Review
Data Science and Plant Metabolomics

Anna Kisiel [1,2], Adrianna Krzemińska [2], Danuta Cembrowska-Lech [2,3] and Tymoteusz Miller [1,2,*]

1 Institute of Marine and Environmental Sciences, University of Szczecin, Wąska 13, 71-415 Szczecin, Poland
2 Polish Society of Bioinformatics and Data Science BIODATA, Popiełuszki 4c, 71-214 Szczecin, Poland
3 Department of Physiology and Biochemistry, Institute of Biology, University of Szczecin, Felczaka 3c, 71-412 Szczecin, Poland
* Correspondence: tymoteuszmiller@ptbd-biodata.org or tymoteusz.miller@usz.edu.pl

Abstract: The study of plant metabolism is one of the most complex tasks, mainly due to the huge amount and structural diversity of metabolites, as well as the fact that they react to changes in the environment and ultimately influence each other. Metabolic profiling is most often carried out using tools that include mass spectrometry (MS), which is one of the most powerful analytical methods. All this means that even when analyzing a single sample, we can obtain thousands of data. Data science has the potential to revolutionize our understanding of plant metabolism. This review demonstrates that machine learning, network analysis, and statistical modeling are some techniques being used to analyze large quantities of complex data that provide insights into plant development, growth, and how they interact with their environment. These findings could be key to improving crop yields, developing new forms of plant biotechnology, and understanding the relationship between plants and microbes. It is also necessary to consider the constraints that come with data science such as quality and availability of data, model complexity, and the need for deep knowledge of the subject in order to achieve reliable outcomes.

Keywords: plant metabolites; data science; machine learning

1. Introduction

Plant metabolism encompasses the chemical reactions and processes happening in a plant's cells to support life, growth, and reproduction. These encompass the transformation of energy and substances, such as water, carbon dioxide, and minerals, into parts of plants, such as starch and other storage products, cellulose, sugars, and various metabolites such as essential oils and allelochemicals [1]. This complex network contains photosynthesis, respiration, transpiration, and biosynthetic pathways which generate required metabolic intermediates, structural components, and various secondary metabolites [2]. The organic compounds produced in this way are usually divided by perspective function into primary metabolites, secondary metabolites (also called specialized metabolites or natural products), and plant hormones [3]. Primary metabolism products derived from glycolysis, the tricarboxylic acid (TCA) cycle, or the shikimate pathway often serve as precursors for the synthesis of the tens of thousands of secondary metabolites that have already been described [4]. Primary metabolites are highly conserved and directly required for plant growth and development [5] and secondary metabolites, including major groups such as phenolic compounds, terpenes, and nitrogen-containing compounds, are often lineage-specific and help plants interact with the biotic and abiotic environment [6].

Photosynthesis captures light energy and converts it into chemical energy in the form of adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide phosphate, (NADPH), which are used to create sugars and starches. The process of respiration, by contrast, releases energy to power cellular processes through the conversion of organic compounds into CO_2 and water while producing ATP [7].

In addition to basic metabolic pathways, plants also produce a wide range of secondary metabolites including allelochemicals, pigments, and essential oils. These compounds

have vital functions in the defense system, signaling, communication, and environmental adaptation of the plant, and are made by complicated biosynthetic pathways which are regulated by multiple environmental and genetic elements [8].

However, it should be remembered that the boundary between primary and secondary metabolism is uncertain, e.g., because many primary metabolism intermediates play similar roles in secondary metabolism. Secondary metabolites were previously thought to only mediate plant-environment interactions, but recent genetic and chemical studies show they can also regulate plant growth and defense, blurring the boundaries between these groups. Combining the roles these compounds play in the plant provides a close link between primary and secondary metabolism, and the distinctions between these processes must be made with increasing caution. It may be necessary to revisit the existing functional division [9–11]. Viewing secondary metabolites as integrated components of metabolic networks shaped by environmental selection pressures can improve our understanding of plant metabolism and plant-environment interactions.

In addition, it is important to remember that plant metabolism is regulated by a complex system of enzymes and pathways that regulates plant metabolism, which is influenced by genetic and environmental conditions such as temperature, light, and nutrient availability [12]. Unraveling plant metabolism is critical for a variety of purposes, from establishing sustainable agronomic practices to increasing crop productivity and resilience to environmental extremes, to discovering new products and remedies derived from plants [2]. Overall, plant metabolism studies give us invaluable insight on how plants maintain their existence, how they adjust to environmental changes, and how these revelations can guide the creation of innovative approaches to boost productivity and sustainability with regards to vegetation-based products and medications [13,14].

Data science requires taking multiple steps, such as collecting and organizing data, exploring and visualizing it, engineering features, and constructing and validating models, to finally deploy and supervise their outcomes. For this purpose, data scientists rely on many tools and strategies such as machine learning algorithms, statistical models, data mining, and big data technologies, in addition to data visualization tools [15].

Data science has an extensive range of applications in many industries, such as finance, retail, healthcare, transportation, and the environment. It is employed to manage difficult issues, such as foreshadowing patron conduct, discovering frauds and irregularities, optimizing logistics networks, and rearing public health [16,17]. The field of data science is constantly progressing, necessitating an aptitude in mathematics and computation as well as a thorough awareness of the principles involved in data analysis and modeling [18]. All the while, new technologies, algorithms, and strategies are being devised for uses concerning all manners of raw information [19]. Domain knowledge holds immense importance in data science. This type of knowledge refers to expertise and familiarity with a particular field or industry, which is key when comprehending the intentions and aims of a data science endeavor. Furthermore, strong domain knowledge aids in interpreting and disseminating the outcomes of data science projects, as well as detecting any potential biases or constraints. Ultimately, domain knowledge is vital for enabling successful, impactful data science results [20].

The purpose of this study is to present the possibility of using different data science methods and techniques such as machine learning, network analysis, and statistical modeling to evaluate data from plant metabolism studies. As we know, plant metabolism is one of the most difficult areas of plant research due to the large number of metabolic pathways, their mutual interactions, and dependence not only on the genotype but also on the environment. Any tool that will facilitate the assessment of the huge amount of data obtained during the analysis of metabolism is worth the interest of scientists. The possibilities of using artificial intelligence in the study of plant metabolism, as well as understanding the interaction between plant metabolism and the environment, have not been sufficiently understood. The use of artificial intelligence should provide the ability to predict the impact of environmental factors on plant metabolism and optimize plant

breeding programs. Based on the collected literature, plant metabolites were characterized in the context of their functions in the biology of plant systems and application possibilities, and an outline of methods used for plant metabolic profiling was provided. The following sections discuss the possibilities of using data science methods for mathematical modeling and explain the software tools available for simulation purposes. We then review the possibilities that recent discoveries in data science have opened up.

2. Plant Metabolites

2.1. Characteristics of Plant Metabolites and Its Applications

Primary metabolism is responsible for the production of appropriate compounds necessary for the survival of the plant, referred to as primary metabolites. As a result of reactions involving many enzymes, a wide range of molecules from the category of carbohydrates, amino acids, fatty acids, nucleic acids, and polymers derived from them (polysaccharides, proteins, lipids, etc.) are synthesized and used. Importantly, primary metabolites are identical in all living plant cells and are responsible for basic life functions such as respiration, growth, cell division, and reproduction [5]. Plant secondary metabolites, on the other hand, are formed from primary metabolites under the influence of various environmental stresses, such as light, temperature, and various metals, through several metabolic pathways. The formation of secondary metabolites is very specific to each family of plants, which from the same primary metabolites produce a large number of different secondary metabolites with different functions. They are mainly responsible for the interaction of the plant with the environment, hence their role in plant defense against biotic (viruses, bacteria, fungi, and insects) and abiotic (metals, temperature, light) stresses [21,22]. Among plant secondary metabolites, we can distinguish several basic sections based on chemical structure and functional groups. They include major groups such as terpenes, phenolic, polysaccharides, hydrocarbons, nitrogen-containing compounds, and sulfur-containing compounds [23] (Table 1).

Table 1. Secondary metabolites in plants.

Classification	Types	Examples	Function in Plant	Ref.
Terpenes	Monoterpenes	geraniol, limonene, carvone, linalool, linalyl acetate, camphora	attract pollinating insects, deterring pests, antifungal and antibacterial activity, plant communication	[19–22]
	Sesquiterpenes	humulene, farnesol, bisabolol, caryophyllene, helenalin	plant communication, antibacterial, antifungal and antiprotozoal activity, healing	[19,23,24]
	Diterpenes	cafestol, placytaksol, ginogolide, taxane, aconane	plant growth and development, defense from pathogens	[19,25]
	Sesterterpenes	geranylfarneso, ophiobolin A, genepolide, gentianelloid A	defense fromst pathogens	[26]
	Triterpenes	squalene, cucurbitacin, oleane, ursolic acid, chamaecydin	signaling molecules	[27,28]
	Polyterpenes	gutta-percha	defense from herbivores	[29,30]

Table 1. *Cont.*

Classification	Types	Examples	Function in Plant	Ref.
Phenolic	Simple phenolic	phenol, gallic acid, salicylic, ferulic and caffeic acids, hydroquinone	antimicrobial	[31]
	Coumarin	hydroxycoumarins, umbelliferone, esculetin, scopoletin	defense from insects, antifungical activity, role in plant nutrition, response in Fe stress	[19,30]
	Furanocoumarins	psoralin, angelicin, bergapten, methoxsalen	defense mechanism against mammals and insects, antifungal activity	[32]
	Lignin	resveratrol, wikstromal, matairesinol, dibenzyl, butyrolactol	antimicrobial, antifungal and cytotoxic effects	[19,33]
	Flavonoids	quercetins, luteolin, apigenin, peonidin, delphinidin	UV protection, pigmentation, antimicrobial defense, antioxidant activity, signal transduction, allelopathy, defense against herbivory, regulating gene expression, mediating symbiotic interactions	[18]
	Isoflavonoids	genistein, daidzein, ekwol, doumestrol, pueraryne	defense mechanism, mainly against fungi	[34]
	Tannins	tannic acid, geranilin, tellmagrandin 1 and 2	plant defense mechanisms against herbivorous, mammals, and insects	[19,35]
N containing compounds	Alkaloids	cocaine, nicotine, morphine, strychnine, codeine	role in germination, plants defense from predators	[18,34,36]
	Cyanogenic glucosides	linamarin, dhurrin, amygdalin, frunasin	ward off herbivores and pathogens	[37]
	Non-protein amino acids	L-Mimosine, L-Canavanine, 5-Hydroxy-L-tryptophan, L-3,4-Dihydroxyphenylolanine	interactions with bacteria, fungi, herbivores and other plants	[38]
S containing compounds		glutathione, glucosinolates, phytoalexins, thionins, defensins, allinim	physiological od abiotic stress, antibacterial and antifungal activity	[39,40]
Polysaccharides		pectin, celulose, inuline, alginian, starch	antibacterial and antifungal activity, plant cell walls components and starch components	[41,42]
Hydrocarbons		ethylene, march gas, methane	plant hormone—plant development	[34]

Terpenes are a large, diverse group of plant secondary metabolites. Among them are such important substances as insect attractants, essential oils, growth inhibitors, and plant hormones such as gibberellic acid and abscisic acid. All terpenes have nascent five-carbon isoprene units [24]. They are classified according to the number of isoprene units in the molecule and the prefix in the name indicates the number of terpene units. Monoterpenes are allelochemicals present in the essential oil in plants. They can be found in such plant organs as fruits, leaves, bark, or stems of herbaceous plants [25]. These substances are therefore responsible for the appropriate smell of plants, which in this way attract pollinating insects and for deterring pests with it [26]. Some monoterpenes have antifungal and antibacterial properties. Many of them have an intense taste and smell [27]. They also have a special role in plant communication as infochemicals, enabling the propagation of defense signals between plants [28]. Camphor and menthol are used as anti-irritants, analgesics, and antipruritics; others have a coronary vasodilator effect [43]. Sesquiterpenes

are compounds that are also attractive for pollinators and substances involved in plant communication using volatiles [29]. They exhibit antibacterial, antifungal, and antiprotozoal effects. They are also responsible for the healing properties of some plants, e.g., *Atracylodis macrocephala* [25,30]. In addition to therapeutic effects, sesquiterpenes may also have toxic and allergenic effects for humans and animals, which indicates their defensive function in the plant organism [30]. Diterpenes are very important compounds in herbal medicine. They exhibit a number of pharmacological properties: analgesic, antibacterial, diaphoretic, anti-inflammatory, and many others [25,44]. They are also involved in the basic elements of plant life, such as plant growth and development, and in the defense mechanism against pathogens [45]. Sesterterpens are a group of terpenes not yet fully explored, but researchers find these compounds in various parts of plants. It is assumed that they play an important role in the defense mechanism of plants against pathogens [32]. Some show cytotoxic activity in leukemic cells [46]. Triterpenes are compounds that can be divided into many subgroups, including saponins. Saponins are characterized by numerous therapeutic applications, but they are also surface-active substances that, in contact with water, form a foamy solution [33]. Simple triterpenes are components of surface waxes and specialized membranes and have the potential to act as signaling molecules [34]. Triterpene compounds are steroid precursors in both animals and plants and have anti-inflammatory and antirheumatic effects [47]. Polyterpenes are complex chemical compounds that occur in plants in the form of so-called resins and rubber. Rubber is a widely used material, an important element of the economy [35]. Polyterpene resins are used as a binder for various types of adhesives and in the production of paints [36].

Currently, it is believed that phenols are the largest group of secondary metabolites. They include both simple compounds with single aromatic rings and complex compounds such as tannins or lignins, which are polymers [24], and they share the presence of one or more phenol groups. These are compounds that perform many important functions in plants: among others, they are responsible for the color, taste, and smell of many plants. Phenolics are valued substances used in herbal medicine due to their anti-inflammatory effect [25]. Phenolic acids are ubiquitous in plants and the member gallic acid is well known for its astringent properties, although it has many other properties, including antiviral, antibacterial, antifungal, anti-inflammatory, and anticancer properties. Salicylates have anti-inflammatory properties [31] and phenol was used as the first antiseptic because it has antimicrobial activity [48]. Coumarins are chemical compounds that are very common in many plant species. Coumarin has been found in about 150 species belonging to more than 30 different families. They exhibit anti-inflammatory, antithrombotic, and anticancer properties, which makes them important substances in herbal medicine [25]. They are of great importance in plant nutrition. They are also responsible for the sweet smell and taste of plant organs, which is supposed to protect the plant from being eaten [49]. Furocoumarins act as a defense mechanism against mammals and insects. They also exhibit antifungal activity. These compounds can be found in roots, fruits, and leaves, often as components of essential oils. They are also responsible for the phytotoxic properties of plants such as *Datura* sp. or *Ruta* sp. [37]. Lignins are chemical compounds that are the main component of the cell walls of plant cells [38]. They are dimeric compounds found in many different species of plants. Many of the lignins have antimicrobial, antifungal, and cytotoxic effects such as wikstromal, matairesinol [25,50]. Resveratrol, on the other hand, has estrogen-like effects [51]. Flavonoids are a large group of compounds with considerable structural diversity. More than 2000 flavonoids are already known. The most common of these are anthocyanins, flavones, and flavonols. They are characterized by antioxidant, anti-inflammatory, antiallergic, antiviral, and anticancer effects. Therefore, they are of great importance in herbal medicine, and the research and use of these substances in dietetics and natural medicine are becoming more and more numerous. In plants, they are responsible for the colors of flowers and fruits [24,52,53]. Isoflavonoids are compounds that are very similar to flavonoids in their use as antioxidants, but these substances can also be classified as phytoestrogens that have the ability to bind to estrogen receptors. Phytoestrogenic

activity excerpt as genistein and daidzein. This action can have a good effect on the body, but research on all aspects of their function is still in progress. In plants, they play a role in the defense mechanism, mainly against fungi [39,54]. Tannins are polyphenols that are very common in the plant kingdom. These compounds have been used for centuries to transform animal skins because they have the ability to precipitate protein [55]. Drugs containing tannins have an antidiarrheal effect and have been used as antidotes in poisoning with heavy metals and alkaloids, and as an antiseptic [25]. In plants, tannins can be found in the leaves, bark, or wood itself. They are closely related to plant defense mechanisms against herbivorous mammals and insects [40].

Nitrogen-containing compounds are substances having one or more nitrogen atoms, usually in a heterocyclic ring. These compounds are easily soluble in water, optically active and have a significant effect on animals [24]. Among these substances are well-known alkaloids such as caffeine, nicotine, cocaine, and morphine [41,56]. It is estimated that 50% of drugs and pharmaceuticals of plant origin are alkaloids [42]. In the plant they have a role in germination and of course protecting plants from predators. Alkaloids have a very pronounced effect on animals, including humans. Many of the alkaloids act on the nervous system, because highly addictive opioids are alkaloids. In addition to the negative effect, they are used in herbal medicine as analgesics [39]. Cyanogenic glucosides are very common substances in plants. Their function in plants depends on activation by β-glucosidases to release toxic volatile hydrogen cyanide (HCN) as well as ketones and aldehydes to ward off herbivores and pathogens [44]. Non-protein amino acids are structures used by plants in their interactions with bacteria, fungi, herbivores, and other plants. They are found in the plant flower nectar and rhizosphere. They are also classified as allelochemicals [57].

Sulfur-containing compounds are organic substances that, even in small amounts, promote, inhibit, or modify physiological or abiotic stress in plants [58]. They also exhibit antibacterial and antifungal activity, which indicates their role in the defense mechanism of plants against pathogens [59].

Polysaccharides are widely distributed in plant organs such as roots, leaves, shoots, and seeds with anticancer, antioxidant, hypoglycemic, antibacterial, and antiviral effects [48]. In plant organisms, they occur in the composition of cell walls as cellulose or pectin and as reserve substances in the form of starch or inulin [60].

Hydrocarbons are very simple chemical compounds, made of hydrogen and carbon. They exist as simple chains or rings and form the basic backbone of more complex molecules. The waxes that build the coating on leaves and fruits contain many unsaturated hydrocarbons that are insoluble in water. They prevent water from sticking to the surface of the leaves. Hydrocarbons are also found in olive oil. An important hydrocarbon in plant development is ethylene, which plays the role of a plant hormone. It causes the fruit to ripen, the leaves to drop, and the neighboring flowers to wilt [39].

Primary and secondary plant metabolites are of great economic importance. They have some common features, they can usually be extracted from plant material by steam distillation, organic, or aqueous solvents, and are low molecular weight (>2000 Da) compounds with the exception of i.a. starch, gums, pectins, and natural rubber biopolymers, condensed tannins [61,62]. Plant metabolites are used in many industries, including pharmacology and medicine [63–65], agriculture [66,67], food industries [68–70], and other industries, including textiles and cosmetics [71,72].

2.2. Methods of Testing Plant Metabolites

Metabolomics is the study of the composition of the pool of metabolites (metabolic profiling) present in every organism, including plants. Thanks to metabolomics, it is possible to understand of phenotypic expression of plants as well as study changes and the regulation of plant metabolism in order to understand their adaptive and defensive responses to environmental stress [73,74]. Metabolomics has been divided into two distinct approaches, untargeted (which is a less specific analysis of all measurable analytes in the sample) and

targeted metabolomics (specific and sensitive analysis of defined and biochemically annotated metabolites). The quantity and complexity of metabolites and their characteristics make metabolomic studies extremely complex. It is necessary to use methodology and instruments to comprehensively identify and measure each metabolite [75,76].

The analysis of metabolites (primary or secondary) starts with sample preparation, which includes the extraction of metabolites by various methods. Among the extraction methods promising in metabolomic analysis are the methods of quenching, and mechanical and ultrasonic extraction, sometimes integrated [77,78]. The selection of solvents is also of key importance, of which chloroform, methanol, and water are most often mentioned [79–81]; it is necessary to extract and enrich the sample with interesting metabolites and to remove impurities such as proteins and salts that hinder the analysis. Extraction is performed using various methods, selecting the proportions of organic solvents, and also based on liquid-liquid extraction or solid phase extraction [82].

Many tools and techniques are used in metabolomics, and usually a combination of them. Mass spectrometry (MS) is one of most powerful and commonly used analytical methods in metabolomics, allowing a choice of sensitivity and resolution performance using either single (MS) or tandem (MS/MS) mass analyzers. A variety of MS-based techniques are now available for untargeted and targeted metabolic profiling using LC (liquid chromatography)-MS, GC (gas chromatography)-MS, CE (capillary electrophoresis)-MS, FTICR (Fourier transform ion cyclotron resonance)-MS, MALDI (matrix-assisted laser desorption/ionization), IMS (ion mobility spectrometry) and NMR (nuclear magnetic resonance). GC-MS achieves a higher separation of metabolites than LC-MS and avoids ion suppression by taking advantage of the gaseous phase and the nature of its MS ionization. However, otherwise for LC-MS, GC-MS requires chemical derivatization of the metabolic prior to the analysis. In turn ion mobility mass spectrometry is a gas-phase ion separation technique, which takes advantage of differences in the mobilities of ions by size, shape, charge, and the interaction with the inert gas under the influence of an electric field. Mass spectrometry with CE is a very good technique for separating polar ionic and charged substances that are separated based on their charge and size ratio in an aqueous medium. It enables significant efficiency in the analysis of metabolites in biological samples, especially for compounds with high polarity and water solubility. Furthermore, CE-MS is fast, uses small amounts of sample and solvents per analysis, and requires little time for sample preparation in comparison to GC-MS.

As mentioned previously, the main goal of metabolome profiling is the analysis of small molecules. Beyond mass spectrometry, nuclear magnetic resonance (NMR) is a good analytical platform used to analyze small molecules in metabolomics. NMR uses molecules of nuclear spin energy in the presence of a magnetic field. Moreover, NMR is very fast and non-destructive, requires little time for sample preparation, and provides highly repeatable results. Another spectroscopy method that allows for metabolite analysis is FTICR which, with MS, provides measures of metabolites in a couple min with minimal pre-detector separation and without ion dissociation. MALDI-MS is one of the mass spectrometric ionization techniques often chosen for the analysis of large biomolecules, especially proteins. However, it has been recognized as a potentially high-throughput method for the metabolome profiling [76,83–86].

Metabolomic studies can detect even thousands of possible specialized metabolites, which leads to the creation of large data sets. That is why it is a huge challenge to extract information about specialized metabolites from the huge amount of data generated during analyzes. This requires transforming the raw data into a numerical matrix and then applying statistical methods that will facilitate the comparison of all results across all samples. Several programs are available for the in silico analysis of the large amount of metabolite spectral data generated by various analytical instruments. They are often proposed by manufacturers of apparatus for metabolomic analyses. Many bioinformatics tools and spectral libraries are available for data pre-processing, including XCMS, METLIN, PRIMe, AMDIS, MetaboAnalyst, MetAlign, and others [76,85,87–91].

3. Does Data Science Can Help in Studying Plant Metabolites?

3.1. Data Science Methods

Data science methods are an essential part of addressing the challenges posed by dangerous plant metabolites and environmental issues. These methods involve using statistical, computational, and mathematical techniques to analyze large and complex datasets, enabling the identification of patterns, relationships, and trends that may not be immediately apparent from the raw data [92].

In order to elaborate on the subject of plant metabolites, it should be mentioned that a summary of statistical methods and dedicated software has already been published [76,93].

In a recent review article by Johnson et al. [93], various statistical methods and software tools commonly used in plant metabolomics research were summarized. The authors highlighted the importance of the preprocessing and normalization of raw data, as well as the selection of appropriate statistical tests for analyzing metabolite abundance changes between different samples or treatments.

Among the statistical methods and software packages discussed by the authors were multivariate analysis tools such as principal component analysis (PCA), partial least squares (PLS), and orthogonal projections to latent structures (OPLS). These techniques can help identify patterns in metabolite data that may be associated with specific biological factors, such as treatment conditions or genetic variation.

Piasecka [76] provided a comprehensive review of various analytical methods that can be used to detect changes in plant metabolomics in response to biotic and abiotic stresses. The author highlighted the importance of integrating different types of data, such as transcriptomic and proteomic data, with metabolite data to obtain a more comprehensive understanding of plant stress responses.

In terms of statistical methods and software, Piasecka [76] discussed various multivariate analysis techniques, such as PCA and hierarchical cluster analysis (HCA), which can be used to identify groups of metabolites that are strongly correlated and may be associated with specific stress conditions.

The methods used in data science include machine learning algorithms such as Random Forest, Gradient Boosting, Support Vector Machines, and Neural Networks [94]. Remote sensing technologies like hyperspectral imaging, thermal imaging, Light Detection and Ranging (LiDAR), and satellite imagery are used to analyze the environment [95,96] while predictive modeling techniques such as species distribution models [97], generalized linear models [98], time series models [99], and projection models [100] are used to forecast future trends.

Spatial analysis is another important method used in data science, which involves using tools such as geographic information system (GIS), geostatistics [101], Kriging [102], and spatial clustering [103] to analyze spatial data. Data visualization techniques like heat maps [104], choropleth maps [105], scatter plots [106], and time series [107] plots are used to present the results of data analysis in a more accessible and understandable way.

Data preprocessing and cleaning methods like data imputation [108], outlier detection [109], feature selection [110], and normalization [111] are used to ensure the accuracy and completeness of data. Molecular biology techniques like dPCR [112], qPCR [113], and DNA sequencing [114,115] are also used to analyze genetic and molecular data related to plant metabolites and their impact on the environment.

Climate modeling techniques, including Global Circulation Models [116], Earth System Models [117], and Regional Climate Models [118], are used to study the impact of climate change on the environment. Environmental impact assessments such as Life Cycle Assessment [119], Environmental Impact Assessment [120], and Ecological Footprint Analysis [121] are used to assess the impact of human activities on the environment and develop strategies to mitigate those impacts.

The methods are called "data science methods" because they are part of the broader field of data science, which involves analyzing data using statistics, computational methods,

and mathematics. By analyzing and processing large, complex datasets, these methods can reveal patterns, relationships, and trends that are not readily apparent in the raw data [122].

These data science methods are used to analyze large and diverse datasets, such as satellite imagery [123], climate data [124], and molecular biology [125] data, in order to gain a better understanding of how harmful plant metabolites impact the environment and to develop targeted management strategies in the context of addressing dangerous plant metabolites and environmental issues [126]. By using these methods, scientists and researchers can turn large and complex datasets into actionable information that can inform decision-making and help to mitigate the impacts of dangerous plant metabolites on the environment.

Data science methods offer several ways to study plant metabolites and their impact on the environment. Predictive modeling can be used to predict the growth and development of different plant species and the production of various metabolites, such as allelochemicals and essential oils [127]. Metabolic pathway analysis can be performed using transcriptomics, proteomics, and metabolomics data to understand the biosynthesis and regulation of different plant metabolites. Gene expression analysis can also be used to study the regulation of metabolic pathways and identify the genes responsible for the production of specific metabolites [128,129]. Some plants contain cyanide or other poisons [130–147]. Machine learning-based classification algorithms can be used to classify cells of different plant species based on their metabolic profiles and predict the potential production of harmful metabolites, such as allelochemicals and cyanide. Data visualization tools can be used to visualize and compare the metabolic profiles of different plant species and identify trends and patterns in the data. Network analysis can be used to study the relationships between different metabolites and the enzymes and pathways involved in their biosynthesis and degradation [92,148,149].

Environmental monitoring data, such as satellite imagery and climate data, can be used to study the impact of environmental factors such as temperature and precipitation on plant metabolism and the production of specific metabolites. By applying these data science methods, researchers can gain a deeper understanding of plant metabolism and the impact of specific metabolites on the environment. This knowledge can inform the development of sustainable agriculture and land use practices, helping to mitigate the negative impact of plant metabolites on the environment [150].

The role of AI in the context of plant metabolism and the classification of plant metabolites is still in its early stages. However, there are some potential applications of AI in this field. For example, AI algorithms can be used to analyze large-scale metabolomic datasets, which can help to identify novel secondary metabolites and their functions. Additionally, machine learning algorithms can be used to classify and predict the functions of different metabolites based on their structural properties and other features. This can help to refine our understanding of the roles of different plant metabolites in plant growth, defense, and interactions with the environment.

Moreover, AI can also help in predicting and understanding the impact of environmental factors on plant metabolism. For example, AI can be used to model the impact of climate change on the production of secondary metabolites in plants, which can help us to predict and mitigate the potential effects of climate change on plant communities and ecosystems. Additionally, AI can also assist in designing and optimizing plant breeding programs for developing crop varieties with specific metabolite profiles that confer desirable traits, such as resistance to pests or diseases. Overall, AI has the potential to significantly contribute to the advancement of our understanding of plant metabolism and the classification of plant metabolites [11].

3.2. Data Science Techniques

In addition to studying plant metabolism, there are several other data science approaches that can be employed. These include techniques such as statistical analysis and machine learning algorithms, which can be applied to better understand how plants function.

1. Clustering analysis:

    Clustering algorithms can be used to group plants based on their metabolic profiles and identify metabolic similarities and differences between species. For example, clustering analysis can be used to group plant species based on their production of specific allelochemicals, such as terpenes, and compare the metabolic profiles of invasive and native plant species [151].

2. Dimension reduction techniques:

    Techniques such as principal component analysis (PCA) and multidimensional scaling (MDS) can be used to reduce the complexity of large datasets and identify the most important metabolic pathways and metabolites. For example, PCA can be used to identify the most important metabolic pathways responsible for the production of volatile organic compounds (VOCs) in plants [152].

3. Artificial Neural Networks (ANNs):

    ANNs can be used to model complex relationships between environmental factors, such as temperature and light, and the production of specific metabolites, such as essential oils. For example, ANNs can be used to predict the production of essential oils in different plant species based on environmental variables, such as temperature, light, and precipitation [153].

4. Decision Tree analysis:

    Decision Tree analysis can be used to identify the most important environmental and genetic factors that influence the production of specific metabolites. For example, Decision Tree analysis can be used to identify plant species, which has implications for identifying the most important environmental and genetic factors that influence the production of allelochemicals in different plant species [154].

5. Bayesian networks:

    Bayesian networks can be used to model the relationships between different metabolites and the pathways involved in their biosynthesis and degradation. For example, Bayesian networks can be used to model the relationships between different metabolites in the biosynthesis of secondary metabolites, such as flavonoids, and the enzymes and pathways involved in their production [149] .

    Examples of data science in plant metabolomic scientific field are presented in Table 2.

Table 2. Examples of data science in plant metabolomic scientific field.

Used Methods	Method	Study
Evaluating the physiological and biochemical responses of melon plants to NaCl salinity stress using supervised and unsupervised statistical analysis	OPLS-DA	Use of OPS-DA and PCA to predict melon plant response to salinity [152].
Ionomic and metabolomic analyses reveal the resistance response mechanism to saline-alkali stress in *Malus halliana* seedlings	OPLS-DA	Use of OPLS-DA to determine the nature of metabolic changes in leaves of apple seedlings [128].
Predicting metabolic pathways of plant enzymes without using sequence similarity: models from machine learning	mApLe	Using mApLe to predict metabolic pathways of plant enzymes instead of Enzyme Commission (EC) numbers [155].
Salinity source alters mineral composition and metabolism of *Cichorium spinosum*	OPLS-DA	Use OPLS-DA for the visualization of the fluctuations in the plant's metabolome in response to the various treatments [156].
Physiological and metabolic responses triggered by omeprazole improve tomato plant tolerance to NaCl stress	OPLS-DA	Use OPLS-DA to separate the variability between the groups of samples [157].

Table 2. *Cont.*

Used Methods	Method	Study
Metabolic responses to potassium availability and waterlogging reshape respiration and carbon use efficiency in oil palm	OPLS	Use of OPLS to determine the significance of metabolome and proteome data components in the organs of the studied plants [158].
Comprehensive meta-analysis and machine learning approaches identified the role of novel drought specific genes in *Oryza sativa*	* SVM, kNN, NB, DT, RF	These machine learning techniques were used to identify the distinguishing features between test samples and controls based on accuracy [129,149].
Evaluating the physiological and biochemical responses of melon plants to NaCl salinity stress using supervised and unsupervised statistical analysis	PCA	Use of PCA to predict melon plant response to salinity [152].
	HCA	Use HCA to make a heat map to predict melon plant response of salinity [152].
Ionomic and metabolomic analyses reveal the resistance response mechanism to saline-alkali stress in *Malus halliana* seedlings	PCA	Use of PCA to predict variability in two groups of metabolites in leaf samples [128].
Principal component analysis of hormone profiling data suggests an important role for cytokinins in regulating leaf growth and senescence of salinized tomato	PCA	Using PCA as a mathematical tool to evaluate the relationship between physiological and hormonal variables in tomato research [159].
Salinity source alters mineral composition and metabolism of *Cichorium spinosum*	HCA	Use HCA to support OPLS-DA in the visualization of the fluctuations in the plant's metabolome in response to the various treatments [157].
Physiological and metabolic responses triggered by omeprazole improve tomato plant tolerance to NaCl stress	PCA	Use PCA for obtaining a broad overview of morphological and physiological changes in tomato plants in response to the use of omeprazole in salted and unsalted conditions [157].
Changes in carbohydrates triggered by low temperature waterlogging modify photosynthetic acclimation to cold in *Festuca pratensis*	PCA	Use of PCA to determine the variability between parameters and to highlight the most important ones from the research perspective [160].
Zinc stress affects ionome and metabolome in tea plants	PCA	Using PCA for tissue ionome variation [161].
	HCA	Using HCA to visualize correlations between elements and metabolites in tea leaves [161].

* SVM—support vector machines, kNN—*k*-nearest neighbors algorithm, NB—naive Bayes classifiers, DT—decision tree, RF—random forest.

4. The Advantages and Applications of Data Science in Plant Metabolism Studies

Data science has many potential benefits in the study of plant metabolism. One of the primary advantages is the ability to derive data-driven insights from complex biochemical processes and interactions that would be difficult to discern without the use of advanced analytical tools. Predictive modeling is another advantage of data science, enabling the development of models that can simulate and predict the behavior of plant metabolism under different conditions. These models can help identify new targets for intervention, optimize growth conditions, and improve crop yields [148,152].

Data science also enables the development of personalized solutions that take into account the unique biology and environment of each plant, leading to more targeted and effective interventions. Additionally, data science can integrate data from multiple sources and scales, providing a comprehensive understanding of plant metabolism, from molecular and cellular data to whole-plant and ecosystem data [162,163].

Sustainable solutions can also be developed using data science, including efficient irrigation and fertilizer application, which can help minimize costs and optimize resource

utilization while also improving crop yields. Finally, data science can support decision-making by generating evidence-based recommendations for interventions and management strategies [164].

The benefits of data science in plant metabolism studies demonstrate its potential to revolutionize our understanding of plant biology, inform the development of innovative solutions, and improve decision-making processes. By utilizing data science, we can develop sustainable practices, optimize resource utilization, and improve crop yields, leading to a more efficient and environmentally responsible agricultural industry [165,166].

5. Challenges and Considerations for Data Science in Plant Metabolic Studies

The benefits of data science methods offer numerous benefits in plant metabolic studies, but it is important to consider the challenges and possible problems that come with their use. One of the primary challenges is the quality and availability of data. The difficulty in collecting and processing data at different scales and levels of detail, or the complexity of the underlying biology, can limit data quality and availability, making it challenging to generate accurate models [167].

Model complexity is another consideration, particularly when modeling complex biological systems like plant metabolism. Complex models can be difficult to interpret and communicate to stakeholders [168]. Overfitting is also a potential issue, particularly when data is limited, or the model is too complex to capture the underlying biological processes [169].

A lack of domain knowledge can lead to inaccurate or inappropriate models or result in misinterpreted results. Data science requires a combination of technical and domain knowledge, and a lack of the latter can significantly impact the accuracy of models and analyses [170].

Computational resources can also be a challenge, particularly with large and complex data sets. Data science methods can be computationally intensive, requiring significant resources to process and analyze data. In plant metabolic studies, where data sets can be especially large and complex, this can be particularly challenging [171].

While data science methods have numerous benefits in plant metabolic studies, it is essential to consider the challenges and address them appropriately. By addressing these considerations, it is possible to improve the accuracy, interpretability, and impact of data science methods, leading to more effective and informed plant metabolic research.

6. Conclusions

In this review, we have explored the potential of integrating data science techniques with plant metabolism studies to enhance our understanding of plant biology and its interactions with the environment. Our main findings indicate that machine learning algorithms, network analysis, and statistical modeling can contribute to new insights into plant growth, development, and response to environmental changes. Moreover, the application of these methods can help address the impact of environmental factors on metabolite production and the environmental consequences of plant metabolites.

By discussing various data science methods, including clustering analysis, dimension reduction, artificial neural networks, decision tree analysis, and Bayesian networks, this review expands the current knowledge in the field by demonstrating the versatility of data science techniques in plant metabolism research. Furthermore, we have shown how AI can be used to identify novel secondary metabolites, predict and understand the impact of environmental factors on plant metabolism, and optimize plant breeding programs.

As we look towards the future, several prospects emerge for further research and development. The integration of novel data sources, such as remote sensing and high-throughput phenotyping, can provide additional layers of information to enhance plant metabolism studies. The continued development of advanced machine learning techniques, such as deep learning and reinforcement learning, can lead to the more accurate and efficient modeling of complex biological systems. Additionally, interdisciplinary collaboration

between plant scientists, data scientists, and other stakeholders will be crucial for addressing the challenges of data integration, model interpretability, and the ethical use of AI in plant research.

While some limitations and challenges persist, our review highlights the exciting potential of combining data science and plant metabolism studies. By fostering interdisciplinary collaboration, we can further advance plant biotechnology, sustainable agriculture, and our understanding of the complex interactions between plants and microbes. This synthesis of knowledge ultimately opens up new avenues for research with a promising future in addressing global food security and environmental sustainability.

References

1. Crozier, A.; Clifford, M.N.; Ashihara, H. *Plant Secoundary Metabolites: Occurence, Strucure and Role in the Human Diet*; John Wiley & Sons: New York, NY, USA, 2008.
2. Li, S.; Tian, Y.; Wu, K.; Ye, Y.; Yu, J.; Zhang, J.; Liu, Q.; Hu, M.; Li, H.; Tong, Y.; et al. Modulating Plant Growth–Metabolism Coordination for Sustainable Agriculture. *Nature* **2018**, *560*, 595–600. [CrossRef]
3. Taiz, L.; Zeiger, E.; Møller, I.; Murphy, A. *Plant Physiology and Development*; Sinauer Associates Incorporated: Sunderland, MA, USA, 2015.
4. Kroymann, J. Natural Diversity and Adaptation in Plant Secondary Metabolism. *Curr. Opin. Plant Biol.* **2011**, *14*, 246–251. [CrossRef] [PubMed]
5. Fernie, A.R.; Pichersky, E. Focus Issue on Metabolism: Metabolites, Metabolites Everywhere. *Plant Physiol.* **2015**, *169*, 1421–1423. [CrossRef]
6. Hartmann, T. From Waste Products to Ecochemicals: Fifty Years Research of Plant Secondary Metabolism. *Phytochemistry* **2007**, *68*, 2831–2846. [CrossRef]
7. Geigenberger, P. Response of Plant Metabolism to Too Little Oxygen. *Curr. Opin. Plant Biol.* **2003**, *6*, 247–256. [CrossRef] [PubMed]
8. Pagare, S.; Bhatia, M.; Tripathi, N.; Pagare, S.; Bansal, Y.K.; Pagare, S.; Kasote, D.M.; Acquadro, A.; Rubiolo, P.; Bicchi, C.; et al. Secondary Metabolites of Plants and Their Role: Overview. *Curr. Trends Biotechnol. Pharm.* **2015**, *9*, 293–304.
9. Neilson, E.H.; Goodger, J.Q.D.; Woodrow, I.E.; Møller, B.L. Plant Chemical Defense: At What Cost? *Trends Plant Sci.* **2013**, *18*, 250–258. [CrossRef]
10. Zhou, S.; Richter, A.; Jander, G. Beyond Defense: Multiple Functions of Benzoxazinoids in Maize Metabolism. *Plant Cell Physiol.* **2018**, *59*, 1528–1537. [CrossRef]
11. Erb, M.; Kliebenstein, D.J. Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy. *Plant Physiol.* **2020**, *184*, 39–52. [CrossRef]
12. Fernie, A.R.; Tohge, T. The Genetics of Plant Metabolism. *Annu. Rev. Genet.* **2017**, *51*, 287–310. [CrossRef]
13. Fang, C.; Fernie, A.R.; Luo, J. Exploring the Diversity of Plant Metabolism. *Trends Plant Sci.* **2019**, *24*, 83–98. [CrossRef] [PubMed]
14. Wink, M. Modes of Action of Herbal Medicines and Plant Secondary Metabolites. *Medicines* **2015**, *2*, 251–286. [CrossRef] [PubMed]
15. Agarwal, R.; Dhar, V. Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Inf. Syst. Res.* **2014**, *25*, 443–448. [CrossRef]
16. Chung, S.-H. Applications of Smart Technologies in Logistics and Transport: A Review. *Transp. Res. E Logist. Transp. Rev.* **2021**, *153*, 102455. [CrossRef]
17. Raghupathi, W.; Raghupathi, V. Big Data Analytics in Healthcare: Promise and Potential. *Health Inf. Sci. Syst.* **2014**, *2*, 3. [CrossRef] [PubMed]
18. Vogelsang, A.; Borg, M. Requirements engineering for machine learning: Perspectives from data scientists. In Proceedings of the 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW), Jeju, Republic of Korea, 23–27 September 2019; pp. 245–251.
19. Schoenherr, T.; Speier-Pero, C. Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential. *J. Bus. Logist.* **2015**, *36*, 120–132. [CrossRef]

20. Waller, M.A.; Fawcett, S.E. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *J. Bus. Logist.* **2013**, *34*, 77–84. [CrossRef]

21. Ramírez-Gómez, X.S.; Jiménez-García, S.N.; Beltrán Campos, V.; Lourdes García Campos, M. Plant metabolites in plant defense against pathogens. In *Plant Diseases—Current Threats and Management Trends*; IntechOpen: Rijeka, Croatia, 2020.

22. Anjitha, K.S.; Sameena, P.P.; Puthur, J.T. Functional Aspects of Plant Secondary Metabolites in Metal Stress Tolerance and Their Importance in Pharmacology. *Plant Stress* **2021**, *2*, 100038. [CrossRef]

23. Aharoni, A.; Galili, G. Metabolic Engineering of the Plant Primary–Secondary Metabolism Interface. *Curr. Opin. Biotechnol.* **2011**, *22*, 239–244. [CrossRef] [PubMed]

24. Teoh, E.S. Secondary Metabolites of Plants. In *Medicinal Orchids of Asia*; Springer International Publishing: Cham, Switzerland, 2016; pp. 59–73.

25. Hussein, R.A.; El-Anssary, A.A. Plants secondary metabolites: The key drivers of the pharmacological actions of medicinal plants. In *Herbal Medicine*; IntechOpen: Rijeka, Croatia, 2019.

26. Feng, L.; Chen, C.; Li, T.; Wang, M.; Tao, J.; Zhao, D.; Sheng, L. Flowery Odor Formation Revealed by Differential Expression of Monoterpene Biosynthetic Genes and Monoterpene Accumulation in Rose (*Rosa rugosa* Thunb.). *Plant Physiol. Biochem.* **2014**, *75*, 80–88. [CrossRef]

27. Banchio, E.; Bogino, P.C.; Santoro, M.; Torres, L.; Zygadlo, J.; Giordano, W. Systemic Induction of Monoterpene Biosynthesis in *Origanum* × *Majoricum* by Soil Bacteria. *J. Agric. Food Chem.* **2010**, *58*, 650–654. [CrossRef] [PubMed]

28. Riedlmeier, M.; Ghirardo, A.; Wenig, M.; Knappe, C.; Koch, K.; Georgii, E.; Dey, S.; Parker, J.E.; Schnitzler, J.-P.; Vlot, A.C. Monoterpenes Support Systemic Acquired Resistance within and between Plants. *Plant Cell* **2017**, *29*, 1440–1459. [CrossRef]

29. Algarra Alarcon, A.; Lazazzara, V.; Cappellin, L.; Bianchedi, P.L.; Schuhmacher, R.; Wohlfahrt, G.; Pertot, I.; Biasioli, F.; Perazzolli, M. Emission of Volatile Sesquiterpenes and Monoterpenes in Grapevine Genotypes Following *Plasmopara viticola* Inoculation In Vitro. *J. Mass Spectrom.* **2015**, *50*, 1013–1022. [CrossRef]

30. Padilla-Gonzalez, G.F.; dos Santos, F.A.; da Costa, F.B. Sesquiterpene Lactones: More Than Protective Plant Compounds with High Toxicity. *CRC Crit. Rev. Plant Sci.* **2016**, *35*, 18–37. [CrossRef]

31. Amann, R.; Peskar, B.A. Anti-Inflammatory Effects of Aspirin and Sodium Salicylate. *Eur. J. Pharmacol.* **2002**, *447*, 1–9. [CrossRef]

32. Chen, Q.; Li, J.; Ma, Y.; Yuan, W.; Zhang, P.; Wang, G. Occurrence and Biosynthesis of Plant Sesterterpenes (C25), a New Addition to Terpene Diversity. *Plant Commun.* **2021**, *2*, 100184. [CrossRef]

33. Faizal, A.; Geelen, D. Saponins and Their Role in Biological Processes in Plants. *Phytochem. Rev.* **2013**, *12*, 877–893. [CrossRef]

34. Thimmappa, R.; Geisler, K.; Louveau, T.; O'Maille, P.; Osbourn, A. Triterpene Biosynthesis in Plants. *Annu. Rev. Plant Biol.* **2014**, *65*, 225–257. [CrossRef] [PubMed]

35. Jan, R.; Asaf, S.; Numan, M.L.; Lubna, L.; Kim, K.-M. Plant Secondary Metabolite Biosynthesis and Transcriptional Regulation in Response to Biotic and Abiotic Stress Conditions. *Agronomy* **2021**, *11*, 968. [CrossRef]

36. Zhaobang, S. *Production and Standards for Chemical Non-Wood Forest Products in China*; CIFOR: Bogor, Indonesia, 1995.

37. Del Río, J.A.; Díaz, L.; García-Bernal, D.; Blanquer, M.; Ortuño, A.; Correal, E.; Moraleda, J.M. Furanocoumarins: Biomolecules of therapeutic interest. *Stud. Nat. Prod. Chem.* **2014**, *43*, 145–195.

38. Weng, J.; Chapple, C. The Origin and Evolution of Lignin Biosynthesis. *New Phytol.* **2010**, *187*, 273–285. [CrossRef]

39. Miadoková, E. Isoflavonoids—An Overview of Their Biological Activities and Potential Health Benefits. *Interdiscip. Toxicol.* **2009**, *2*, 211–218. [CrossRef]

40. Shahin, H.; Naser, M.-S.; Behrad, E.; Farad, B.M. Plants and Secondary Metabolites (Tannins): A Review. *Int. J. For. Soil Eros.* **2011**, *1*, 47–53.

41. Dewick, P.M. *Medicinal Natural Products: A Biosynthetic Approach*; John Wiley & Sons: New York, NY, USA, 2002.

42. Zhu, Z.H.; Chao, C.J.; Lu, X.Y.; Xiong, Y.G. *Paulownia in China: Cultivation and Utilization*; International Development Research Centre: Ottawa, ON, Canada, 1986.

43. Yi, Z.; Wang, Z.; Li, H.; Liu, M. Inhibitory Effect of Tellimagrandin I on Chemically Induced Differentiation of Human Leukemia K562 Cells. *Toxicol. Lett.* **2004**, *147*, 109–119. [CrossRef]

44. Bagci, E.; Yazgın, A.; Hayta, S.; Cakılcıoglu, U. Composition of the Essential Oil of *Teucrium chamaedrys* L. (Lamiaceae) from Turkey. *J. Med. Plants Res.* **2010**, *4*, 2588–2590.

45. Zerbe, P.; Bohlmann, J. Plant Diterpene Synthases: Exploring Modularity and Metabolic Diversity for Bioengineering. *Trends Biotechnol.* **2015**, *33*, 419–428. [CrossRef]

46. Ishikura, H.; Mochizuki, T.; Izumi, Y.; Usui, T.; Sawada, H.; Uchino, H. Differentiation of Mouse Leukemic M1 Cells Induced by Polyprenoids. *Leuk. Res.* **1984**, *8*, 843–852. [CrossRef] [PubMed]

47. Culioli, G. A Lupane Triterpene from Frankincense (*Boswellia* Sp., Burseraceae). *Phytochemistry* **2003**, *62*, 537–541. [CrossRef]

48. Pelczar, M.J.; Chan, E.C.S.; Krieg, N.R. Control of Microorganisms, the Control of Microorganisms by Physical Agents. *Microbiology* **1988**, *469*, 509.

49. Robe, K.; Izquierdo, E.; Vignols, F.; Rouached, H.; Dubos, C. The Coumarins: Secondary Metabolites Playing a Primary Role in Plant Nutrition and Health. *Trends Plant Sci.* **2021**, *26*, 248–259. [CrossRef] [PubMed]

50. Sharma, P.R.; Shanmugavel, M.; Saxena, A.K.; Qazi, G.N. Induction of Apoptosis by a Synergistic Lignan Composition from *Cedrus Deodara* in Human Cancer Cells. *Phytother. Res.* **2008**, *22*, 1587–1594. [CrossRef]

51. Gehm, B.D.; McAndrews, J.M.; Chien, P.-Y.; Jameson, J.L. Resveratrol, a Polyphenolic Compound Found in Grapes and Wine, Is an Agonist for the Estrogen Receptor. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 14138–14143. [CrossRef]
52. Montanher, A.B.; Zucolotto, S.M.; Schenkel, E.P.; Fröde, T.S. Evidence of Anti-Inflammatory Effects of *Passiflora edulis* in an Inflammation Model. *J. Ethnopharmacol.* **2007**, *109*, 281–288. [CrossRef]
53. Serafini, M.; Peluso, I.; Raguzzini, A. Flavonoids as Anti-Inflammatory Agents. *Proc. Nutr. Soc.* **2010**, *69*, 273–278. [CrossRef]
54. Goławska, S.; Sprawka, I.; Łukasik, I.; Goławski, A. Are Naringenin and Quercetin Useful Chemicals in Pest-Management Strategies? *J. Pest Sci.* **2014**, *87*, 173–180. [CrossRef]
55. Hagerman, A.E.; Butler, L.G. The Specificity of Proanthocyanidin-Protein Interactions. *J. Biol. Chem.* **1981**, *256*, 4494–4497. [CrossRef] [PubMed]
56. Benowitz, N.L. Farmacología de La Nicotina: Adicción, Enfermedad Inducida Por El Tabaquismo y Terapéutica. *Revisión Anu. De Farmacol. Y Toxicol.* **2009**, *49*, 57–71.
57. Nepi, M. Beyond Nectar Sweetness: The Hidden Ecological Role of Non-Protein Amino Acids in Nectar. *J. Ecol.* **2014**, *102*, 108–115. [CrossRef]
58. Khan, M.I.R.; Asgher, M.; Iqbal, N.; Khan, N.A. Potentiality of Sulphur-Containing Compounds in Salt Stress Tolerance. In *Ecophysiology and Responses of Plants under Salt Stress*; Springer: New York, NY, USA, 2013; pp. 443–472.
59. Kim, S.; Kubec, R.; Musah, R.A. Antibacterial and Antifungal Activity of Sulfur-Containing Compounds from *Petiveria alliacea* L. *J. Ethnopharmacol.* **2006**, *104*, 188–192. [CrossRef]
60. Liu, J.; Willför, S.; Xu, C. A Review of Bioactive Plant Polysaccharides: Biological Activities, Functionalization, and Biomedical Applications. *Bioact. Carbohydr. Diet. Fibre* **2015**, *5*, 31–61. [CrossRef]
61. Bourgaud, F.; Gravot, A.; Milesi, S.; Gontier, E. Production of Plant Secondary Metabolites: A Historical Perspective. *Plant Sci.* **2001**, *161*, 839–851. [CrossRef]
62. Danova, K.; Pistelli, L. Plant Tissue Culture and Secondary Metabolites Production. *Plants* **2022**, *11*, 3312. [CrossRef]
63. Twaij, B.M.; Hasan, M.N. Bioactive Secondary Metabolites from Plant Sources: Types, Synthesis, and Their Therapeutic Uses. *Int. J. Plant Biol.* **2022**, *13*, 4–14. [CrossRef]
64. Atanasov, A.G.; Waltenberger, B.; Pferschy-Wenzig, E.-M.; Linder, T.; Wawrosch, C.; Uhrin, P.; Temml, V.; Wang, L.; Schwaiger, S.; Heiss, E.H. Discovery and Resupply of Pharmacologically Active Plant-Derived Natural Products: A Review. *Biotechnol. Adv.* **2015**, *33*, 1582–1614. [CrossRef] [PubMed]
65. Ullrich, C.I.; Aloni, R.; Saeed, M.E.M.; Ullrich, W.; Efferth, T. Comparison between Tumors in Plants and Human Beings: Mechanisms of Tumor Development and Therapy with Secondary Plant Metabolites. *Phytomedicine* **2019**, *64*, 153081. [CrossRef] [PubMed]
66. Freeman, B.; Beattie, G. An Overview of Plant Defenses against Pathogens and Herbivores. *Plant Health Instr.* **2008**. [CrossRef]
67. Yue, Z.; Singh, V.; Argenta, J.; Segbefia, W.; Miller, A.; Ming Tseng, T. Use of Plant Secondary Metabolites to Reduce Crop Biotic and Abiotic Stresses: A Review. In *Secondary Metabolites—Trends and Reviews*; IntechOpen: Rijeka, Croatia, 2022.
68. Kallscheuer, N.; Classen, T.; Drepper, T.; Marienhagen, J. Production of Plant Metabolites with Applications in the Food Industry Using Engineered Microorganisms. *Curr. Opin. Biotechnol.* **2019**, *56*, 7–17. [CrossRef] [PubMed]
69. Heitefuss, R. Functions and Biotechnology of Plant Secondary Metabolites, 2nd edn, Annual Plant Reviews, Vol 39. *J. Phytopathol.* **2010**, *1*, 72. [CrossRef]
70. Zhou, S.; Ma, Y.; Shang, Y.; Qi, X.; Huang, S.; Li, J. Functional Diversity and Metabolic Engineering of Plant Specialized Metabolites. *Life Metab.* **2022**, loac019. [CrossRef]
71. Boeriu, C.G. *Plants4Cosmetics: Perspectives for Plant Ingredients in Cosmetics*; Wageningen UR-Food & Biobased Research: Wageningen, The Netherlands, 2015.
72. Brudzyńska, P.; Sionkowska, A.; Grisel, M. Plant-Derived Colorants for Food, Cosmetic and Textile Industries: A Review. *Materials* **2021**, *14*, 3484. [CrossRef]
73. Razzaq, A.; Sadia, B.; Raza, A.; Khalid Hameed, M.; Saleem, F. Metabolomics: A Way Forward for Crop Improvement. *Metabolites* **2019**, *9*, 303. [CrossRef] [PubMed]
74. Peters, K.; Worrich, A.; Weinhold, A.; Alka, O.; Balcke, G.; Birkemeyer, C.; Bruelheide, H.; Calf, O.; Dietz, S.; Dührkop, K.; et al. Current Challenges in Plant Eco-Metabolomics. *Int. J. Mol. Sci.* **2018**, *19*, 1385. [CrossRef]
75. Hong, J.; Yang, L.; Zhang, D.; Shi, J. Plant Metabolomics: An Indispensable System Biology Tool for Plant Science. *Int. J. Mol. Sci.* **2016**, *17*, 767. [CrossRef] [PubMed]
76. Piasecka, A.; Kachlicki, P.; Stobiecki, M. Analytical Methods for Detection of Plant Metabolomes Changes in Response to Biotic and Abiotic Stresses. *Int. J. Mol. Sci.* **2019**, *20*, 379. [CrossRef] [PubMed]
77. Parida, A.K.; Panda, A.; Rangani, J. Metabolomics-Guided Elucidation of Abiotic Stress Tolerance Mechanisms in Plants. In *Plant Metabolites and Regulation under Environmental Stress*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 89–131.
78. Gong, Z.-G.; Hu, J.; Wu, X.; Xu, Y.-J. The Recent Developments in Sample Preparation for Mass Spectrometry-Based Metabolomics. *Crit. Rev. Anal. Chem.* **2017**, *47*, 325–331. [CrossRef] [PubMed]
79. Corrales, A.; Carrillo, L.; Lasierra, P.; Nebauer, S.G.; Dominguez-Figueroa, J.; Renau-Morata, B.; Pollmann, S.; Granell, A.; Molina, R.; Vicente-Carbajosa, J.; et al. Multifaceted Role of Cycling DOF Factor 3 (CDF3) in the Regulation of Flowering Time and Abiotic Stress Responses in *Arabidopsis*. *Plant Cell Environ.* **2017**, *40*, 748–764. [CrossRef]

80. T'Kindt, R.; Morreel, K.; Deforce, D.; Boerjan, W.; van Bocxlaer, J. Joint GC–MS and LC–MS Platforms for Comprehensive Plant Metabolomics: Repeatability and Sample Pre-Treatment. *J. Chromatogr. B* **2009**, *877*, 3572–3580. [CrossRef]

81. Lehmann, T.; Janowitz, T.; Sánchez-Parra, B.; Alonso, M.-M.P.; Trompetter, I.; Piotrowski, M.; Pollmann, S. Arabidopsis NITRILASE 1 Contributes to the Regulation of Root Growth and Development through Modulation of Auxin Biosynthesis in Seedlings. *Front. Plant Sci.* **2017**, *8*, 36. [CrossRef]

82. Bojko, B.; Reyes-Garcés, N.; Bessonneau, V.; Goryński, K.; Mousavi, F.; Souza Silva, E.A.; Pawliszyn, J. Solid-Phase Microextraction in Metabolomics. *TrAC Trends Anal. Chem.* **2014**, *61*, 168–180. [CrossRef]

83. Patel, M.K.; Mishra, A.; Jha, B. Untargeted Metabolomics of Halophytes. In *Marine OMICS*; CRC Press: Boca Raton, FL, USA, 2016; pp. 329–346. ISBN 1315372304.

84. Bianchi, F.; Ilag, L.; Termopoli, V.; Mendez, L. Advances in MS-Based Analytical Methods: Innovations and Future Trends. *J. Anal. Methods Chem.* **2018**, *2018*, 1–2. [CrossRef]

85. Patel, M.; Pandey, S.; Kumar, M.; Haque, M.; Pal, S.; Yadav, N. Plants Metabolome Study: Emerging Tools and Techniques. *Plants* **2021**, *10*, 2409. [CrossRef] [PubMed]

86. Nakabayashi, R.; Saito, K. Metabolomics for Unknown Plant Metabolites. *Anal. Bioanal. Chem.* **2013**, *405*, 5005–5011. [CrossRef]

87. Tautenhahn, R.; Patti, G.J.; Rinehart, D.; Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **2012**, *84*, 5035–5039. [CrossRef]

88. Smith, C.A.; Maille, G.O.; Want, E.J.; Qin, C.; Trauger, S.A.; Brandon, T.R.; Custodio, D.E.; Abagyan, R.; Siuzdak, G. METLIN. *Ther. Drug Monit.* **2005**, *27*, 747–751. [CrossRef]

89. Behrends, V.; Tredwell, G.D.; Bundy, J.G. A Software Complement to AMDIS for Processing GC-MS Metabolomic Data. *Anal. Biochem.* **2011**, *415*, 206–208. [CrossRef]

90. Xia, J.; Psychogios, N.; Young, N.; Wishart, D.S. MetaboAnalyst: A Web Server for Metabolomic Data Analysis and Interpretation. *Nucleic Acids Res.* **2009**, *37*, W652–W660. [CrossRef] [PubMed]

91. Sakurai, T.; Yamada, Y.; Sawada, Y.; Matsuda, F.; Akiyama, K.; Shinozaki, K.; Hirai, M.Y.; Saito, K. PRIMe Update: Innovative Content for Plant Metabolomics and Integration of Gene Expression and Metabolite Accumulation. *Plant Cell Physiol.* **2013**, *54*, e5. [CrossRef] [PubMed]

92. Sampaio, M.; Rocha, M.; Dias, O. Exploring Synergies between Plant Metabolic Modelling and Machine Learning. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 1885–1900. [CrossRef]

93. Johnson, S.R.; Lange, B.M. Open-access metabolomics databases for natural product research: Present capabilities and future potential. *Front. Bioeng. Biotechnol.* **2015**, *3*, 3–22. [CrossRef]

94. Fan, J.; Zheng, J.; Wu, L.; Zhang, F. Estimation of Daily Maize Transpiration Using Support Vector Machines, Extreme Gradient Boosting, Artificial and Deep Neural Networks Models. *Agric. Water Manag.* **2021**, *245*, 106547. [CrossRef]

95. Lenart, C.; Burai, P.; Smailbegovic, A.; Biro, T.; Katona, Z.; Andricevic, R. Multi-Sensor Integration and Mapping Strategies for the Detection and Remediation of the Red Mud Spill in Kolontar, Hungary: Estimating the Thickness of the Spill Layer Using Hyperspectral Imaging and Lidar. In Proceedings of the 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, Portugal, 6–9 June 2011; pp. 1–4.

96. Dalponte, M.; Bruzzone, L.; Gianelle, D. Tree Species Classification in the Southern Alps Based on the Fusion of Very High Geometrical Resolution Multispectral/Hyperspectral Images and LiDAR Data. *Remote Sens. Environ.* **2012**, *123*, 258–270. [CrossRef]

97. Porfirio, L.L.; Harris, R.M.B.; Lefroy, E.C.; Hugh, S.; Gould, S.F.; Lee, G.; Bindoff, N.L.; Mackey, B. Improving the Use of Species Distribution Models in Conservation Planning and Management under Climate Change. *PLoS ONE* **2014**, *9*, e113749. [CrossRef]

98. Mulder, H.A.; Rönnegård, L.; Fikse, W.F.; Veerkamp, R.F.; Strandberg, E. Estimation of Genetic Variance for Macro- and Micro-Environmental Sensitivity Using Double Hierarchical Generalized Linear Models. *Genet. Sel. Evol.* **2013**, *45*, 23. [CrossRef]

99. Wen, J.; Yang, J.; Jiang, B.; Song, H.; Wang, H. Big Data Driven Marine Environment Information Forecasting: A Time Series Prediction Network. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 4–18. [CrossRef]

100. Beckage, B.; Gross, L.J.; Lacasse, K.; Carr, E.; Metcalf, S.S.; Winter, J.M.; Howe, P.D.; Fefferman, N.; Franck, T.; Zia, A.; et al. Linking Models of Human Behaviour and Climate Alters Projected Climate Change. *Nat. Clim. Change* **2018**, *8*, 79–84. [CrossRef]

101. Shit, P.K.; Bhunia, G.S.; Maiti, R. Spatial Analysis of Soil Properties Using GIS Based Geostatistics Models. *Model Earth Syst. Environ.* **2016**, *2*, 107. [CrossRef]

102. Ihl, T.; Bautista, F.; Cejudo Ruíz, F.R.; Delgado, M.d.C.; Quintana Owen, P.; Aguilar, D.; Goguitchaichvili, A. Concentration of Toxic Elements in Topsoils of the Metropolitan Area of Mexico City: A Spatial Analysis Using Ordinary Kriging and Indicator Kriging. *Rev. Int. Contam. Ambient.* **2015**, *31*, 47–62.

103. Spencer, P.S.; Lagrange, E.; Camu, W. ALS and Environment: Clues from Spatial Clustering? *Rev. Neurol.* **2019**, *175*, 652–663. [CrossRef] [PubMed]

104. Pleil, J.D.; Stiegel, M.A.; Madden, M.C.; Sobus, J.R. Heat Map Visualization of Complex Environmental and Biomarker Measurements. *Chemosphere* **2011**, *84*, 716–723. [CrossRef] [PubMed]

105. Neset, T.-S.; Opach, T.; Lion, P.; Lilja, A.; Johansson, J. Map-Based Web Tools Supporting Climate Change Adaptation. *Prof. Geogr.* **2016**, *68*, 103–114. [CrossRef]

106. Iturbide, M.; Gutiérrez, J.M.; Alves, L.M.; Bedia, J.; Cerezo-Mota, R.; Cimadevilla, E.; Cofiño, A.S.; di Luca, A.; Faria, S.H.; Gorodetskaya, I.V. An Update of IPCC Climate Reference Regions for Subcontinental Analysis of Climate Model Data: Definition and Aggregated Datasets. *Earth Syst. Sci. Data* **2020**, *12*, 2959–2970. [CrossRef]

107. Mudelsee, M. Trend Analysis of Climate Time Series: A Review of Methods. *Earth Sci. Rev.* **2019**, *190*, 310–322. [CrossRef]

108. Wei, R.; Wang, J.; Su, M.; Jia, E.; Chen, S.; Chen, T.; Ni, Y. Missing Value Imputation Approach for Mass Spectrometry-Based Metabolomics Data. *Sci. Rep.* **2018**, *8*, 663. [CrossRef]

109. Hamamoto, A.H.; Carvalho, L.F.; Sampaio, L.D.H.; Abrão, T.; Proença, M.L., Jr. Network Anomaly Detection System Using Genetic Algorithm and Fuzzy Logic. *Expert Syst. Appl.* **2018**, *92*, 390–402. [CrossRef]

110. Bermingham, M.L.; Pong-Wong, R.; Spiliopoulou, A.; Hayward, C.; Rudan, I.; Campbell, H.; Wright, A.F.; Wilson, J.F.; Agakov, F.; Navarro, P. Application of High-Dimensional Feature Selection: Evaluation for Genomic Prediction in Man. *Sci. Rep.* **2015**, *5*, 10312. [CrossRef]

111. Manor, O.; Borenstein, E. MUSiCC: A Marker Genes Based Framework for Metagenomic Normalization and Accurate Profiling of Gene Abundances in the Microbiome. *Genome Biol.* **2015**, *16*, 53. [CrossRef]

112. Gong, A.-D.; Lian, S.-B.; Wu, N.-N.; Zhou, Y.-J.; Zhao, S.-Q.; Zhang, L.-M.; Cheng, L.; Yuan, H.-Y. Integrated Transcriptomics and Metabolomics Analysis of Catechins, Caffeine and Theanine Biosynthesis in Tea Plant (*Camellia sinensis*) over the Course of Seasons. *BMC Plant Biol.* **2020**, *20*, 294. [CrossRef]

113. Li, Y.; Fang, J.; Qi, X.; Lin, M.; Zhong, Y.; Sun, L.; Cui, W. Combined Analysis of the Fruit Metabolome and Transcriptome Reveals Candidate Genes Involved in Flavonoid Biosynthesis in *Actinidia arguta*. *Int. J. Mol. Sci.* **2018**, *19*, 1471. [CrossRef]

114. Unamba, C.I.N.; Nag, A.; Sharma, R.K. Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants. *Front. Plant Sci.* **2015**, *6*, 1074. [CrossRef] [PubMed]

115. Todaka, D.; Zhao, Y.; Yoshida, T.; Kudo, M.; Kidokoro, S.; Mizoi, J.; Kodaira, K.-S.; Takebayashi, Y.; Kojima, M.; Sakakibara, H.; et al. Temporal and Spatial Changes in Gene Expression, Metabolite Accumulation and Phytohormone Content in Rice Seedlings Grown under Drought Stress Conditions. *Plant J.* **2017**, *90*, 61–78. [CrossRef]

116. Walsh, K.J.E.; McBride, J.L.; Klotzbach, P.J.; Balachandran, S.; Camargo, S.J.; Holland, G.; Knutson, T.R.; Kossin, J.P.; Lee, T.; Sobel, A. Tropical Cyclones and Climate Change. *Wiley Interdiscip. Rev. Clim. Change* **2016**, *7*, 65–89. [CrossRef]

117. Anderson, T.R.; Hawkins, E.; Jones, P.D. CO_2, the Greenhouse Effect and Global Warming: From the Pioneering Work of Arrhenius and Callendar to Today's Earth System Models. *Endeavour* **2016**, *40*, 178–187. [CrossRef]

118. Demaria, E.M.C.; Palmer, R.N.; Roundy, J.K. Regional Climate Change Projections of Streamflow Characteristics in the Northeast and Midwest US. *J. Hydrol. Reg. Stud.* **2016**, *5*, 309–323. [CrossRef]

119. Cucurachi, S.; Scherer, L.; Guinée, J.; Tukker, A. Life Cycle Assessment of Food Systems. *One Earth* **2019**, *1*, 292–297. [CrossRef]

120. Ivanova, D.; Stadler, K.; Steen-Olsen, K.; Wood, R.; Vita, G.; Tukker, A.; Hertwich, E.G. Environmental Impact Assessment of Household Consumption. *J. Ind. Ecol.* **2016**, *20*, 526–536. [CrossRef]

121. Ahmad, M.; Jiang, P.; Majeed, A.; Umar, M.; Khan, Z.; Muhammad, S. The Dynamic Impact of Natural Resources, Technological Innovations and Economic Growth on Ecological Footprint: An Advanced Panel Data Estimation. *Resour. Policy* **2020**, *69*, 101817. [CrossRef]

122. Schleder, G.R.; Padilha, A.C.M.; Acosta, C.M.; Costa, M.; Fazzio, A. From DFT to Machine Learning: Recent Approaches to Materials Science—A Review. *J. Phys. Mater.* **2019**, *2*, 032001. [CrossRef]

123. Rolnick, D.; Donti, P.L.; Kaack, L.H.; Kochanski, K.; Lacoste, A.; Sankaran, K.; Ross, A.S.; Milojevic-Dupont, N.; Jaques, N.; Waldman-Brown, A.; et al. Tackling Climate Change with Machine Learning. *ACM Comput. Surv.* **2023**, *55*, 1–96. [CrossRef]

124. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L. Integrating Satellite and Climate Data to Predict Wheat Yield in Australia Using Machine Learning Approaches. *Agric. For. Meteorol.* **2019**, *274*, 144–159. [CrossRef]

125. Hudson, I.L. Data Integration Using Advances in Machine Learning in Drug Discovery and Molecular Biology. *Artif. Neural Netw.* **2021**, *2190*, 167–184.

126. Moore, B.M.; Wang, P.; Fan, P.; Leong, B.; Schenck, C.A.; Lloyd, J.P.; Lehti-Shiu, M.D.; Last, R.L.; Pichersky, E.; Shiu, S.-H. Robust Predictions of Specialized Metabolism Genes through Machine Learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 2344–2353. [CrossRef]

127. Hernández-Alvarado, R.B.; Madariaga-Mazón, A.; Martinez-Mayorga, K. Prediction of Toxicity of Secondary Metabolites. *Phys. Sci. Rev.* **2019**, *4*. [CrossRef]

128. Jia, X.; Zhu, Y.; Zhang, R.; Zhu, Z.; Zhao, T.; Cheng, L.; Gao, L.; Liu, B.; Zhang, X.; Wang, Y. Ionomic and Metabolomic Analyses Reveal the Resistance Response Mechanism to Saline-Alkali Stress in *Malus halliana* Seedlings. *Plant Physiol. Biochem.* **2020**, *147*, 77–90. [CrossRef]

129. Thanmalagan, R.R.; Roy, A.; Jayaprakash, A.; Lakshmi, P.T.V. Comprehensive Meta-Analysis and Machine Learning Approaches Identified the Role of Novel Drought Specific Genes in *Oryza Sativa*. *Plant Gene* **2022**, *32*, 100382. [CrossRef]

130. Teles, F.F.F. Chronic Poisoning by Hydrogen Cyanide in Cassava and Its Prevention in Africa and Latin America. *Food Nutr. Bull.* **2002**, *23*, 407–412. [CrossRef]

131. Craig, D.B.; Guimond, M.S. Analysis of Cyanide Using Fluorogenic Derivatization and Capillary Electrophoresis. *Food Chem.* **2022**, *370*, 131377. [CrossRef]

132. Sahin, S. Cyanide Poisoning in Children Caused by Apricot Seeds. *J. Health Med. Inform.* **2011**, *2*, 1.

133. Thodberg, S.; del Cueto, J.; Mazzeo, R.; Pavan, S.; Lotti, C.; Dicenta, F.; Jakobsen Neilson, E.H.; Møller, B.L.; Sánchez-Pérez, R. Elucidation of the Amygdalin Pathway Reveals the Metabolic Basis of Bitter and Sweet Almonds (*Prunus Dulcis*). *Plant Physiol.* **2018**, *178*, 1096–1111. [CrossRef]

134. Pelentir, N.; Block, J.M.; Monteiro Fritz, A.R.; Reginatto, V.; Amante, E.R. Production and chemical characterization of peach (*Prunus persica*) kernel flour. *J. Food Process Eng.* **2011**, *34*, 1253–1265. [CrossRef]

135. Touré, A.; Xueming, X. Flaxseed Lignans: Source, Biosynthesis, Metabolism, Antioxidant Activity, Bio-Active Components, and Health Benefits. *Compr. Rev. Food Sci. Food Saf.* **2010**, *9*, 261–269. [CrossRef]

136. Pramitha, A.R.; Harijono, H.; Wulan, S.N. Comparison of Cyanide Content in Arbila Beans (*Phaseolus Lunatus* L.) of East Nusa Tenggara Using Picrate and Acid Hydrolysis Methods. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *924*, 012031. [CrossRef]

137. Keeler, R.F.; van Kampen, K.R.; James, L.F. *Effects of Poisonous Plants on Livestock*; Elsevier: Amsterdam, The Netherlands, 2013; ISBN 1483270181.

138. Liu, Q.; Zhuo, L.; Liu, L.; Zhu, S.; Sunnassee, A.; Liang, M.; Zhou, L.; Liu, Y. Seven Cases of Fatal Aconite Poisoning: Forensic Experience in China. *Forensic Sci. Int.* **2011**, *212*, e5–e9. [CrossRef] [PubMed]

139. Disel, N.R.; Yilmaz, M.; Kekec, Z.; Karanlik, M. Poisoned after Dinner: Dolma with *Datura Stramonium. Turk. J. Emerg. Med.* **2015**, *15*, 51–55. [CrossRef]

140. Thornton, S.L.; Darracq, M.; Lo, J.; Cantrell, F.L. Castor Bean Seed Ingestions: A State-Wide Poison Control System's Experience. *Clin. Toxicol.* **2014**, *52*, 265–268. [CrossRef]

141. Cortinovis, C.; Caloni, F. Alkaloid-Containing Plants Poisonous to Cattle and Horses in Europe. *Toxins* **2015**, *7*, 5301–5307. [CrossRef]

142. Kerchner, A.; Farkas, Á. Worldwide Poisoning Potential of Brugmansia and Datura. *Forensic Toxicol.* **2020**, *38*, 30–41. [CrossRef]

143. Lee, M.R.; Dukan, E.; Milne, I. Three Poisonous Plants (*Oenanthe*, *Cicuta* and *Anamirta*) That Antagonise the Effect of Γ-Aminobutyric Acid in Human Brain. *J. R. Coll. Physicians Edinb.* **2020**, *50*, 80–86. [CrossRef] [PubMed]

144. Orlygsson, J.; Scully, S.M. Influence of Inhibitory Compounds on Biofuel Production from Oxalate-Rich Rhubarb Leaf Hydrolysates Using Thermoanaerobacter Thermohydrosulfuricus Strain AK91. *Fuels* **2021**, *2*, 71–86. [CrossRef]

145. Kingma, J.S.; Frenay, I.M.; Meinders, A.J.; van Dijk, V.F.; Harmsze, A.M. A Poisonous Spring Smoothie with Wild Herbs: Accidental Intoxication with Foxglove (*Digitalis Purpurea*). *Ned. Tijdschr. Geneeskd.* **2020**, *164*, D5306.

146. Bai, X.; Wang, G.; Ren, Y.; Han, J. Detection of Highly Poisonous *Nerium oleander* Using Quantitative Real-Time PCR with Specific Primers. *Toxins* **2022**, *14*, 776. [CrossRef]

147. Woolery, S.; Willner, J.; Prahlow, J.A.; Douglas, E. Death After Poison Ivy Smoke Inhalation. *Am. J. Forensic Med. Pathol.* **2022**, *43*, 359–362. [CrossRef]

148. Wang, P.; Schumacher, A.M.; Shiu, S.-H. Computational Prediction of Plant Metabolic Pathways. *Curr. Opin. Plant Biol.* **2022**, *66*, 102171. [CrossRef]

149. Toubiana, D.; Fernie, A.R.; Nikoloski, Z.; Fait, A. Network Analysis: Tackling Complex Data to Study Plant Metabolism. *Trends Biotechnol.* **2013**, *31*, 29–36. [CrossRef] [PubMed]

150. Hino, M.; Benami, E.; Brooks, N. Machine Learning for Environmental Monitoring. *Nat. Sustain.* **2018**, *1*, 583–588. [CrossRef]

151. Schläpfer, P.; Zhang, P.; Wang, C.; Kim, T.; Banf, M.; Chae, L.; Dreher, K.; Chavali, A.K.; Nilo-Poyanco, R.; Bernard, T.; et al. Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. *Plant Physiol.* **2017**, *173*, 2041–2059. [CrossRef] [PubMed]

152. Sarabi, B.; Ghashghaie, J. Evaluating the Physiological and Biochemical Responses of Melon Plants to NaCl Salinity Stress Using Supervised and Unsupervised Statistical Analysis. *Plant Stress* **2022**, *4*, 100067. [CrossRef]

153. Akbar, A.; Kuanar, A.; Patnaik, J.; Mishra, A.; Nayak, S. Application of Artificial Neural Network Modeling for Optimization and Prediction of Essential Oil Yield in Turmeric (*Curcuma longa* L.). *Comput. Electron. Agric.* **2018**, *148*, 160–178. [CrossRef]

154. Almeida, B.K.; Garg, M.; Kubat, M.; Afkhami, M.E. Not That Kind of Tree: Assessing the Potential for Decision Tree–Based Plant Identification Using Trait Databases. *Appl. Plant Sci.* **2020**, *8*. [CrossRef] [PubMed]

155. de Oliveira Almeida, R.; Valente, G.T. Predicting Metabolic Pathways of Plant Enzymes without Using Sequence Similarity: Models from Machine Learning. *Plant Genome* **2020**, *13*, e11379. [CrossRef] [PubMed]

156. Ntatsi, G.; Aliferis, K.A.; Rouphael, Y.; Napolitano, F.; Makris, K.; Kalala, G.; Katopodis, G.; Savvas, D. Salinity Source Alters Mineral Composition and Metabolism of *Cichorium Spinosum. Environ. Exp. Bot.* **2017**, *141*, 113–123. [CrossRef]

157. Rouphael, Y.; Raimondi, G.; Lucini, L.; Carillo, P.; Kyriacou, M.C.; Colla, G.; Cirillo, V.; Pannico, A.; El-Nakhel, C.; de Pascale, S. Physiological and Metabolic Responses Triggered by Omeprazole Improve Tomato Plant Tolerance to NaCl Stress. *Front. Plant Sci.* **2018**, *9*, 249. [CrossRef]

158. Cui, J.; Davanture, M.; Zivy, M.; Lamade, E.; Tcherkez, G. Metabolic Responses to Potassium Availability and Waterlogging Reshape Respiration and Carbon Use Efficiency in Oil Palm. *New Phytol.* **2019**, *223*, 310–322. [CrossRef] [PubMed]

159. Albacete, A.; Ghanem, M.E.; Dodd, I.C.; Pérez-Alfocea, F. Principal Component Analysis of Hormone Profiling Data Suggests an Important Role for Cytokinins in Regulating Leaf Growth and Senescence of Salinized Tomato. *Plant Signal. Behav.* **2010**, *5*, 45–48. [CrossRef] [PubMed]

160. Jurczyk, B.; Rapacz, M.; Pociecha, E.; Kościelniak, J. Changes in Carbohydrates Triggered by Low Temperature Waterlogging Modify Photosynthetic Acclimation to Cold in *Festuca Pratensis. Environ. Exp. Bot.* **2016**, *122*, 60–67. [CrossRef]

161. Zhang, Y.; Wang, Y.; Ding, Z.; Wang, H.; Song, L.; Jia, S.; Ma, D. Zinc Stress Affects Ionome and Metabolome in Tea Plants. *Plant Physiol. Biochem.* **2017**, *111*, 318–328. [CrossRef] [PubMed]

162. Zheng, Y.-Y.; Kong, J.-L.; Jin, X.-B.; Wang, X.-Y.; Zuo, M. CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture. *Sensors* **2019**, *19*, 1058. [CrossRef]

163. Ma, C.; Zhang, H.H.; Wang, X. Machine Learning for Big Data Analytics in Plants. *Trends Plant Sci.* **2014**, *19*, 798–808. [CrossRef]

164. Ayoub Shaikh, T.; Rasool, T.; Rasheed Lone, F. Towards Leveraging the Role of Machine Learning and Artificial Intelligence in Precision Agriculture and Smart Farming. *Comput. Electron. Agric.* **2022**, *198*, 107119. [CrossRef]

165. Storm, H.; Baylis, K.; Heckelei, T. Machine Learning in Agricultural and Applied Economics. *Eur. Rev. Agric. Econ.* **2020**, *47*, 849–892. [CrossRef]

166. Jane, J.B.; Ganeshi, E.N. A Review on Big Data with Machine Learning and Fuzzy Logic for Better Decision Making. *Int. J. Sci. Technol. Res.* **2019**, *8*, 1121–1125.

167. Scalbert, A.; Brennan, L.; Fiehn, O.; Hankemeier, T.; Kristal, B.S.; van Ommen, B.; Pujos-Guillot, E.; Verheij, E.; Wishart, D.; Wopereis, S. Mass-Spectrometry-Based Metabolomics: Limitations and Recommendations for Future Progress with Particular Focus on Nutrition Research. *Metabolomics* **2009**, *5*, 435–458. [CrossRef] [PubMed]

168. Hogarth, R.M.; Soyer, E. Communicating Forecasts: The Simplicity of Simulated Experience. *J. Bus. Res.* **2015**, *68*, 1800–1809. [CrossRef]

169. Westerhuis, J.A.; Hoefsloot, H.C.J.; Smit, S.; Vis, D.J.; Smilde, A.K.; van Velzen, E.J.J.; van Duijnhoven, J.P.M.; van Dorsten, F.A. Assessment of PLSDA Cross Validation. *Metabolomics* **2008**, *4*, 81–89. [CrossRef]

170. Bacardit, J.; Widera, P.; Lazzarini, N.; Krasnogor, N. Hard Data Analytics Problems Make for Better Data Analysis Algorithms: Bioinformatics as an Example. *Big Data* **2014**, *2*, 164–176. [CrossRef] [PubMed]

171. Najafabadi, M.M.; Villanustre, F.; Khoshgoftaar, T.M.; Seliya, N.; Wald, R.; Muharemagic, E. Deep Learning Applications and Challenges in Big Data Analytics. *J. Big Data* **2015**, *2*, 1. [CrossRef]