A. H. (2001). *Relationships between root characteristics and seed size in two contrasting floras*.

Guida, G., Sellami, M. H., Mistretta, C., Oliva, M., Buonomo, R., De Mascellis, R., Patanè, C., Rouphael, Y., Albrizio, R., & Giorio, P. (2017). Agronomical, physiological and fruit quality responses of two Italian long-storage tomato landraces under rain-fed and full irrigation conditions. *Agricultural Water Management*, *180*, 126–135. https://doi.org/10.1016/j.agwat.2016.11.004

Guimarães, P. H. R., de Lima, I. P., de Castro, A. P., Lanna, A. C., Guimarães Santos Melo, P., & de Raïssac, M. (2020). Phenotyping Root Systems in a Set of Japonica Rice Accessions: Can Structural Traits Predict the Response to Drought? *Rice*, *13*(1). https://doi.org/10.1186/s12284-020-00404-5

Happ, M. M., Graef, G. L., Wang, H., Howard, R., Posadas, L., & Hyten, D. L. (2021). Comparing a Mixed Model Approach to Traditional Stability Estimators for Mapping Genotype by Environment Interactions and Yield Stability in Soybean [Glycine max (L.) Merr.]. *Frontiers in Plant Science, 12*. https://doi.org/10.3389/fpls.2021.630175

Haq Chowdhury, R., Sultana Eti, F., Atiqur Rahman Bhuiyan, M., Das Gupta, S., & Hassan Rubel, M. (2023). *Drought-Responsive Genes in Tomato: Meta-Analysis of Gene Expression using Machine Learning*. https://doi.org/10.20944/preprints202305.1530.v1

He, M., & Dijkstra, F. A. (2014). Drought effect on plant nitrogen and phosphorus: A meta-analysis. *New Phytologist*, *204*(4), 924–931. https://doi.org/10.1111/nph.12952

Helyes, L., Lugasi, A., & Pék, Z. (2007). Effect of natural light on surface temperature and lycopene content of vine ripened tomato fruit. *Canadian Journal of Plant Science*, *87*(4), 927–929. https://doi.org/https://doi.org/10.4141/cjps07022

Helyes, L., Lugasi, A., & Pék, Z. (2012). Effect of irrigation on processing tomato yield and antioxidant components. *Turkish Journal of Agriculture and Forestry*, *36*(6), 702–709. https://doi.org/10.3906/tar-1107-9

Hernandez-Espinoza, L. H., & Barrios-Masias, F. H. (2020). Physiological and anatomical changes in tomato roots in response to low water stress. *Scientia Horticulturae*, *265*. https://doi.org/10.1016/j.scienta.2020.109208

Herrera, J. M., Stamp, P., & Liedgens, M. (2007). *DYNAMICS OF ROOT DEVELOPMENT OF SPRING WHEAT GENOTYPES VARYING IN NITROGEN USE EFFICIENCY*.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Ilahy, R., Siddiqui, M. W., Tlili, I., Montefusco, A., Piro, G., Hdider, C., & Lenucci, M. S. (2018). When Color Really Matters: Horticultural Performance and Functional Quality of High-Lycopene Tomatoes. *Critical Reviews in Plant Sciences*, *37*(1), 15–53. https://doi.org/10.1080/07352689.2018.1465631

International Commission for Uniform Methods of Sugar Analysis. (1974). *ICUMSA Proceedings 1974: 16th Session*. Verlag Dr. Albert Bartens.

Isaac Abiodun, O., Jantan, A., Esther Omolara, A., Victoria Dada, K., AbdElatif Mohamed, N., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*, 938. https://doi.org/10.1016/j.heliyon.2018

Jaiswal, A. K., Mengiste, T. D., Myers, J. R., Egel, D. S., & Hoagland, L. A. (2020). Tomato Domestication Attenuated Responsiveness to a Beneficial Soil Microbe for Plant Growth Promotion and Induction of Systemic Resistance to Foliar Pathogens. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.604566

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*.

Jat, M. L., Jat, R. K., Singh, P., Jat, S. L., Sidhu, H. S., Jat, H. S., Bijarniya, D., Parihar, C. M., & Gupta, R. (2017). Predicting Yield and Stability Analysis of Wheat under Different Crop Management Systems across Agro-Ecosystems in India. *American Journal of Plant Sciences*, *08*(08), 1977–2012. https://doi.org/10.4236/ajps.2017.88133

Jaywant, S. A., Singh, H., & Arif, K. M. (2022). Sensors and Instruments for Brix Measurement: A Review. *Sensors*, *22*(6). https://doi.org/10.3390/s22062290

Jha, S. K., Gao, Y., Liu, H., Huang, Z., Wang, G., Liang, Y., & Duan, A. (2017). Root development and water uptake in winter wheat under different irrigation methods and scheduling for North China. *Agricultural Water Management*, *182*, 139–150. https://doi.org/10.1016/j.agwat.2016.12.015

Jin, H. (2022). *Hyperparameter Importance for Machine Learning Algorithms*. http://arxiv.org/abs/2201.05132

Joshi, B. K., Gardner, R. G., & Panthee, D. R. (2011). Gge biplot analysis of tomato F 1 hybrids evaluated across years for marketable fruit yield. *Journal of Crop Improvement*, *25*(5), 488–496. https://doi.org/10.1080/15427528.2011.587138

Jürkenbeck, K., Spiller, A., & Meyerding, S. G. H. (2020). Tomato attributes and consumer preferences – a consumer segmentation approach. *British Food Journal*, *122*(1), 328–344. https://doi.org/10.1108/BFJ-09-2018-0628

Kafkafi, U. (2008). FUNCTIONS OF THE ROOT SYSTEM. In M. Raviv & H. J. Lieth (Eds.), *Soilless Culture: Theory and Practice* (Elsevier, pp. 13–40). https://doi.org/org/10.1016/B978-044452975-6.50004-6

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Kang, M., Ko, E., & Mersha, T. B. (2022). A roadmap for multi-omics data integration using deep learning. In *Briefings in Bioinformatics* (Vol. 23, Issue 1). Oxford University Press. https://doi.org/10.1093/bib/bbab454

Karlova, R., Boer, D., Hayes, S., & Testerink, C. (2021). Root plasticity under abiotic stress. *Plant Physiology*, *187*(3), 1057–1070. https://doi.org/10.1093/plphys/kiab392

Katerji, N., Campi, P., & Mastrorilli, M. (2013). Productivity, evapotranspiration, and water use efficiency of corn and tomato crops simulated by AquaCrop under contrasting water stress conditions in the Mediterranean region. *Agricultural Water Management*, *130*, 14–26. https://doi.org/10.1016/j.agwat.2013.08.005

Kaul, M., Hill, R. L., & Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction. *Agricultural Systems*, *85*(1), 1–18. https://doi.org/10.1016/j.agsy.2004.07.009

Khapte, P. S., Kumar, P., Burman, U., & Kumar, P. (2019). Deficit irrigation in tomato: Agronomical and physio-biochemical implications. In *Scientia Horticulturae* (Vol. 248, pp. 256–264). Elsevier B.V. https://doi.org/10.1016/j.scienta.2019.01.006

Kim, D. S., Lee, D. U., Lim, J. H., Kim, S., & Choi, J. H. (2020). Agreement between visual and model-based classification of tomato fruit ripening. *Transactions of the ASABE*, *63*(3), 667–674. https://doi.org/10.13031/TRANS.13812

Kim, M., Nguyen, T. T. P., Ahn, J. H., Kim, G. J., & Sim, S. C. (2021). Genome-wide association study identifies QTL for eight fruit traits in cultivated tomato (Solanum lycopersicum L.). *Horticulture Research*, *8*(1). https://doi.org/10.1038/s41438-021-00638-4

Kim, Y. X., Son, S. Y., Lee, S., Lee, Y., Sung, J., & Lee, C. H. (2022). Effects of limited water supply on metabolite composition in tomato fruits (Solanum lycopersicum L.) in two soils with different nutrient conditions. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.983725

Koevoets, I. T., Venema, J. H., Elzenga, J. T. M., & Testerink, C. (2016). Roots withstanding their environment: Exploiting root system architecture responses to abiotic stress to improve crop tolerance. In *Frontiers in Plant Science* (Vol. 7, Issue AUG2016). Frontiers Media S.A. https://doi.org/10.3389/fpls.2016.01335

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling* (Vol. 26). New York: Springer.

Kulus, D. (2018). Genetic resources and selected conservation methods of tomato. *Journal of Applied Botany and Food Quality, 91*, 135–144. https://doi.org/10.5073/JABFQ.2018.091.019

Kung, H. Y., Kuo, T. H., Chen, C. H., & Tsai, P. Y. (2016). Accuracy analysis mechanism for agriculture data using the ensemble neural network method. *Sustainability (Switzerland), 8*(8). https://doi.org/10.3390/su8080735

Kuti, J. O., & Konuru, H. B. (2005). Effects of genotype and cultivation environment on lycopene content in red-ripe tomatoes. *Journal of the Science of Food and Agriculture, 85*(12), 2021–2026. https://doi.org/10.1002/jsfa.2205

Kuwata, K., & Shibasaki, R. (2015). Estimating crop yields with deep learning and remotely sensed data. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 858–861.

Kwabena Osei, M., Annor, B., Adjebeng- Danquah, J., Danquah, A., Danquah, E., Blay, E., & Adu-Dapaah, H. (2019). Genotype × Environment Interaction: A Prerequisite for Tomato Variety Development. In *Recent Advances in Tomato Breeding and Production*. IntechOpen. https://doi.org/10.5772/intechopen.76011

Lantini, L., Holleworth, R., Egyir, D., Giannakis, I., Tosti, F., & Alani, A. M. (2018). Use of Ground Penetrating Radar for Assessing Interconnections between Root Systems of Different Matured Tree Species. *Metrology for Archaeology and Cultural Heritage (MetroArchaeo), IEEE*, 22–26.

Le Marié, C., Kirchgessner, N., Marschall, D., Walter, A., & Hund, A. (2014). Rhizoslides: Paper-based growth system for non-destructive, high throughput phenotyping of root development by means of image analysis. *Plant Methods, 10*(1). https://doi.org/10.1186/1746-4811-10-13

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436–444). Nature Publishing Group. https://doi.org/10.1038/nature14539

Lee, D.-K., Jung, H., Jang, G., Seo Jeong, J., Shic Kim, Y., Ha, S.-H., Do Choi, Y., Kim, J.-K., D-kl, K., & conceived, J. (2016). Overexpression of the OsERF71 transcription factor Alters Rice Root Structure and Drought Resistance. *Plant Physiology Preview*. www.plantphysiol.org

Li, Q., Dong, B., Qiao, Y., Liu, M., & Zhang, J. (2010). Root growth, available soil water, and water-use efficiency of winter wheat under different irrigation regimes applied at different growth stages in North China. *Agricultural Water Management, 97*(10), 1676–1682. https://doi.org/10.1016/j.agwat.2010.05.025

Li, X., Šimůnek, J., Shi, H., Yan, J., Peng, Z., & Gong, X. (2017). Spatial distribution of soil water, soil temperature, and plant roots in a drip-irrigated intercropping field with plastic mulch. *European Journal of Agronomy, 83,* 47–56. https://doi.org/10.1016/j.eja.2016.10.015

Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems, 96*. https://doi.org/10.1016/j.compenvurbsys.2022.101845

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. In *Sensors (Switzerland)* (Vol. 18, Issue 8). MDPI AG. https://doi.org/10.3390/s18082674

Lima, E., Hyde, R., & Green, M. (2021). Model selection for inferential models with high dimensional data: synthesis and graphical representation of multiple techniques. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-020-79317-8

Lima, G. P. P., Gómez, H. A. G., Seabra Junior, S., Maraschin, M., Tecchio, M. A., & Borges, C. V. (2022). Functional and Nutraceutical Compounds of Tomatoes as Affected by Agronomic Practices, Postharvest Management, and Processing Methods: A Mini Review. In *Frontiers in Nutrition* (Vol. 9). Frontiers Media S.A. https://doi.org/10.3389/fnut.2022.868492

Liu, C., Liu, W., Chen, W., Yang, J., & Zheng, L. (2015). Feasibility in multispectral imaging for predicting the content of bioactive compounds in intact tomato fruit. *Food Chemistry, 173*, 482–488. https://doi.org/10.1016/j.foodchem.2014.10.052

Liu, J., Hu, T., Feng, P., Yao, D., Gao, F., & Hong, X. (2021). Effect of potassium fertilization during fruit development on tomato quality, potassium uptake, water and potassium use efficiency under deficit irrigation regime. *Agricultural Water Management, 250*. https://doi.org/10.1016/j.agwat.2021.106831

Liu, Z., Huang, Y., Tan, F., Chen, W., & Ou, L. (2021). Effects of Soil Type on Trace Element Absorption and Fruit Quality of Pepper. *Frontiers in Plant Science, 12*. https://doi.org/10.3389/fpls.2021.698796

Lobell, D. B., & Gourdji, S. M. (2012). The influence of climate change on global crop productivity. *Plant Physiology*, *160*(4), 1686–1697. https://doi.org/10.1104/pp.112.208298

Lobet, G., Pagès, L., & Draye, X. (2011). A novel image-analysis toolbox enabling quantitative analysis of root system architecture. *Plant Physiology*, *157*(1), 29–39. https://doi.org/10.1104/pp.111.179895

Lopes, M. S., & Reynolds, M. P. (2010). Partitioning of assimilates to deeper roots is associated with cooler canopies and increased yield under drought in wheat. *Functional Plant Biology*, *37*(2), 147–156. https://doi.org/10.1071/FP09121

Lum, M. S., Hanafi, M. M., Rafii, Y. M., & Akmar, A. S. N. (2014). EFFECT OF DROUGHT STRESS ON GROWTH, PROLINE AND ANTIOXIDANT ENZYME ACTIVITIES OF UPLAND RICE. In *J. Anim. Plant Sci* (Vol. 24, Issue 5).

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). *Consistent Individualized Feature Attribution for Tree Ensembles*. http://arxiv.org/abs/1802.03888

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems, 30*. https://github.com/slundberg/shap

Luo, H., Xu, H., Chu, C., He, F., & Fang, S. (2020). High Temperature can Change Root System Architecture and Intensify Root Interactions of Plant Seedlings. *Frontiers in Plant Science, 11*. https://doi.org/10.3389/fpls.2020.00160

Luypaert, J., Massart, D. L., & Vander Heyden, Y. (2007). Near-infrared spectroscopy applications in pharmaceutical analysis. In *Talanta* (Vol. 72, Issue 3, pp. 865–883). Elsevier. https://doi.org/10.1016/j.talanta.2006.12.023

Ma, L., Shi, Y., Siemianowski, O., Yuan, B., Egner, T. K., Mirnezami, S. V., Lind, K. R., Ganapathysubramanian, B., Venditti, V., & Cademartiri, L. (2019). Hydrogel-based transparent soils for root phenotyping in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *166*(22), 11063–11068. https://doi.org/10.1073/pnas.1820334116

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). *Rectifier Nonlinearities Improve Neural Network Acoustic Models*.

Mairhofer, S., Zappala, S., Tracy, S., Sturrock, C., Bennett, M. J., Mooney, S. J., & Pridmore, T. P. (2013). Recovering complete plant root system architectures from soil via X-ray μ-Computed Tomography. *Plant Methods*, *9*(1). https://doi.org/10.1186/1746-4811-9-8

Malosetti, M., Ribaut, J. M., & van Eeuwijk, F. A. (2013). The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology*, *4 MAR*. https://doi.org/10.3389/fphys.2013.00044

Manschadi, A. M., Christopher, J., Devoil, P., & Hammer, G. L. (2006). The role of root architectural traits in adaptation of wheat to water-limited environments. *Functional Plant Biology*, *33*(9), 823–837. https://doi.org/10.1071/FP06055

Mao, H., Meng, J., Ji, F., Zhang, Q., & Fang, H. (2019). Comparison of machine learning regression algorithms for cotton leaf area index retrieval using Sentinel-2 spectral bands. *Applied Sciences (Switzerland)*, *9*(7). https://doi.org/10.3390/app9071459

Marouelli, W. A., & Silva, W. L. C. (2007). Water tension thresholds for processing tomatoes under drip irrigation in Central Brazil. *Irrigation Science*, *25*(4), 411–418. https://doi.org/10.1007/s00271-006-0056-6

McDowell, N. G., Beerling, D. J., Breshears, D. D., Fisher, R. A., Raffa, K. F., & Stitt, M. (2011). The interdependence of mechanisms underlying climate-driven vegetation mortality. In *Trends in Ecology and Evolution* (Vol. 26, Issue 10, pp. 523–532). https://doi.org/10.1016/j.tree.2011.06.003

McGrail, R. K., Van Sanford, D. A., & McNear, D. H. (2020). Trait-based root phenotyping as a necessary tool for crop selection and improvement. In *Agronomy* (Vol. 10, Issue 9). MDPI. https://doi.org/10.3390/agronomy10091328

Meeusen, R. (2014). Exercise, nutrition and the brain. *Sports Medicine*, *44*(SUPPL.1). https://doi.org/10.1007/s40279-014-0150-5

Mehra, L. K., Cowger, C., Gross, K., & Ojiambo, P. S. (2016). Predicting pre-planting risk of Stagonospora nodorum blotch in winter wheat using machine learning models. *Frontiers in Plant Science*, *7*(MAR2016). https://doi.org/10.3389/fpls.2016.00390

Mendoza, F., Lu, R., Ariana, D., Cen, H., & Bailey, B. (2011). Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content. *Postharvest Biology and Technology*, *62*(2), 149–160. https://doi.org/10.1016/j.postharvbio.2011.05.009

Metzner, R., Eggert, A., van Dusschoten, D., Pflugfelder, D., Gerth, S., Schurr, U., Uhlmann, N., & Jahnke, S. (2015). Direct comparison of MRI and X-ray CT technologies for 3D imaging of root systems in soil: Potential and challenges for root trait quantification. *Plant Methods*, *11*(1). https://doi.org/10.1186/s13007-015-0060-z

Mirzaee, M., & Ghanati, F. (2013). Effects of drought stress on the lipid peroxidation and antioxidant enzyme activities in two canola (Brassica napus L.) cultivars Phytochemistry View project Signaling pathways involved in secondary metabolites biosynthesis View project. *Article in Journal of Agricultural Science and Technology*. https://www.researchgate.net/publication/267037219

Mishra, K. B., Iannacone, R., Petrozza, A., Mishra, A., Armentano, N., La Vecchia, G., Trtílek, M., Cellini, F., & Nedbal, L. (2012). Engineered drought tolerance in tomato plants is reflected in chlorophyll fluorescence emission. *Plant Science*, *182*(1), 79–86. https://doi.org/10.1016/j.plantsci.2011.03.022

Montgomery, D. R., & Biklé, A. (2021). Soil Health and Nutrient Density: Beyond Organic vs. Conventional Farming. In *Frontiers in Sustainable Food Systems* (Vol. 5). Frontiers Media S.A. https://doi.org/10.3389/fsufs.2021.699147

Moore, S., Vrebalov, J., Payton, P., & Giovannoni, J. (2002). Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. *Journal of Experimental Botany*, *53*(377), 2023–2030. https://doi.org/10.1093/jxb/erf057

Moya, C., Oyanedel, E., Verdugo, G., Flores, M. F., Urrestarazu, M., & Álvaro, J. E. (2017). Increased electrical conductivity in nutrient solution management enhances dietary and organoleptic qualities in soilless culture tomato. *HortScience*, *52*(6), 868–872. https://doi.org/10.21273/HORTSCI12026-17

Naeem, M., Zhao, W., Ahmad, N., & Zhao, L. (2023). Beyond green and red: unlocking the genetic orchestration of tomato fruit color and pigmentation. *Functional and Integrative Genomics*, *23*(3). https://doi.org/10.1007/s10142-023-01162-5

Nemeskéri, E., & Helyes, L. (2019). Physiological responses of selected vegetable crop species to water stress. In *Agronomy* (Vol. 9, Issue 8). MDPI AG. https://doi.org/10.3390/agronomy9080447

Nemeskéri, E., Neményi, A., Bocs, A., Pék, Z., & Helyes, L. (2019). Physiological factors and their relationship with the productivity of processing tomato under different water supplies. *Water (Switzerland)*, *11*(3). https://doi.org/10.3390/w11030586

Nguyen, C. V., Vrebalov, J. T., Gapper, N. E., Zheng, Y., Zhong, S., Fei, Z., & Giovannoni, J. J. (2014). Tomato GOLDEN2-LIKE transcription factors reveal molecular gradients that function during fruit development and ripening. *Plant Cell*, *26*(2), 585–601. https://doi.org/10.1105/tpc.113.118794

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. In *Postharvest Biology and Technology* (Vol. 46, Issue 2, pp. 99–118). https://doi.org/10.1016/j.postharvbio.2007.06.024

Nicolas, M. E., Lambers, H., Simpson, R. J., & Dalling, M. J. (1985). Effect of Drought on Metabolism and Partitioning of Carbon in Two Wheat Varieties Differing in Drought-tolerance. In *Annals of Botany* (Vol. 55).

Nketiah, E. A., Chenlong, L., Yingchuan, J., & Aram, S. A. (2023). Recurrent neural network modeling of multivariate time series and its application in temperature forecasting. *PLoS ONE*, *18*(5 May). https://doi.org/10.1371/journal.pone.0285713

Oh, M.-M., Carey, E. E., & Rajashekar, C. B. (2010). Regulated Water Deficits Improve Phytochemical Concentration in Lettuce. *J. AMER. SOC. HORT. SCI*, *135*(3), 223–229.

Oh, S. (2019). Feature interaction in terms of prediction performance. *Applied Sciences (Switzerland)*, *9*(23). https://doi.org/10.3390/app9235191

Oliveira, A. B., Moura, C. F. H., Gomes-Filho, E., Marco, C. A., Urban, L., & Miranda, M. R. A. (2013). The Impact of Organic Farming on Quality of Tomatoes Is Associated to Increased Oxidative Stress during Fruit Development. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0056354

Osmont, K. S., Sibout, R., & Hardtke, C. S. (2007). Hidden branches: Developments in root system architecture. In *Annual Review of Plant Biology* (Vol. 58, pp. 93–113). https://doi.org/10.1146/annurev.arplant.58.032806.104006

Palumbo, M., Pace, B., Cefola, M., Montesano, F. F., Serio, F., Colelli, G., & Attolico, G. (2021). Self-configuring cvs to discriminate rocket leaves according to cultivation practices and to correctly attribute visual quality level. *Agronomy*, *11*(7). https://doi.org/10.3390/agronomy11071353

Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, *121*, 57–65. https://doi.org/10.1016/j.compag.2015.11.018

Panthee, D. R., Cao, C., Debenport, S. J., Rodríguez, G. R., Labate, J. A., Robertson, L. D., Breksa III, A. P., van der Knaap, E., & McSpadden Gardener, B. B. (2012). Magnitude of Genotype · Environment Interactions Affecting Tomato Fruit Quality. *HORTSCIENCE*, *47*(6), 721–726. http://solgenomics.

Paola Carrieri, A., Rowe, W. P., Winn, M., & Pyzer-Knapp, E. O. (2019). A Fast Machine Learning Workflow for Rapid Phenotype Prediction from Whole Shotgun Metagenomes. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 9434–9439. www.aaai.org

Patanè, C., Tringali, S., & Sortino, O. (2011). Effects of deficit irrigation on biomass, yield, water productivity and fruit quality of processing tomato under semi-arid Mediterranean climate conditions. *Scientia Horticulturae*, *129*(4), 590–596. https://doi.org/10.1016/j.scienta.2011.04.030

Pathak, T. B., & Stoddard, C. S. (2018). Climate change effects on the processing tomato growing season in California using growing degree day model. *Modeling Earth Systems and Environment*, *4*(2), 765–775. https://doi.org/10.1007/s40808-018-0460-y

Perkons, U., Kautz, T., Uteau, D., Peth, S., Geier, V., Thomas, K., Lütke Holz, K., Athmann, M., Pude, R., & Köpke, U. (2014). Root-length densities of various annual crops following crops with contrasting root systems. *Soil and Tillage Research*, *137*, 50–57. https://doi.org/10.1016/j.still.2013.11.005

Petropoulos, S. A., Xyrafis, E., Polyzos, N., Antoniadis, V., Fernandes, Â., Barros, L., & Ferreira, I. C. F. R. (2020). The optimization of nitrogen fertilization regulates crop performance and quality of processing tomato (solanum lycopersicum l. cv. heinz 3402). *Agronomy, 10*(5). https://doi.org/10.3390/agronomy10050715

Pfeifer, J., Faget, M., Walter, A., Blossfeld, S., Fiorani, F., Schurr, U., & Nagel, K. A. (2014). Spring barley shows dynamic compensatory root and shoot growth responses when exposed to localised soil compaction and fertilisation. *Functional Plant Biology, 41*(6), 581–597. https://doi.org/10.1071/FP13224

Picon-Cochard, C., Pilon, R., Revaillot, S., Jestin, M., & Dawson, L. (2009). Use of near-infrared reflectance spectroscopy to predict the percentage of dead versus living grass roots. *Plant and Soil, 317*(1–2), 309–320. https://doi.org/10.1007/s11104-008-9810-2

Pigliucci, M., Murren, C. J., & Schlichting, C. D. (2006). Phenotypic plasticity and evolution by genetic assimilation. In *Journal of Experimental Biology* (Vol. 209, Issue 12, pp. 2362–2367). https://doi.org/10.1242/jeb.02070

Pound, M. P., French, A. P., Atkinson, J. A., Wells, D. M., Bennett, M. J., & Pridmore, T. (2013). RootNav: Navigating images of complex root architectures. *Plant Physiology, 162*(4), 1802–1814. https://doi.org/10.1104/pp.113.221531

Prinzenberg, A., Gf Visser, R., Marcelis, L., & Heuvelink, E. (2021). Genetic Mapping of the Tomato Quality Traits Brix and Blossom-end Rot Under Supplemental LED and HPS Lighting Conditions. *Euphytica*, 213–217. https://doi.org/10.21203/rs.3.rs-387667/v1

Prudent, M., Bertin, N., Génard, M., Muños, S., Rolland, S., Garcia, V., Petit, J., Baldet, P., Rothan, C., & Causse, M. (2010). Genotype-dependent response to carbon availability in growing tomato fruit. *Plant, Cell and Environment, 33*(7), 1186–1204. https://doi.org/10.1111/j.1365-3040.2010.02139.x

Quinet, M., Angosto, T., Yuste-Lisbona, F. J., Blanchard-Gros, R., Bigot, S., Martinez, J. P., & Lutts, S. (2019). Tomato Fruit Development and Metabolism. *Frontiers in Plant Science, 10*. https://doi.org/10.3389/fpls.2019.01554

Radočaj, D., Jurišić, M., Gašparović, M., Plaščak, I., & Antonić, O. (2021). Cropland suitability assessment using satellite-based biophysical vegetation properties and machine learning. *Agronomy, 11*(8). https://doi.org/10.3390/agronomy11081620

Rahmat, F., Zulkafli, Z., Ishak, A. J., Abdul Rahman, R. Z., Stercke, S. De, Buytaert, W., Tahir, W., Ab Rahman, J., Ibrahim, S., & Ismail, M. (2023). Supervised feature selection using principal component analysis. *Knowledge and Information Systems*. https://doi.org/10.1007/s10115-023-01993-5

Ramos, P. J., Prieto, F. A., Montoya, E. C., & Oliveros, C. E. (2017). Automatic fruit count on coffee branches using computer vision. *Computers and Electronics in Agriculture, 137*, 9–22. https://doi.org/10.1016/j.compag.2017.03.010

Rao, A. V., Young, G. L., & Rao, L. G. (2018). *Lycopene and Tomatoes in Human Nutrition and Health*. https://doi.org/org/10.1201/9781351110877

Ray, D. K., Gerber, J. S., Macdonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature Communications, 6*. https://doi.org/10.1038/ncomms6989

Raza, A., Razzaq, A., Mehmood, S. S., Zou, X., Zhang, X., Lv, Y., & Xu, J. (2019). Impact of climate change on crops adaptation and strategies to tackle its outcome: A review. In *Plants* (Vol. 8, Issue 2). MDPI AG. https://doi.org/10.3390/plants8020034

Ristova, D., & Busch, W. (2014). Natural variation of root traits: From development to nutrient uptake. *Plant Physiology*, *166*(2), 518–527. https://doi.org/10.1104/pp.114.244749

Roberts, D. P., & Mattoo, A. K. (2019). Sustainable Crop Production Systems and Human Nutrition. In *Frontiers in Sustainable Food Systems* (Vol. 3). Frontiers Media S.A. https://doi.org/10.3389/fsufs.2019.00072

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. In *Ecography* (Vol. 40, Issue 8, pp. 913–929). Blackwell Publishing Ltd. https://doi.org/10.1111/ecog.02881

Rogers, E. D., & Benfey, P. N. (2015). Regulation of plant root system architecture: Implications for crop advancement. In *Current Opinion in Biotechnology* (Vol. 32, pp. 93–98). Elsevier Ltd. https://doi.org/10.1016/j.copbio.2014.11.015

Ruder, S. (2016). *An overview of gradient descent optimization algorithms*. http://arxiv.org/abs/1609.04747

Rusu, O. R., Mangalagiu, I., Amăriucăi-Mantu, D., Teliban, G. C., Cojocaru, A., Burducea, M., Mihalache, G., Roșca, M., Caruso, G., Sekara, A., & Stoleru, V. (2023). Interaction Effects of Cultivars and Nutrition on Quality and Yield of Tomato. *Horticulturae*, *9*(5). https://doi.org/10.3390/horticulturae9050541

Salekdeh, G. H., Reynolds, M., Bennett, J., & Boyer, J. (2009). Conceptual framework for drought phenotyping during molecular breeding. In *Trends in Plant Science* (Vol. 14, Issue 9, pp. 488–496). https://doi.org/10.1016/j.tplants.2009.07.007

Salman, S., & Liu, X. (2019). Overfitting Mechanism and Avoidance in Deep Neural Networks. *ArXiv Preprint*. https://www.youtube.com/watch?v=Qi1Yry33TQE.

Schneider, J. R., Caverzan, A., & Chavarria, G. (2019). Water deficit stress, ROS involvement, and plant performance. In *Archives of Agronomy and Soil Science* (Vol. 65, Issue 8, pp. 1160–1181). Taylor and Francis Ltd. https://doi.org/10.1080/03650340.2018.1556789

Scholander, P. F., Hammel, H. T., Bradstreet, E. D., & Hemmingsen, E. A. (1965). Sap Pressure in Vascular Plants. *Science, 148*(3668), 339–346.

Sengupta, S., & Lee, W. S. (2014). Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions. *Biosystems Engineering, 117*(C), 51–61. https://doi.org/10.1016/j.biosystemseng.2013.07.007

Senthilnath, J., Dokania, A., Kandukuri, M., K.N., R., Anand, G., & Omkar, S. N. (2016). Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosystems Engineering*, *146*, 16–32. https://doi.org/10.1016/j.biosystemseng.2015.12.003

Shahhosseini, M., Hu, G., & Pham, H. (2022). Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Machine Learning with Applications*, *7*, 100251. https://doi.org/10.1016/j.mlwa.2022.100251

Shahid, N. (2023). Comparison of hierarchical clustering and neural network clustering: an analysis on precision dominance. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-32790-3

Sharma, A., Tiwari, A. D., Kumari, M., Kumar, N., Saxena, V., & Kumar, R. (2023). Artificial intelligence-based prediction of lycopene content in raw tomatoes using physicochemical attributes. *Phytochemical Analysis*, *34*(7), 729–744. https://doi.org/10.1002/pca.3185

Sharma, S. P., Leskovar, D. I., Crosby, K. M., Volder, A., & Ibrahim, A. M. H. (2014). Root growth, yield, and fruit quality responses of reticulatus and inodorus melons (Cucumis melo L.) to deficit subsurface drip irrigation. *Agricultural Water Management*, *136*, 75–85. https://doi.org/10.1016/j.agwat.2014.01.008

Shi, X., Hsieh, S. J., & Romero, R. A. F. (2021). FEA and machine learning techniques for hidden structure analysis. *Sensors*, *21*(15). https://doi.org/10.3390/s21155159

Shin, Y., Ryu, J. A., Liu, R. H., Nock, J. F., & Watkins, C. B. (2008). Harvest maturity, storage temperature and relative humidity affect fruit quality, antioxidant contents and activity, and inhibition of cell proliferation of strawberry fruit. *Postharvest Biology and Technology*, *49*(2), 201–209. https://doi.org/10.1016/j.postharvbio.2008.02.008

Silva, R. S., Kumar, L., Shabani, F., & Picanço, M. C. (2017). Assessing the impact of global warming on worldwide open field tomato cultivation through CSIRO-Mk3.0 global climate model. *Journal of Agricultural Science*, *155*(3), 407–420. https://doi.org/10.1017/S0021859616000654

Singh, C., Rai, P., & BK, R. (2017). Genetic Engineering for Drought Resistance in Rice. *Agricultural Research & Technology: Open Access Journal*, *11*(5). https://doi.org/10.19080/artoaj.2017.11.555824

Singh, J., Heeren, D. M., Rudnick, D. R., Woldt, W. E., Bai, G., Ge, Y., & Luck, J. D. (2020). Soil structure and texture effects on the precision of soil water content measurements with a capacitance-based electromagnetic sensor. *Transactions of the ASABE*, *63*(1), 141–152. https://doi.org/10.13031/trans.13496

Slack, S., York, L. M., Roghazai, Y., Lynch, J., Bennett, M., & Foulkes, J. (2018). Wheat shovelomics II: Revealing relationships between root crown traits and crop growth. *BioRxiv*, 280917.

Slavin, J. L., & Lloyd, B. (2012). Health benefits of fruits and vegetables. In *Advances in Nutrition* (Vol. 3, Issue 4, pp. 506–516). https://doi.org/10.3945/an.112.002154

Song, K., Yan, F., Ding, T., Gao, L., & Lu, S. (2020). A steel property optimization model based on the XGBoost algorithm and improved PSO. *Computational Materials Science*, *174*. https://doi.org/10.1016/j.commatsci.2019.109472

Srivastava, S., & Srivastava, A. K. (2015). Lycopene; chemistry, biosynthesis, metabolism and degradation under various abiotic parameters. *Journal of Food Science and Technology*, *52*(1), 41–53. https://doi.org/10.1007/s13197-012-0918-2

Stewart, A. L., & Ahmed, S. (2019). Effects of climate change on fruit nutrition. In *Fruit Crops: Diagnosis and Management of Nutrient Constraints* (pp. 77–93). Elsevier. https://doi.org/10.1016/B978-0-12-818732-6.00007-1

Su, Y. xue, Xu, H., & Yan, L. jiao. (2017). Support vector machine-based open crop model (SBOCM): Case of rice production in China. *Saudi Journal of Biological Sciences*, *24*(3), 537–547. https://doi.org/10.1016/j.sjbs.2017.01.024

Suresh, S., Newton, D. T., Everett, T. H., Lin, G., & Duerstock, B. S. (2022). Feature Selection Techniques for a Machine Learning Model to Detect Autonomic Dysreflexia. *Frontiers in Neuroinformatics*, *16*. https://doi.org/10.3389/fninf.2022.901428

Swarup, S., Cargill, E. J., Crosby, K., Flagel, L., Kniskern, J., & Glenn, K. C. (2021). Genetic diversity is indispensable for plant breeding to improve crops. In *Crop Science* (Vol. 61, Issue 2, pp. 839–852). John Wiley and Sons Inc. https://doi.org/10.1002/csc2.20377

Takács, S., Csengeri, E., Pék, Z., Bíró, T., Szuvandzsiev, P., Palotás, G., & Helyes, L. (2021). Performance evaluation of aquacrop model in processing tomato biomass, fruit yield and water stress indicator modelling. *Water (Switzerland)*, *13*(24). https://doi.org/10.3390/w13243587

Takács, S., Pék, Z., Csányi, D., Daood, H. G., Szuvandzsiev, P., Palotás, G., & Helyes, L. (2020). Influence of water stress levels on the yield and lycopene content of tomato. *Water (Switzerland)*, *12*(8). https://doi.org/10.3390/W12082165

Takahashi, N., Yamazaki, Y., Kobayashi, A., Higashitani, A., & Takahashi, H. (2003). Hydrotropism Interacts with Gravitropism by Degrading Amyloplasts in Seedling Roots of Arabidopsis and Radish. *PLANT PHYSIOLOGY*, *132*(2), 805–810. https://doi.org/10.1104/pp.102.018853

Thole, V., Vain, P., Yang, R. Y., Almeida Barros da Silva, J., Enfissi, E. M. A., Nogueira, M., Price, E. J., Alseekh, S., Fernie, A. R., Fraser, P. D., Hanson, P., & Martin, C. (2020). Analysis of Tomato Post-Harvest Properties: Fruit Color, Shelf Life, and Fungal Susceptibility. *Current Protocols in Plant Biology*, *5*(2). https://doi.org/10.1002/cppb.20108

Thompson, A., Maskery, I., & Leach, R. K. (2016). X-ray computed tomography for additive manufacturing: A review. In *Measurement Science and Technology* (Vol. 27, Issue 7). Institute of Physics Publishing. https://doi.org/10.1088/0957-0233/27/7/072001

Trachsel, S., Kaeppler, S. M., Brown, K. M., & Lynch, J. P. (2011). Shovelomics: High throughput phenotyping of maize (Zea mays L.) root architecture in the field. *Plant and Soil*, *341*(1–2), 75–87. https://doi.org/10.1007/s11104-010-0623-8

Tracy, S. R., Nagel, K. A., Postma, J. A., Fassbender, H., Wasson, A., & Watt, M. (2020). Crop Improvement from Phenotyping Roots: Highlights Reveal Expanding Opportunities. In *Trends in Plant Science* (Vol. 25, Issue 1, pp. 105–118). Elsevier Ltd. https://doi.org/10.1016/j.tplants.2019.10.015

Tripodi, P., D’Alessandro, A., & Francese, G. (2023). An integrated genomic and biochemical approach to investigate the potentiality of heirloom tomatoes: Breeding resources for food quality and sustainable agriculture. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.1031776

Tsai, H. Y., Janss, L. L., Andersen, J. R., Orabi, J., Jensen, J. D., Jahoor, A., & Jensen, J. (2020). Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-60203-2

Udriște, A. A., Iordachescu, M., Ciceoi, R., & Bădulescu, L. (2022). Next-Generation Sequencing of Local Romanian Tomato Varieties and Bioinformatics Analysis of the Ve Locus. *International Journal of Molecular Sciences*, *23*(17). https://doi.org/10.3390/ijms23179750

Umrani, M. A. T., Mrs, C., & Aitwade, S. A. (2021). APPLICATION OF MACHINE LEARNING AND DEEP LEARNING IN SMART AGRICULTURE. In *International Journal of Engineering Applied Sciences and Technology* (Vol. 6). http://www.ijeast.com

Vadez, V., Rao, J. S., Bhatnagar-Mathur, P., & Sharma, K. K. (2013). DREB1A promotes root development in deep soil layers and increases water extraction under water stress in groundnut. *Plant Biology*, *15*(1), 45–52. https://doi.org/10.1111/j.1438-8677.2012.00588.x

Vallarino, J. G., Kubiszewski-Jakubiak, S., Ruf, S., Rößner, M., Timm, S., Bauwe, H., Carrari, F., Rentsch, D., Bock, R., Sweetlove, L. J., & Fernie, A. R. (2020). Multi-gene metabolic engineering of tomato plants results in increased fruit yield up to 23%. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-73709-6

Van Dusschoten, D., Metzner, R., Kochs, J., Postma, J. A., Pflugfelder, D., Bühler, J., Schurr, U., & Jahnke, S. (2016). Quantitative 3D analysis of plant roots growing in soil using magnetic resonance imaging. *Plant Physiology*, *170*(3), 1176–1188. https://doi.org/10.1104/pp.15.01388

Vela-Hinojosa, C., Escalona-Buendía, H. B., Mendoza-Espinoza, J. A., Villa-Hernández, J. M., Lobato-Ortíz, R., Rodríguez-Pérez, J. E., & Pérez-Flores, L. J. (2019). Antioxidant balance and regulation in tomato genotypes of different color. *Journal of the American Society for Horticultural Science*, *144*(1), 45–54. https://doi.org/10.21273/JASHS04525-18

Vélez-Terreros, P. Y., Romero-Estévez, D., Yánez-Jácome, G. S., Simbaña-Farinango, K., & Navarrete, H. (2021). Comparison of major nutrients and minerals between organic and conventional tomatoes. A review. In *Journal of Food Composition and Analysis* (Vol. 100). Academic Press Inc. https://doi.org/10.1016/j.jfca.2021.103922

Venkatesan, S., Lim, J., Ko, H., & Cho, Y. (2022). A Machine Learning Based Model for Energy Usage Peak Prediction in Smart Farms. *Electronics (Switzerland)*, *11*(2). https://doi.org/10.3390/electronics11020218

Waines, J. G., & Ehdaie, B. (2007). Domestication and crop physiology: Roots of green-revolution wheat. In *Annals of Botany* (Vol. 100, Issue 5, pp. 991–998). https://doi.org/10.1093/aob/mcm180

Waisel, Yoav., Eshel, A., & Kafkafi, U. (1996). *Plant roots : the hidden half*. Marcel Dekker.

Wang, H., Inukai, Y., & Yamauchi, A. (2006). Root development and nutrient uptake. In *Critical Reviews in Plant Sciences* (Vol. 25, Issue 3, pp. 279–301). https://doi.org/10.1080/07352680600709917

Wang, M. W. H., Goodman, J. M., & Allen, T. E. H. (2020). Machine learning in predictive toxicology: recent applications and future directions for classification models. *Chemical Research in Toxicology*, *34*(2), 217–239.

Wang, X., Jiang, D., & Lang, X. (2017). Future extreme climate changes linked to global warming intensity. *Science Bulletin*, *62*(24), 1673–1680. https://doi.org/10.1016/j.scib.2017.11.004

Wang, X., Yun, J., Shi, P., Li, Z., Li, P., & Xing, Y. (2019). Root Growth, Fruit Yield and Water Use Efficiency of Greenhouse Grown Tomato Under Different Irrigation Regimes and Nitrogen Levels. *Journal of Plant Growth Regulation*, *38*(2), 400–415. https://doi.org/10.1007/s00344-018-9850-7

Wasaya, A., Zhang, X., Fang, Q., & Yan, Z. (2018). Root phenotyping for drought tolerance: A review. *Agronomy*, *8*(11). https://doi.org/10.3390/agronomy8110241

Waskom, M. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

White, P. J., George, T. S., Gregory, P. J., Bengough, A. G., Hallett, P. D., & McKenzie, B. M. (2013). Matching roots to their environment. *Annals of Botany*, *112*(2), 207–222. https://doi.org/10.1093/aob/mct123

Wolpert, D. H. (1992). Stacked Generalization. *Neural Networks*, *5*(2), 241–259.

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, *14*(7), 507–515. https://doi.org/10.1038/nrg3457

Wu, B., Zhang, L., Tian, J., Zhang, G., & Zhang, W. (2023). Fine root dynamics, longevity, and canopy characteristics of cotton under varying water and nitrogen levels. *Plant and Soil*, *482*(1–2), 191–209. https://doi.org/10.1007/s11104-022-05681-1

Wyatt, S. E., & Kiss, J. Z. (2013). Plant tropisms: From darwin to the international space station. *American Journal of Botany*, *100*(1), 1–3. https://doi.org/10.3732/ajb.1200591

Xiao, L., Shibuya, T., Kato, K., Nishiyama, M., & Kanayama, Y. (2022). Effects of light quality on plant development and fruit metabolism and their regulation by plant growth regulators in tomato. In *Scientia Horticulturae* (Vol. 300). Elsevier B.V. https://doi.org/10.1016/j.scienta.2022.111076

Xu, C., Tao, H., Tian, B., Gao, Y., Ren, J., & Wang, P. (2016). Limited-irrigation improves water use efficiency and soil reservoir capacity through regulating root and canopy growth of winter wheat. *Field Crops Research*, *196*, 268–275. https://doi.org/10.1016/j.fcr.2016.07.009

Xu, Y., Deng, S., Ma, L., Li, M., Xie, B., Gao, J., Shao, M., & Chen, Y. (2023). Effects of Soil Properties and Nutrients on the Fruit Economic Parameters and Oil Nutrient Contents of Camellia oleifera. *Forests*, *14*(9). https://doi.org/10.3390/f14091786

Yadav, N., Md Alfayeed, S., & Wadhawan, A. (2020). MACHINE LEARNING IN AGRICULTURE: TECHNIQUES AND APPLICATIONS. In *International Journal of Engineering Applied Sciences and Technology* (Vol. 5). http://www.ijeast.com

Yan, W. (2001). GGEbiplot-A Windows Application for Graphical Analysis of Multienvironment Trial Data and Other Types of Two-Way Data. *Agronomy Journal*, *93*, 1111–1118. http://www.ptinet.

Yan, W. (2019). LG biplot: a graphical method for mega-environment investigation using existing crop variety trial data. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-43683-9

Yan, W., Hunt, L. A., Sheng, Q., & Szlavnics, Z. (2000). Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. *Crop Science*, *40*(3), 697–605.

Yan, W., & Kang, M. S. (2003). *GGE Biplot Analysis; A Graphical Tool for Breeders, Geneticists, and Agronomists*. CRC Press, Boca Raton, FL. https://doi.org/10.4324/9780429122729

Yan, W., Kang, M. S., Ma, B., Woods, S., & Cornelius, P. L. (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Science*, *47*(2), 643–655. https://doi.org/10.2135/cropsci2006.06.0374

Yan, W., & Rajcan, I. (2002). Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Science*, *42*(1), 11–20. https://doi.org/10.2135/cropsci2002.1100

Yan, W., & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *CANADIAN JOURNAL OF PLANT SCIENCE*, *86*, 623–645.

You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1). www.aaai.org

Yu, G., Zhang, S., Li, S., Zhang, M., Benli, H., & Wang, Y. (2022). Numerical investigation for effects of natural light and ventilation on 3D tomato body heat distribution in a Venlo greenhouse. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2022.05.006

Zhang, P., Jia, Y., & Shang, Y. (2022). Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*, *18*(6). https://doi.org/10.1177/15501329221106935

Zhang, X., Derival, M., Albrecht, U., & Ampatzidis, Y. (2019). Evaluation of a ground penetrating radar to map the root architecture of HLB-infected citrus trees. *Agronomy*, *9*(7). https://doi.org/10.3390/agronomy9070354

Zhang, Y., Kang, M. S., & Lamkey, K. R. (2005). DIALLEL-SAS05: A Comprehensive Program for Griffing's and Gardner–Eberhart Analyses. *Agronomy Journal*, *97*(4), 1097–1106. https://doi.org/10.2134/agronj2004.0260

Zhang, Y., Ma, Y., Zhao, D., Tang, Z., Zhang, T., Zhang, K., Dong, J., & Zhang, H. (2022). Genetic regulation of lateral root development. In *Plant Signaling and Behavior*. Taylor and Francis Ltd. https://doi.org/10.1080/15592324.2022.2081397

Zhao, W., Li, Y., Fan, S., Wen, T., Wang, M., Zhang, L., & Zhao, L. (2021). The transcription factor WRKY32 affects tomato fruit colour by regulating YELLOW FRUITED-TOMATO 1, a core component of ethylene signal transduction. *Journal of Experimental Botany*, *72*(12), 4269–4282. https://doi.org/10.1093/jxb/erab113

Zheng, Y., Yang, Z., Wei, T., & Zhao, H. (2022). Response of Tomato Sugar and Acid Metabolism and Fruit Quality under Different High Temperature and Relative Humidity Conditions. *Phyton-International Journal of Experimental Botany*, *91*(9), 2033–2054. https://doi.org/10.32604/phyton.2022.019468

Zhu, F., Wen, W., Cheng, Y., & Fernie, A. R. (2022). The metabolic changes that effect fruit quality during tomato fruit ripening. *Molecular Horticulture*, *2*(1), 1–19. https://doi.org/10.1186/s43897-022-00024-1

Zhu, J., Ingram, P. A., Benfey, P. N., & Elich, T. (2011). From lab to field, new approaches to phenotyping root system architecture. In *Current Opinion in Plant Biology* (Vol. 14, Issue 3, pp. 310–317). https://doi.org/10.1016/j.pbi.2011.03.020

Zhu, Y. H., Weiner, J., Yu, M. X., & Li, F. M. (2019). Evolutionary agroecology: Trends in root architecture during wheat breeding. *Evolutionary Applications*, *12*(4), 733–743. https://doi.org/10.1111/eva.12749

Zorrilla-Fontanesi, Y., Rambla, J. L., Cabeza, A., Medina, J. J., Sánchez-Sevilla, J. F., Valpuesta, V., Botella, M. A., Granell, A., & Amaya, I. (2012). Genetic analysis of strawberry fruit aroma and identification of O-methyltransferase FaOMT as the locus controlling natural variation in mesifurane content. *Plant Physiology*, *159*(2), 851–870. https://doi.org/10.1104/pp.111.188318

Zotarelli, L., Scholberg, J. M., Dukes, M. D., Muñoz-Carpena, R., & Icerman, J. (2009). Tomato yield, biomass accumulation, root distribution and irrigation water use efficiency on a sandy soil, as affected by nitrogen rate and irrigation scheduling. *Agricultural Water Management*, *96*(1), 23–34. https://doi.org/10.1016/j.agwat.2008.06.007

Zsögön, A., Čermák, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., Freschi, L., Voytas, D. F., Kudla, J., & Peres, L. E. P. (2018). De novo domestication of wild tomato using genome editing. *Nature Biotechnology*, *36*(12), 1211–1216. https://doi.org/10.1038/nbt.4272

Zuo, H. (2022). *Analysis and Detection of Tomatoes Quality using Machine Learning Algorithm and Image Processing*. https://doi.org/10.21203/rs.3.rs-2016895/v1

10 APPENDICES

Appendix 1: General Results Regarding Root Count and Root Length

ANOVA - 2020 Root_Count ~ Treatment

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	121816.4	60908.19	26.02065	0
Residuals	159	372181.4	2340.764	NA	NA

Tukey HSD test - 2020 Root_Count ~ Treatment

	diff	lwr	upr	P
I50-I100	60.7963	38.76767	82.82493	0
Control-I100	55.12963	33.101	77.15826	1E-07
Control-I50	-5.66667	-27.6953	16.36196	0.815637

ANOVA - 2020 Root_Length ~ Treatment

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	1.13E+08	56519043	22.7319	0
Residuals	159	3.95E+08	2486332	NA	NA

Tukey HSD test - 2020 Root_Length ~ Treatment

	diff	lwr	upr	p
I50-I100	1871.13	1153.19	2589.07	0
Control-I100	1652.574	934.634	2370.514	6E-07
Control-I50	-218.556	-936.496	499.3845	0.751897

ANOVA - 2021 Root_Count ~ Treatment

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	20198.53	10099.27	7.189406	0.001025
Residuals	159	223354.1	1404.743	NA	NA

Tukey HSD test - 2021 Root_Count ~ Treatment

	diff	lwr	upr	p
I50-I100	-2.66667	-19.7317	14.39836	0.927467
Control-I100	22.24074	5.17571	39.30577	0.006784
Control-I50	24.90741	7.842376	41.97244	0.002042

ANOVA - 2021 Root_Length ~ Treatment

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	8155425	4077712	1.445669	0.238668
Residuals	159	4.48E+08	2820641	NA	NA

Tukey HSD test - 2021 Root_Length ~ Treatment

	diff	lwr	upr	p
I50-I100	-255.019	-1019.7	509.6665	0.710327
Control-I100	294.1111	-470.574	1058.796	0.634821
Control-I50	549.1296	-215.555	1313.815	0.20876

ANOVA - 2020 Root_Count ~ Layer

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	103899.4	51949.71	21.17416	0
Residuals	159	390098.4	2453.449	NA	NA

Tukey HSD test - 2020 Root_Count ~ Layer

	diff	lwr	upr	p
30-50/10-30cm	56.055556	33.50293	78.60819	0.0000001
50-70/10-30cm	51.037037	28.48441	73.58967	0.0000009
50-70/30-50cm	-5.018519	-27.57115	17.53411	0.8584969

ANOVA - 2020 Root_Length ~ Layer

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	144710084	72355042	31.63564	0
Residuals	159	363654734	2287137	NA	NA

Tukey HSD test - 2020 Root_Length ~ Layer

	diff	lwr	upr	p
30-50/10-30cm	1969.77778	1281.1972	2658.3583	0
50-70/10-30cm	2038.31481	1349.7343	2726.8954	0
50-70/30-50cm	68.53704	-620.0435	757.1176	0.9698992

ANOVA - 2021 Root_Count ~ Layer

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	12296.35	6148.173	4.227169	0.016267
Residuals	159	231256.28	1454.442	NA	NA

Tukey HSD test - 2020 Root_Count ~ Layer

	diff	lwr	upr	p
30-50/10-30cm	10.18519	-7.179099	27.54947	0.3497452
50-70/10-30cm	21.33333	3.969049	38.69762	0.0115706
50-70/30-50cm	11.14815	-6.216136	28.51243	0.2847866

ANOVA - 2021 Root_Length ~ Layer

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	50499006	25249503	9.884983	8.99E-05
Residuals	159	406138366	2554329	NA	NA

Tukey HSD test - 2021 Root_Length ~ Treatment

	diff	lwr	upr	p
30-50/10-30cm	536.7593	-190.93196	1264.45	0.1917719
50-70/10-30cm	1357.7222	630.03101	2085.413	0.0000552
50-70/30-50cm	820.963	93.27175	1548.654	0.0227258

Appendix 2: Evaluation of the Layer Scale for the Monitored Root Zone

2020 ANOVA for Root_Count between Treatment within Layer 10-30cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	19522	9761	18.55	8.82E-07
Residuals	51	26831	526		

Tukey HSD test - 2020 Root_Count between Treatment within Layer 10-30cm

	diff	lwr	upr	p
I50-I100	35.333333	16.876913	53.78975	0.000077
Control-I100	43.944444	25.488024	62.40086	0.0000015
Control-I50	8.611111	-9.845309	27.06753	0.5025163

2020 ANOVA for Root_Count between Treatment within Layer 30-50cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	49507	24754	10.46	0.000156
Residuals	51	120726	2367		

Tukey HSD test - 2020 Root_Count between Treatment within Layer 30-50cm

	diff	lwr	upr	p
I50-I100	74.16667	35.017029	113.316304	0.0000906
Control-I100	36.77778	-2.371859	75.927415	0.0696615
Control-I50	-37.38889	-76.538526	1.760748	0.0640538

2020 ANOVA for Root_Count between Treatment within Layer 50-70cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	75720	37860	19.75	4.47E-07
Residuals	51	97792	1917		

Tukey HSD test - 2020 Root_Count between Treatment within Layer 50-70cm

	diff	lwr	upr	p
I50-I100	72.88889	37.65355	108.12422	0.0000216
Control-I100	84.66667	49.43133	119.902	0.0000012
Control-I50	11.77778	-23.45756	47.01311	0.7004982

2020 ANOVA for Root_Count between Layer within Treatment I100

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	18690	9345	6.603	8.82E-07
Residuals	51	72175	1415		

Tukey HSD test - 2020 Root_Count between Layer within Treatment I100

	diff	lwr	upr	p
30-50/10-30cm	45.5	15.229346	75.770654	0.0018779
50-70/10-30cm	24.94444	-5.326209	55.215098	0.1251788
50-70/30-50cm	-20.55556	-50.826209	9.715098	0.2387139

2020 ANOVA for Root_Count between Layer within Treatment I50

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	68970	34485	13.72	1.71E-05
Residuals	51	128176	2513		

Tukey HSD test - 2020 Root_Count between Layer within Treatment I50

	diff	lwr	upr	p
30-50/10-30cm	84.33333	43.99388	124.67279	0.0000179
50-70/10-30cm	62.5	22.16054	102.83946	0.0013357
50-70/30-50cm	-21.83333	-62.17279	18.50612	0.398126

2020 ANOVA for Root_Count between Layer within Treatment Control

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	39172	19586	22.2	1.16E-07
Residuals	51	44998	882		

Tukey HSD test - 2020 Root_Count between Layer within Treatment Control

	diff	lwr	upr	p
30-50/10-30cm	38.33333	14.431906	62.23476	0.0008883
50-70/10-30cm	65.66667	41.765239	89.56809	0.0000001
50-70/30-50cm	27.33333	3.431906	51.23476	0.0214497

2020 ANOVA for Root_Length between Treatment within Layer 10-30cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	14147014	7073507	18.73	7.96E-07
Residuals	51	19259253	377632		

Tukey HSD test - 2020 Root_ Length between Treatment within Layer = 1

	diff	lwr	upr	p
I50-I100	825.5556	331.0776	1320.0335	0.0005379
Control-I100	1229.9444	735.4665	1724.4224	0.0000006
Control-I50	404.3889	-90.0891	898.8669	0.12898

2020 ANOVA for Root_ Length between Treatment within Layer 30-50cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	32358213	16179107	8.14	8.55E-04
Residuals	51	101362227	198749		

Tukey HSD test - 2020 Root_ Length between Treatment within Layer 30-50cm

	diff	lwr	upr	p
I50-I100	1845.2222	710.8246	2979.6198	0.0007473
Control-I100	544.6111	-589.7865	1679.0087	0.4829137
Control-I50	-1300.6111	-2435.0087	-166.2135	0.0210642

2020 ANOVA for Root_ Length between Treatment within Layer 50-70cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	113096263	56548131	34.57	3.25E-10
Residuals	51	83431763	1635917		

Tukey HSD test - 2020 Root_ Length between Treatment within Layer 50-70cm

	diff	lwr	upr	p
I50-I100	2942.6111	1913.4273	3971.795	0
Control-I100	3183.1667	2153.9829	4212.35	0
Control-I50	240.5556	-788.6283	1269.739	0.8396278

2020 ANOVA for Root_ Length between Layer within Treatment I100

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	31816360	15908180	6.603	8.82E-07
Residuals	51	53562366	1050242		

Tukey HSD test - 2020 Root_ Length between Layer within Treatment I100

	diff	lwr	upr	p
30-50/10-30cm	1858.3333	1033.7072	2682.9595	0.0000045
50-70/10-30cm	681.5556	-143.0706	1506.1817	0.123707
50-70/30-50cm	-1176.7778	-2001.4039	-352.1516	0.0032487

2020 ANOVA for Root_ Length between Layer within Treatment I50

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	96728464	48364232	21.72	1.50E-07
Residuals	51	113581786	2227094		

Tukey HSD test - 2020 Root_Count between Layer within Treatment I50

	diff	lwr	upr	p
30-50/10-30cm	2878	1677.17	4078.83	0.0000013
50-70/10-30cm	2798.61111	1597.781	3999.441	0.0000023
50-70/30-50cm	-79.38889	-1280.219	1121.441	0.9860605

2020 ANOVA for Root_ Length between Layer within Treatment Control

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	62728663	31364332	43.34	1.00E-11
Residuals	51	36909092	723708		

Tukey HSD test - 2020 Root_ Length between Layer within Treatment Control

	diff	lwr	upr	p
30-50/10-30cm	1173	488.4677	1857.532	0.0003825
50-70/10-30cm	2634.778	1950.2455	3319.31	0
50-70/30-50cm	1461.778	777.2455	2146.31	0.0000123

2021 ANOVA for Root_Count between Treatment within Layer 10-30cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	7092	3546	5.281	8.23E-03
Residuals	51	34247	672		

Tukey HSD test - 2021 Root_Count between Treatment within Layer 10-30cm

	diff	lwr	upr	p
I50-I100	0.6111111	-20.240565	21.46279	0.9972444
Control-I100	24.6111111	3.759435	45.46279	0.0170578
Control-I50	24	3.148324	44.85168	0.0204897

2021 ANOVA for Root_Count between Treatment within Layer 30-50cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	3489	1744	1.51	2.31E-01
Residuals	51	58936	1156		

Tukey HSD test - 2020 Root_Count between Treatment within Layer 30-50cm

	diff	lwr	upr	p
I50-I100	-4.444444	-31.798302	22.90941	0.9188371
Control-I100	14.388889	-12.964969	41.74275	0.4185481
Control-I50	18.833333	-8.520525	46.18719	0.2296275

2021 ANOVA for Root_Count between Treatment within Layer 50-70cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	10817	5408	2.364	1.04E-01
Residuals	51	116675	2288		

Tukey HSD test - 2021 Root_Count between Treatment within Layer 50-70cm

	diff	lwr	upr	p
I50-I100	-4.166667	-42.653769	34.32044	0.9630771
Control-I100	27.722222	-10.76488	66.20932	0.2008332
Control-I50	31.888889	-6.598213	70.37599	0.122484

2021 ANOVA for Root_Count between Layer within Treatment I100

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	4537	2269	6.603	8.82E-07
Residuals	51	73218	1436		

Tukey HSD test - 2021 Root_Count between Layer within Treatment I100

	diff	lwr	upr	p
30-50/10-30cm	15.277778	-15.210707	45.76626	0.4530108
50-70/10-30cm	21.888889	-8.599596	52.37737	0.2028766
50-70/30-50cm	6.611111	-23.877373	37.0996	0.8602547

2021 ANOVA for Root_Count between Layer within Treatment I50

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	2668	1334	1.175	1.71E-05
Residuals	51	57891	1135		

Tukey HSD test - 2021 Root_Count between Layer within Treatment I50

	diff	lwr	upr	p
30-50/10-30cm	10.222222	-16.887954	37.3324	0.6363921
50-70/10-30cm	17.111111	-9.999065	44.22129	0.2885011
50-70/30-50cm	6.888889	-20.221288	33.99907	0.8134734

2021 ANOVA for Root_Count between Layer within Treatment Control

	Df	SumSq	MeanSq	Fvalue	Pr(>F)

Layer	2	6290	3145	2.037	1.41E-01
Residuals	51	78749	1544		

Tukey HSD test - 2021 Root_Count between Layer within Treatment Control

	diff	lwr	upr	p
30-50/10-30cm	5.055556	-26.563641	36.67475	0.9212962
50-70/10-30cm	25	-6.619197	56.6192	0.146596
50-70/30-50cm	19.944444	-11.674752	51.56364	0.288942

2021 ANOVA for Root_Length between Treatment within Layer 10-30cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	4911029	2455514	2.758	7.96E-07
Residuals	51	45403026	890255		

Tukey HSD test - 2021 Root_ Length between Treatment within Layer 10-30cm

	diff	lwr	upr	p
I50-I100	-581.9444	-1341.16793	177.279	0.1637863
Control-I100	103.0556	-656.16793	862.279	0.9426012
Control-I50	685	-74.22349	1444.223	0.0847053

2021 ANOVA for Root_ Length between Treatment within Layer 30-50cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	3091728	1545864	8.14	8.55E-04
Residuals	51	108632413	2130047		

Tukey HSD test - 2021 Root_ Length between Treatment within Layer 30-50cm

	diff	lwr	upr	p
I50-I100	-333.9444	-1508.3199	840.431	0.7724347
Control-I100	250.1667	-924.2087	1424.542	0.8647716
Control-I50	584.1111	-590.2643	1758.487	0.4582512

2021 ANOVA for Root_ Length between Treatment within Layer 50-70cm

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Treatment	2	2674820	1337410	0.283	7.55E-01
Residuals	51	241425351	4733830		

Tukey HSD test - 2021 Root_ Length between Treatment within Layer 50-70cm

	diff	lwr	upr	p
I50-I100	150.8333	-1599.895	1901.561	0.9764481
Control-I100	529.1111	-1221.617	2279.839	0.7471857
Control-I50	378.2778	-1372.45	2129.006	0.8611709

2021 ANOVA for Root_ Length between Layer within Treatment I100

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	8571424	4285712	1.494	2.34E-01
Residuals	51	146332593	2869267		

Tukey HSD test - 2021 Root_ Length between Layer within Treatment I100

	diff	lwr	upr	p
30-50/10-30cm	405.0556	-957.9505	1768.062	0.7543826
50-70/10-30cm	971.4444	-391.5616	2334.45	0.2074481
50-70/30-50cm	566.3889	-796.6171	1929.395	0.5783

2021 ANOVA for Root_ Length between Layer within Treatment I50

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	26614838	13307419	5.27	8.31E-03
Residuals	51	128785799	2525212		

Tukey HSD test - 2021 Root_Count between Layer within Treatment I50

	diff	lwr	upr	p
30-50/10-30cm	653.0556	-625.6225	1931.734	0.4395622
50-70/10-30cm	1704.2222	425.5441	2982.9	0.0062621
50-70/30-50cm	1051.1667	-227.5114	2329.845	0.1263722

2020 ANOVA for Root_ Length between Layer within Treatment Control

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Layer	2	17834896	8917448	3.779	2.95E-02
Residuals	51	120342397	2359655		

Tukey HSD test - 2020 Root_ Length between Layer within Treatment Control

	diff	lwr	upr	p
30-50/10-30cm	552.1667	-683.8848	1788.218	0.5317047
50-70/10-30cm	1397.5	161.4485	2633.551	0.0232304
50-70/30-50cm	845.3333	-390.7182	2081.385	0.2340297

Appendix 3: Comparison of the Root Development in the Two Years

Root_Count between 2020 and 2021

	t	df	p-value
Treatment1	-2.6466	105.36	0.009376
Treatment2	-8.8288	82.754	1.47e-13
Treatment3	-6.9193	106	3.583e-10

Root_Length between 2020 and 2021

	t	df	p-value
Treatment1	-0.4447	97.811	0.6575
Treatment2	-6.3083	103.66	7.047e-09
Treatment3	-5.1596	103.29	1.202e-06

Root_Count between 2020 and 2021

	t	df	p-value
Layer1	-4.9212	105.65	3.182e-06
Layer2	-8.1088	87.263	2.92e-12
Layer3	-5.5527	103.58	2.185e-07

Root_Length between 2020 and 2021

	t	df	p-value
Layer1	-3.4254	101.85	0.0008863
Layer2	-6.8939	105.16	4.165e-10
Layer3	-3.2276	104.78	0.001666

Appendix 4: Effect of Different Treatments on Relative Chlorophyll Content (SPAD) and Chlorophyll Fluorescence (Fv/Fm)

ANOVA of SPAD between Treatment in each measurement date in 2020

		Df	SumSq	MeanSq	Fvalue	Pr(>F)
08/07/2020	Treatment	2	72.2	36.11	2.846	6.85E-02
	Residuals	45	570.8	12.68		
15/07/2020	Treatment	2	96.9	48.45	5.429	7.73E-03
	Residuals	45	401.6	8.92		
21/07/2020	Treatment	2	175.5	87.73	6.729	2.77E-03
	Residuals	45	586.7	13.04		
29/07/2020	Treatment	2	94.7	47.33	2.789	7.21E-02
	Residuals	45	763.6	16.97		
06/08/2020	Treatment	2	18.4	9.181	0.672	5.16E-01
	Residuals	45	615.1	13.669		
12/08/2020	Treatment	2	72.3	36.17	2.199	1.23E-01
	Residuals	45	740.1	16.45		

ANOVA of SPAD between Treatment in each measurement date in 2021

		Df	SumSq	MeanSq	Fvalue	Pr(>F)
24/06/2021	Treatment	2	407.4	203.68	9.209	3.42E-04
	Residuals	57	1260.7	22.12		
30/06/2021	Treatment	2	89.5	44.74	1.788	1.76E-01
	Residuals	57	1426	25.02		
07/07/2021	Treatment	2	775.9	388	24.53	2.06E-08
	Residuals	57	901.4	15.8		

14/07/2021	Treatment	2	1736	868.1	46.63	1.01E-12
	Residuals	57	1061	18.6		

ANOVA of chlorophyll fluorescence between Treatment in each measurement date in 2020

		Df	SumSq	MeanSq	Fvalue	Pr(>F)
08/07/2020	Treatment	2	0.001474	0.000737	2.157	1.58E-01
	Residuals	12	0.004101	0.000341		
15/07/2020	Treatment	2	0.002967	0.001484	1.472	1.58E-01
	Residuals	12	0.012094	0.001008		
21/07/2020	Treatment	2	0.002713	0.001356	1.17	3.43E-01
	Residuals	12	0.013908	0.001159		
29/07/2020	Treatment	2	0.006564	0.003282	6.544	1.20E-02
	Residuals	12	0.006018	0.000502		
06/08/2020	Treatment	2	0.003868	0.0019339	2.937	9.15E-02
	Residuals	12	0.0079	0.0006584		
12/08/2020	Treatment	2	0.007645	0.003823	3.446	6.57E-02
	Residuals	12	0.01331	0.001109		

ANOVA of chlorophyll fluorescence between Treatment in each measurement date in 2021

		Df	SumSq	MeanSq	Fvalue	Pr(>F)
24/06/2021	Treatment	2	0.0456	0.02278	0.829	4.42E-01
	Residuals	57	1.5656	0.0274		
30/06/2021	Treatment	2	0.01099	0.005495	4.758	1.23E-02
	Residuals	57	0.06583	0.001155		
07/07/2021	Treatment	2	0.0725	0.03624	3.4	4.03E-02
	Residuals	57	0.6076	0.01066		
14/07/2021	Treatment	2	0.00914	0.004569	2.133	1.28E-01
	Residuals	57	0.1221	0.002142		

List of Figures