

Review

Metabolomics-Guided Elucidation of Plant Abiotic Stress Responses in the 4IR Era: An Overview

Morena M. Tinte ¹, Kekeletso H. Chele ¹, Justin J. J. van der Hooft ^{2,*}  and Fidele Tugizimana ^{1,3,*} 

¹ Department of Biochemistry, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa; morenatinte@gmail.com (M.M.T.); ckekeletso@gmail.com (K.H.C.)

² Bioinformatics Group, Wageningen University, 6708 PB Wageningen, The Netherlands

³ International Research and Development Division, Omnia Group, Ltd., Johannesburg 2021, South Africa

* Correspondence: justin.vanderhooft@wur.nl (J.J.J.v.d.H.); Fidele.Tugizimana@omnia.co.za (F.T.)

Abstract: Plants are constantly challenged by changing environmental conditions that include abiotic stresses. These are limiting their development and productivity and are subsequently threatening our food security, especially when considering the pressure of the increasing global population. Thus, there is an urgent need for the next generation of crops with high productivity and resilience to climate change. The dawn of a new era characterized by the emergence of fourth industrial revolution (4IR) technologies has redefined the ideological boundaries of research and applications in plant sciences. Recent technological advances and machine learning (ML)-based computational tools and omics data analysis approaches are allowing scientists to derive comprehensive metabolic descriptions and models for the target plant species under specific conditions. Such accurate metabolic descriptions are imperatively essential for devising a roadmap for the next generation of crops that are resilient to environmental deterioration. By synthesizing the recent literature and collating data on metabolomics studies on plant responses to abiotic stresses, in the context of the 4IR era, we point out the opportunities and challenges offered by omics science, analytical intelligence, computational tools and big data analytics. Specifically, we highlight technological advancements in (plant) metabolomics workflows and the use of machine learning and computational tools to decipher the dynamics in the chemical space that define plant responses to abiotic stress conditions.

Keywords: abiotic stress; metabolomics; 4IR technologies; automation; machine learning



Citation: Tinte, M.M.; Chele, K.H.; van der Hooft, J.J.J.; Tugizimana, F. Metabolomics-Guided Elucidation of Plant Abiotic Stress Responses in the 4IR Era: An Overview. *Metabolites* **2021**, *11*, 445. <https://doi.org/10.3390/metabo11070445>

Academic Editor: Peter Meikle

Received: 25 May 2021

Accepted: 3 July 2021

Published: 8 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction—A Dawn of a New Era and a Prime to Plant Defenses

1.1. The Fourth Industrial Revolution (4IR) Era

The Fourth Industrial Revolution (4IR) era entails the integration of advanced technologies in the physical, digital and biological domains. This includes the confluence and convergence of emerging technologies such as artificial intelligence (AI), the Internet of Things (IoT), big data analytics, cloud computing, robotics and wireless telecommunications [1,2]. These innovative technologies have brought about paradigm shifts and are disruptively boosting many industries globally by encouraging new models that enable the acquisition, sharing, and use of data and resources to produce improved products/services in a faster, cheaper, more effective and sustainable manner [3]. In life sciences, particularly in the field of metabolomics—a multidisciplinary omics science that studies metabolism (Section 2.1)—some of these 4IR technologies have been and are integral components of metabolomics workflows. It suffices to highlight the use of analytical platforms that are equipped with analytical and artificial intelligence (A/AI), generation of big data, application and development of big data analytics involving the use of machine and deep learning (ML and DL) algorithms (Figure 1).

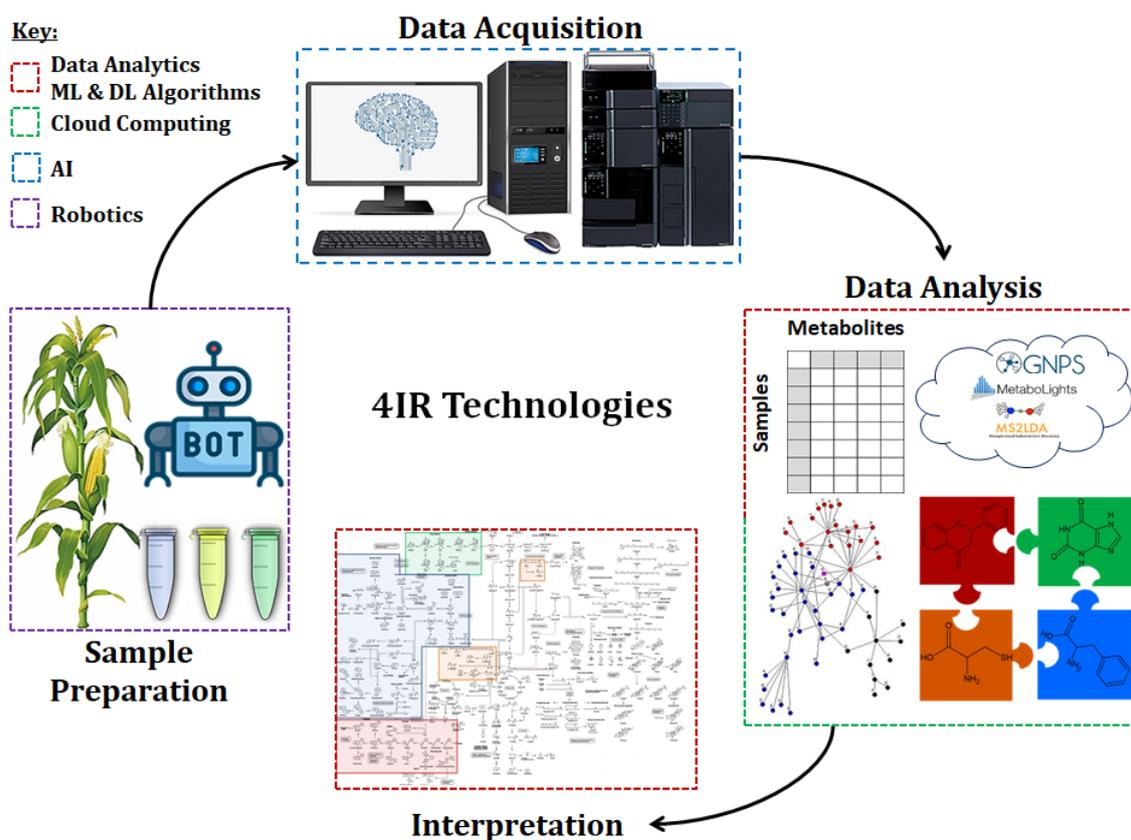


Figure 1. 4IR technologies in the plant metabolomics workflow. The 4IR technologies and their implementation within the plant metabolomics workflows are indicated based on colour, as highlighted by the key. An illustration of the preparation of samples with the assistance of robotics, advancements in analytical platforms, equipped with A/AI for sample analysis. The generated (big) data can be uploaded on cloud-based servers, e-infrastructures for data analysis, storage and sharing. Some of these web-based suits include MetaboAnalyst, XCMS Online, MetExplore, PhenoMeNal and GNPS. Computational tools in these e-infrastructures involve the use of chemometrics methods, ML and DL algorithms. Metabolic pathway reconstruction and network analysis are often used for biological interpretation of metabolomics data. The IoT is an indispensable component supporting most of these cloud metabolomics frameworks.

As the field matures, with advancements in technologies, development and applications of state-of-the-art bioinformatics and computational tools, equipped with ML algorithms, are gaining momentum for data mining and interpretation [4,5]. Typical widely adopted examples include the Global Natural Product Social Molecular Networking (GNPS), an ecosystem of tandem mass spectrometry (MS/MS) data storage, analysis and sharing [6], MetaboLights, a cloud computing based repository that enables the sharing and re-use of data and meta-data [7], MS2LDA, a software tool that extracts co-occurring mass fragments and neutral losses from MS/MS spectra using an unsupervised ML algorithm [8], MetaboAnalyst, a web-based service consisting of modules for data pretreatment, mining and pathway analysis, XCMS, a cloud-based data analysis suite for preprocessing untargeted liquid chromatography-mass spectrometry (LC-MS) data, statistical analysis, pathway analysis and multi-omic data integration, MetExplore, an environment for the curation of metabolic networks, and PhenoMeNal, a e-infrastructure with a collection of workflows and tools for metabolomics analysis pipelines [5] (Figure 1). The digitization of mass spectra to aid in biological interpretation of plant metabolomics data is illustrated in a study of more than 70 *Rhamnaceae* plant extracts [9] where the authors illuminate clade-specific chemical signatures annotated through an integrative computational metabolomics workflow.

The increasing momentum and use of 4IR technologies in life sciences, particularly in plant metabolomics, which is the focus of this review, is redefining the ideological

boundaries of research and applications in the field. Recent advances in generating comprehensive biological (metabolomics) datasets at high throughput, in combination with enhanced capabilities to mine and interpret these datasets are increasingly allowing scientists to derive a comprehensive understanding for crop plants under consideration [10,11]. Such accurate and predictive models that describe the metabolism of plants under specific conditions would provide novel insights that identify key biological bottlenecks in regard to plant growth and productivity [11,12].

This review focuses on the use of metabolomics in interrogating plant responses to adverse environmental conditions, with a particular attention to the 4IR technologies in this multidisciplinary omics science. Metabolomics is increasingly enabling the decoding of the language of cells at molecular level, advancing the understanding of regulatory network rules and mechanistic events at cellular and chemical space of the plant under consideration. Plants are naturally sessile organisms, and are thus susceptible to changing environmental conditions such as abiotic stress factors that include drought, salinity, extreme high and low temperatures, heavy metals, light and radiation [13,14]. These abiotic stress factors can negatively affect plant growth, development and productivity, and subsequently the agricultural yield. It is, therefore, imperative to comprehensively and predictively understand the plant metabolism under abiotic stresses, as such fundamental and actionable insights (adding to the current knowledgebase, Section 1.2) will contribute to the development of plants with enhanced resilience and productivity, and support strategies that promote plant growth under abiotic stress conditions [15,16]. To logically articulate these aspects, the review is structured to comprise four main components. In Section 1.1. and Section 1.2, the 4IR era is briefly defined and introduced as well as the current models of plant defense mechanisms. The second main section then elaborates on 4IR technologies in the context of (plant) metabolomics workflows, from sample preparation step to the annotation of metabolites. Automation and technological advancements in analytical techniques, the use of machine learning and computational tools to aid in deciphering the dynamics in the chemical space that define plant defense responses are highlighted. The application of metabolomics to decode plant responses to adverse environmental conditions is increasing, and this is highlighted in the third main section. The review concludes with the fourth main section that contains an outlook on expected developments in the plant metabolomics, driven by advancements of 4IR technologies.

1.2. Plant Defense Mechanisms—Current Models

Evolutionally, plants have developed eminently intricate immune systems and defense mechanisms to respond to biotic and abiotic stresses [17]. These responses are stress-dependent, at cellular and molecular levels, but there are also some overlaps in biochemical and physiological events that define plant response to stresses [18,19]. Over the years, various studies have generated a substantial knowledge-base and understanding of plant responses to different adverse environmental stresses, formulating models that explain mechanistic events that govern plant responses [20–22]. It suffices here to highlight some of the elucidated molecular events that define plant defense mechanisms, involving the danger perception to activation of downstream molecular and cellular phenomonologies [23]. Perception of abiotic stresses generally results in the generation of reactive oxygen species (ROS) that act as early signaling molecules. Elevated levels of ROS act as a signaling wave that is involved in a defense-related reconfiguration of the hormonal network comprising abscisic acid (ABA), gibberellins (GAs), auxins, jasmonic acid (JA), salicylic acid (SA) ethylene, cytokinins (CKs) and brassinosteroids (BRs). These phytohormones are primary signaling molecules that trigger the expression of stress-related genes and induction of metabolic reprogramming and physiological changes that result in abiotic stress tolerance or resistance [16,21]. The outcome is determined by the directional shift of metabolic reconfigurations and fluxes—either towards an irreversible damage by the stress factor or effective stress resistance and acclimation of the plants. Some of the (general) negative impacts of abiotic stresses on the plant physiology include a reduction in both transduction

and photosynthesis rates, a decrease in stomatal conductance and in leaf water content and, subsequently, a reduced growth rate [24,25].

The current understanding of the stress signaling and responses is still the tip of an iceberg. Comprehensive predictive models that describe the activation of different signals, sensing mechanisms, downstream changes in gene expression, metabolism, physiology, growth and development are still lacking. Systems biology approaches, particularly omics sciences (genomics, transcriptomics, proteomics and metabolomics) carried out separately or in integrated manner, hold unique opportunities to generate novel insights describing, comprehensively and predictively the metabolism of plants under specific conditions [10–12]. The plant metabolome can echo the effect of environmental stress conditions, and therefore, metabolomics can be applied to provide a snapshot of the plant metabolism at a cellular and molecular level by monitoring the changes in metabolite levels and fluxes, which reflect the biological and physiological processes in response to the stressful conditions [26,27]. Thus, the focal point of this review is the application and potential of metabolomics with 4IR-inspired tools and technologies to aid in elucidating plant defense responses to abiotic stress conditions by illuminating the plant specialized metabolome [28–30].

2. 4IR Technologies and Plant Metabolomics

Metabolomics is classically defined as an omics science that aims at the analysis of the entire complement of small molecular weight molecules, namely metabolites (≤ 1500 Da in size), within a biological system under given physiological conditions [29,31]. Metabolomics incorporates the domains of biology, chemistry, chemometric, statistics and computer science. This multidisciplinary scientific field has matured over the last two decades and has gained popularity, particularly in the life sciences, and is becoming indispensable in interrogating cellular biochemistry and elucidating the mechanisms responsible for metabolic changes in response to different physiological conditions [5,32–35]. In plant sciences, metabolomics has been successfully applied in a broad spectrum of studies including metabolic pathway studies [29], relating genotype and biochemical phenotype [36], silent phenotype mutations [37], plant-environment interactions [27] and plant priming, a phenomenon that pre-conditions plants for enhanced defense against stresses [38]. Metabolomics, particularly for large scale studies, is, however, limited by various bottlenecks; hence, the focus to overcome these challenges has intensified through the development of computational metabolomics tools and improved technologies [9], most of which are 4IR-driven (Figure 1) and are articulated in the following sub-sections on 4IR technologies in metabolomics workflows.

2.1. Automation in Sample Preparation

Sample preparation is a key aspect of metabolomic studies, as it is responsible for removing proteins and cellular debris from the metabolites of interest. An ideal sample preparation method should maintain sample integrity, and be simple and robust. Traditional sample preparation steps involve the quenching of the metabolism, cellular lysis with organic solvents or detergents, sonication and centrifugation. The preparation of large numbers of samples, utilizing such sample preparation strategies is, however, time consuming and likely to introduce unwanted variation [39,40]. The demand for automated sample preparation strategies (Figure 1) that will diminish human error, improve reproducibility, decrease extraction time and increase throughput for large-scale metabolomics studies is thus evident [41]. Recently, novel automated sample preparation methods (Table 1) have emerged in the field of metabolomics. These methods are solid-phase extraction (SPE), solid-phase microextraction (SPME), liquid-phase microextraction (LPME) methods (dispersive liquid–liquid microextraction (DLLME), hollow fiber liquid–liquid microextraction (HF-LLME), single drop microextraction (SDME) and the recent electromembrane extraction (EME)), accelerated solvent extraction (ASE), supercritical fluid extraction (SFE), microwave-assisted extraction (MAE) and ultrasound-assisted extraction (UAE) [40,42,43].

Table 1. Automated sample preparation methods with their advantages and disadvantages over traditional sample preparation methods.

Method	Description	Advantages	Disadvantages	Originally Automated? (Yes/No)	Reference(s)
Solid-phase extraction (SPE)	Extracts metabolites based on their chemical and physical properties that determine their distribution between the mobile liquid phase and a solid stationary phase. Targeted metabolites are released from stationary phase by changing the mobile phase into the elution solvent.	Enhanced selectivity, rapid, reproducible and economical.	Poor metabolite coverage	No	[43–47]
Solid-phase microextraction (SPME)	Extracts a range of metabolites from a variety of matrices by the insertion of a polymer-coated fiber into either the vial headspace, liquid sample or exposed in vivo. The metabolites diffuse from the sample onto the fiber.	Enhanced sensitivity, minimum invasiveness, enhanced analysis throughput and compatibility with in vivo sampling and extraction.	Time-consuming steps in equilibration of the fibre (30 min) and sample extraction (up to 5 min), low metabolite coverage, and its expensive.	Yes	[40,48–50]
Dispersive liquid–liquid microextraction (DLLME)	Extraction solvent (i.e., water-immiscible organic solvent) is added to a dispersive solvent (i.e., water-miscible solvent), the mixture is then injected into the sample to form a homogenous solution. Induced dispersion increases surface contact between extract and the sample, thus resulting in instantaneous extraction.	Simple, cost-effective, rapid, has high extraction recovery, reduced solvent consumption and has high reproducibility.	Uses halogenated and organic solvents, requires manual/mechanical agitation of the sample for dispersion of the organic solvents in the sample solution, and time-consuming phase separation step.	No	[48,51,52]
Electromembrane extraction (EME)	An electrical field is applied between the sample and the acceptor compartments, separated by a membrane of organic solvent (i.e., the support liquid membrane (SLM)). Charged ionic metabolites are extracted from the sample solution, through the SLM, and into the acceptor compartment. Proteins, salts, etc. are incapable of passing the SLM, thus metabolites are recovered in the aqueous phase.	Enables large-scale automation.	Difficulty of extracting hydrophilic metabolites.	No	[48,52,53]

Table 1. Cont.

Method	Description	Advantages	Disadvantages	Originally Automated? (Yes/No)	Reference(s)
Hollow fiber liquid–liquid microextraction (HF-LLME)	<p>Two-phase mode: The hollow fiber (HF) is soaked with the extraction solvent and exposed to the sample's solution or headspace, an equilibrium between solvent and sample is established, thus resulting in the extraction of metabolites from the sample into the solvent.</p> <p>Three-phase mode: The center of the HF contains an aqueous phase (i.e., acceptor phase), in addition to the soaked HF pores with organic solvent. The HF is exposed to the sample where two equilibria are established. The first is between the sample and the solvent, followed by the second between the solvent and the acceptor phase, thus metabolites are extracted from the sample into the acceptor phase of the HF through the solvent.</p>	Highly selective and concentrate metabolites.		No	[51,52]
Single drop microextraction (SDME)	Similar to HF-LLME. A syringe is used instead of a HF and only a drop of the extractant solvent is required.	Simple, cost-effective and time-saving.	Limited by partial solubility of organic solvents in water, limited extraction volume, metabolite losses due to volatility and dislodgement of the extractant solvent.	No	[51,52,54]
Accelerated solvent extraction (ASE)/Pressurized liquid extraction (PLE)	The solvent's temperature is elevated beyond its boiling point to increase its solubilizing capacity and reduce its viscosity to penetrate into the sample matrix and increase the metabolites' diffusion rate. Additionally, the elevated pressure ensures the solvent remains in the liquid phase and aids it in penetrating through the sample matrix, which maximizes solvent and metabolite contact, and result in effective extraction.	Reduced solvent usage and rapid.		No	[42,55,56]
Supercritical fluid extraction (SFE)	Utilizes gas properties above their critical points as solvents to facilitate the extraction of non-polar to semi-polar metabolites from plant materials.	Enhanced sensitivity and accuracy, reduced extraction time, ideal for thermo-labile metabolites and reduced use of organic solvents.	Very expensive.	Yes	[42,56–58]

Table 1. Cont.

Method	Description	Advantages	Disadvantages	Originally Automated? (Yes/No)	Reference(s)
Microwave-assisted extraction (MAE)	Microwave, electromagnetic radiation with a frequency in the 0.3–300 GHz range, energy is used to extract polar metabolites from plant materials by heating the solvent.	Reduced extraction time (15–20 min), reduced solvent consumption, improved extraction yield and precision.	Operates at relatively high temperature which is problematic for thermally labile metabolites, low extraction yield for non-polar solvents and requires a centrifugation step to remove solid materials from extractant.	No	[57,59,60]
Ultrasound-assisted extraction (UAE)	Utilizes ultrasonic energy and solvents to extract secondary metabolites from various plant materials.	Reduced extraction time, solvent consumption, energy, thermal degradation, extraction temperature and equipment size, enhanced mass transfer, extraction yield and high extract recovery.	Low extraction efficiency.	No	[57,61–63]

The detailed SPME workflow is described in [50]. This method eliminates human error and thus improves precision and sample throughput. In addition, damage to the fragile fiber and consumption of chemicals and their environmental footprint is reduced [64,65]. Automation of dispersive liquid–liquid microextraction based on solidification of floating organic drop (DLLME-SFO) method has been achieved by the integration of a sequential injection analysis (SIA) system with a modified robot that has a 3D printed phase separator [66]. The EME method is automated when performed in a multiwell-format [48]. The HF-LLME method integrates the extraction, purification and concentration steps into a single step and is performed in either a two- or three-phase mode as described in Table 1. Automation of this method is enabled by a 96-well HF-LLME system with an auto-injector integrated to an analytical platform such as high performance liquid-chromatography (HPLC) [51]. Principles of ASE and the factors that determine the efficiency of metabolite extraction are highlighted in [42]. Detailed information on the SFE's system, procedure, solvents and on-line automation with a supercritical fluid chromatography with a triple quadrupole mass spectrometer detector (SFC-QqQ-MS) are highlighted in the following studies [57,58]. The system and procedure of the MAE method is described by [57,67]. This MAE system has been automated by the incorporation of an autosampler that enables extraction sequences of up to 24 samples and thus accelerates method optimization [59,68]. The principles of the UAE method for metabolite extraction are explained in detail in the studies of [57,69,70].

In the context of plant metabolomics, only a few of these methods have been applied. For instance, the use of SPE to obtain insights into complex phenolic composition of tea and other plant samples [44], SPME to extract metabolites that reflect the changes in the 'HoneyCrisp' apples metabolome [71], extraction of phenolics with DLLME for chromatography analysis has been reported by [72], EME with liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to analyze plant hormones in citrus leaf samples [73], ASE was used in the analysis of natural products in green tea (*Camellia sinensis* L.) [55], SFE had been applied to extract tetrahydrocannabinol (THC) from *Cannabis sativa* L. [74], pharmaceutical and nutraceutical natural compounds from *Berberis* species have been extracted with MAE [75], and UAE has been applied in the extraction of phytochemical compounds from apricot by-products (pulp) [76]. These methods, including HF-LLME and SDME, have not yet been reported in metabolomics studies of plant responses to abiotic stresses. In plant metabolomics, these automated sample preparation methods allow for the extraction of metabolites from hundreds of samples with reduced extraction time and solvent consumption per sample, minimization of variation between samples, improved metabolite recovery, and at higher purity levels. Thus, such automation of sample preparation step enables consistency in sample handling, an assurance factor towards the generation of high quality and reproducible metabolomics data. The confidence in the data produced is essential for maximizing the biological output.

Robotic systems have thus been developed in addition to automated sample preparation methods for rapid throughput and reproducibility in large scale metabolomic analysis (Figure 1). These robots can be incorporated into existing automated systems to function as a transport system (i.e., sample handling) or to perform manipulation (i.e., sample preparation) tasks [77,78]. Depending on the type of task(s) performed, a single-arm or dual-arm robot is used. Single-arm robotic systems perform transportation tasks to connect different stations, whereas dual-arm robotic systems perform both transportation and manipulation tasks with the use of standard laboratory equipment and devices. In addition, dual-arm robots can function simultaneously and independently. Dual-arm robots can also transfer labware from one arm to the other without setting it down [77,79]. An example is the CSDA10F dual-arm robot (Yaskawa, Kitakyushu, Japan), that is used to perform multiple tasks including the handling of vials, solvents, pipettes and sample preparation. It is also capable of transferring and feeding vials to the autosampler of an analytical instrument. The use of the dual-arm robot improved the quality and efficiency of the analytical measurements in comparison to manual sample preparation [79]. Remote sample preparations can

thus be achieved with the application of such dual-arm robotic systems, which will save the analyst time and increase reproducibility that will lead to accurate analyses of metabolites. Thus, as one of the core steps in the (plant) metabolomics workflow, sample preparation has a tremendous impact on the final results. Hence, the minimization of errors and improved coverage and reproducibility, by incorporating automation and robotic technologies, will improve the biological insights generated from metabolomics investigations, particularly in abiotic stress studies.

2.2. Automation and Analytical Intelligence in Analytical Platforms

Robust analytical platforms are essential for accurate analyses of metabolites with good reproducibility over a period of time. Nuclear magnetic resonance (NMR) and liquid chromatography-mass spectrometry (LC-MS) are the most popular platforms used in metabolomics [29,80,81]. Automation in these analytical platforms, particularly in LC-MS, has significantly improved the efficiency of data acquisition and its reliability. Therefore, due to the preexistence of automation within these platforms, we thus refer to automation in the context of technological advancements of these platforms. In addition, automation has also reduced the risk of variability or errors related to manual operations. However, the operation of these analytical platforms is dependent on highly skilled and experienced analysts that can identify and avoid problems prior to analysis to increase the chance of producing highly reliable and robust data [78,82].

Platforms with analytical intelligence could contribute to the acquisition of highly reliable data, regardless of the user's skill(s) or experience, but still need highly skilled input to train the algorithms (Figure 1). Analytical intelligence is a concept for analytical instruments and it consists of systems and software that resemble an experienced analyst by automatically identifying good or bad conditions, displaying the results, providing feedback to the user, and solving common problems faced by an experienced analyst. For instance, column equilibration to avoid column damage, constantly monitoring the level and consumption of the mobile phase, management of column performance, manually purging the flow channels and manually picking and integrating peaks (i.e., peak detection and peak integration) are some examples of problems that challenge analyst [78,83,84]. Although different forms of 'analytical intelligence' have been part of advancements in analytical systems, the concept of 'analytical intelligence' has been recently articulated by Shimadzu [83]. This analytical intelligence means automated support functions that utilize digital technology involving IoT and artificial intelligence to enable high productivity and maximizing reliability, regardless of an operator's skill level. Analytical intelligence thus allows a system to monitor and diagnose itself, handling any issues during data acquisition without user input, and its implementation is expected to increase data reliability [83].

2.2.1. Mass Spectrometry (MS)-Based Platforms

Automated mass spectrometry (MS) systems are equipped with autosamplers and in some cases cartesian *xyz* robotics systems (i.e., robotics systems designed to permit the arms with at least three degrees of freedom) for sample preparation and injection into LC- and gas chromatography (GC)-MS instruments, thus enabling higher throughput, multiple and remote sample data acquisition [51,78]. GC-MS autosamplers are preferred over the commonly used manual injection procedure, as they provide accurate, precise, and rapid sample aliquoting for large quantities of samples with frequent wash cycles, which reduce sample carryover, defined by the user. Autosamplers thus aid in enhanced statistical analysis and, therefore, reliable biological interpretations by enabling the acquisition and analysis of multiple biological replicates, which increases the sample size for powerful statistical analysis [78,85]. Despite the advances of automated MS systems for enhancing reliable biological interpretations, these systems are, however, limited by their resolving power for isomers. Hence, automated MS systems are coupled with chromatographic separation techniques to improve the separation of isomers. This solution has proven successful; however, for complex samples such as plant extracts, additional advancements

are required to drastically improve resolution and thereby enhance metabolite coverage and biological interpretations [34,86].

2.2.1.1. Orthogonal Separations

High peak resolution, metabolite coverage and selectivity can be improved by the development of two-dimensional (2D) GC and LC coupled to MS [34,48]. In these techniques, two columns with different stationary phases, which are connected through a modulator (often involving rotary valves), are used to provide further separation of co-eluting metabolites from the initial column, thus enhancing resolution and peak capacity [34,78]. Considering that plants have complex metabolomes with many unique chemical species and overlapping peaks, the application of these techniques contributes to the advancement of plant metabolomics by enabling the detection, quantification and identification of the vast number of unknown metabolites. Additionally, this will aid in the elucidation of the plants cell and molecular mechanisms [87,88].

Ion mobility spectrometry (IMS) is another technology capable of enhancing peak resolution and selectivity with rapid separations in the millisecond range [89]. IMS devices separate ions based on their differences in mobility under the influence of an electrical field in the gas phase caused by their shape, size and charge. The measurement of the drift time (i.e., mobility) can be converted into the collision cross section (CSS), a unique physiochemical property of an ion [48,90]. IMS is divided into various technologies that are grouped into either dispersive or selective. Dispersive IMS technologies are those that enable the analysis of all ions, whereas selective IMS technologies only enable analysis of selected ions. Hence, the dispersive IMS technologies, which include drift tube IMS (DTIMS) and travelling wave IMS (TWIMS), are suitable for untargeted metabolomics. The selective IMS technologies, which include field asymmetric IMS (FAIMS) and differential mobility analyzers (DMA), are suitable for targeted studies and provide better orthogonality to conventional MS data.

Trapped IMS (TIMS) is the most recent IMS technology that is suited for targeted studies due to its high selectivity with regard to the resolving power of analytes with similar mobility; however, it can be utilized for untargeted studies by separating ions in an ion funnel/drift tube, where the ions are carried by the gas flow towards the funnel exit while the opposed electrical field pushes them back to the entrance funnel, resulting in ions with identical charge to experience the same electrical force but different dragging force from the gas flow due to different CSS. This results in better separation and peak resolution [48,91,92]. The applied field strength, gas pressure and flow in the IMS cell varies among the IMS devices [29,91]. The coupling of IMS with mass spectrometry (IMS-MS) achieves separation based on ion mobility/CSS and mass-to-charge (m/z) ratio, which enables separation of isomers and isobars, increases peak capacity, reduces chemical noise and provides structural information through the CCS measurements. IMS-MS thus further enhances resolution, selectivity, and sensitivity [89,91].

IMS can also be integrated into LC-MS systems by incorporating an IMS between the ionization source after LC and before the MS analyzer [93]. For instance, in the study of [94] that resulted in the profiling of 171 metabolites, including phenolics, flavonoids, terpenoids, lipids and nucleotides, in 30 cultivars of leaf and head type lettuces (*Lactuca sativa* L.). An untargeted screening of *Passiflora* leaf extracts with ultra-high performance liquid chromatography ion mobility collision-induced dissociation mass spectrometry (UHPLC-IM-CID-MS) has been used in the identification of flavonoid isomers, i.e., 6-C and 8-C glycosylflavone isomer pairs orientin/isorientin and vitexin/isovitexin [95]. These flavonoid isomers have been identified in plants subjected to abiotic stress. Thus, the application of IMS in plant responses to abiotic stress research has the potential to increase the identification of previously unknown detected metabolites and the resultant metabolite coverage, thereby contributing to the elucidation of plant response mechanisms to abiotic stress [96]. IMS is, however, limited by its inability to separate multiple coexisting isomers with similar drift time and biomolecules such as carbohydrates, with flexible

structure due to lack of reliable CSS calculation protocols and molecular models. Thus, inaccurate CSS calculations and incorrect structural identification may occur. Here, models based on machine and deep learning algorithms which take into consideration collision effects, the long-range interactions between drift gas and the analyte ion, may be helpful to reduce error rates by computing theoretical CSS values for biomolecules and predicting metabolite structures [97,98]. The characterization of both metabolite structures and spatial information in plants provides the opportunity to elucidate physiological mechanisms in plant organisms, hence the development and advancements of spatial metabolomics techniques [99,100].

2.2.1.2. Spatial Metabolomics: Mass Spectrometry Imaging

Current metabolomic techniques require the extraction of metabolites of interest from biological samples prior to analysis. This approach has been widely used and successful in detecting metabolomic changes in organisms; however, information concerning metabolite spatial location within organelles, cells, tissues or organs is lost in the process, thus making it difficult to interpret metabolomics data. This problem can be resolved but requires time-consuming workflows and expert analyst [101,102]. The development of spatial metabolomics has thus enabled in situ metabolomic approaches.

MS imaging (MSI) is the dominant technology in spatial metabolomics that is used to visualize 2D and 3D spatial distribution of metabolites in biological tissues [102,103]. MSI combines the sensitivity and specificity of MS with the detailed metabolite spatial information, to produce mass spectrums/images representing relative ion intensities at specific tissue localizations [101,102]. MSI platforms are comprised of three components, an ionization source, a mass analyzer and an ion detector. The ionization source extracts and ionizes metabolites simultaneously from the surface of the sample at particular x - and y -coordinates. The ionized metabolites are separated in the mass analyzer according to their m/z and their abundance are detected and recorded by the ion detector. Software is used to produce images, also referred to as intensity maps, from the spatially resolved MS data in which each pixel (i.e., peak) is composed of a mass spectrum. In these images, the spatial coordinates are represented by the x - and y -coordinates whereas the signal intensity is represented by the color intensity scale [101,104].

MSI platforms are categorized by their ionization source and mass analyzer, and they vary in speed, sensitivity, and spatial resolution. Matrix-assisted laser desorption/ionization (MALDI) is the dominant platform used in MSI and it extracts metabolites and simultaneously ionizes them with the use of a laser (Figure 2). Emergence of other platforms such as desorption electrospray ionization (DESI), laser ablation electrospray ionization (LAESI), nanostructure-initiator MS (NIMS), nanospray desorption electrospray ionization (nanoDESI), infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI) and secondary ion MS (SIMS), which utilizes an ion beam, increases the development and use of the MSI technology in spatial metabolomics [102,103].

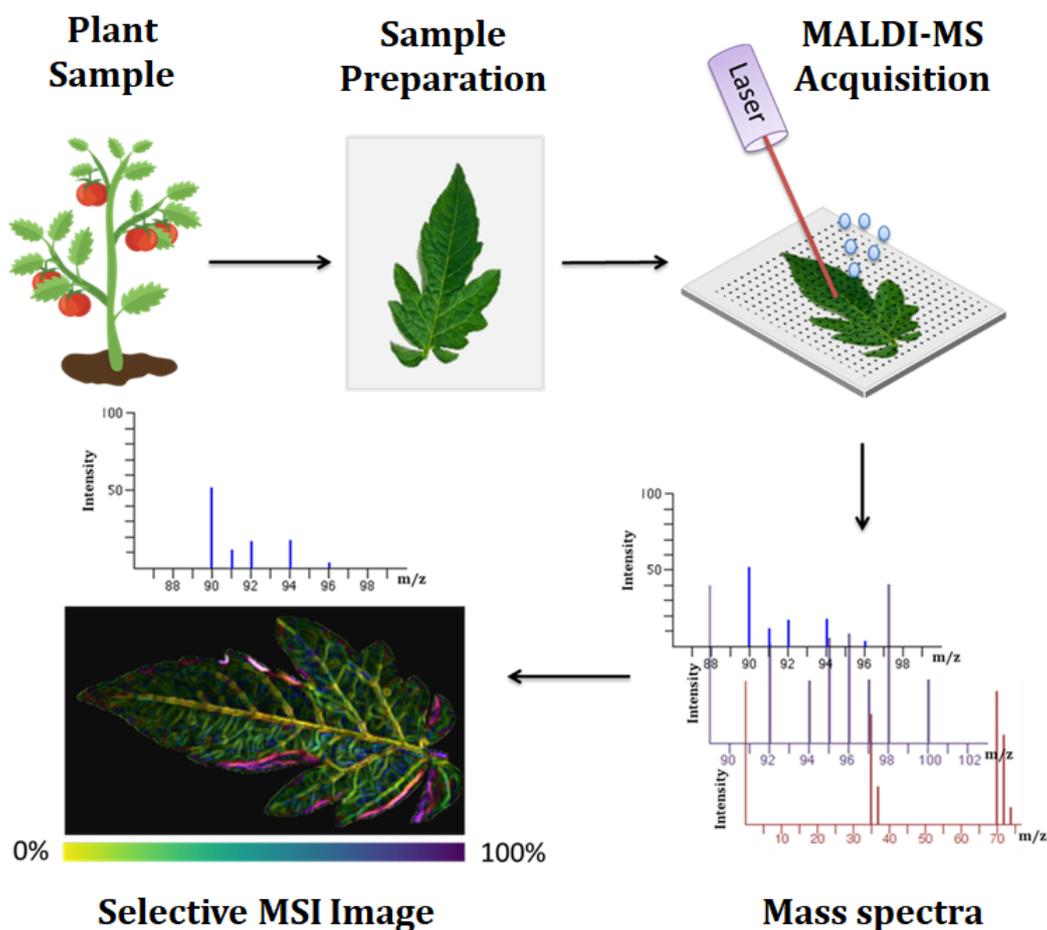


Figure 2. A general overview of a matrix-assisted laser desorption/ionization-mass spectrometry imaging (MALDI-MSI) workflow. A tissue (leaf) is sectioned from the sample (tomato plant) and mounted on a target surface. A matrix is applied to the tissue and a laser beam extracts metabolites into a mass analyzer. Mass spectra are generated at each x,y coordinates across the tissue surface. A software algorithm combines and processes the mass spectra and generates a MSI image.

The detection of many metabolites from a single living cell within 10 min, is MSI greatest advantage over current metabolomic techniques. Hence, MSI is increasingly being applied in a variety of biological and clinical studies, as it enables simultaneous measurements of vast amount of metabolites and their spatial distribution from a variety of samples [105,106]. MALDI-MSI has been successful in imaging various classes of endogenous molecules on the surface of different plant organ tissues, including the root, stem, leaf, flower, fruit and seed, to provide new insight into the molecular in situ analysis through the quantitative changes in molecules during plant growth and development or those induced by environmental stresses [100]. The application of MALDI-MSI with on-tissue chemical derivatization of maize root and leaf, has enabled the identification of 600 metabolites [107]. MSI of *Lychnophora salicifolia* leaves with laser desorption ionization (LDI) and MS/MS, revealed the accumulation of vicenin-2, a di-C-glycosyl flavonoid, in the mesophyll cells and upper epidermis, in response to extreme sunlight [108]. Spatial metabolomics has also been used to explore plant specialized metabolite diversity in the plant genus *Euphorbia L.*, and the diterpenoids were shown to be localized to young stems and roots, thereby suggesting they play a role in defense responses to biotic stresses [109].

The widespread application of spatial metabolomics is likely to benefit from the incorporation of machine and deep learning techniques, as well as cloud computing. For instance, the development of novel mass spectrometry imaging software, in addition to several existing examples of software such as MSiReader [110], with ML capabilities to predict mass accuracy of collected data, will enable reduction in molecular ambiguity,

enhanced data quality and interpretability. Machine learning dimension reduction methods, such as *t*-distributed stochastic neighbor embedding (*t*-SNE) and uniform manifold approximation and projection (UMAP), enable visualization of the large complex image spectral data that will aid in identifying similar pixel or image clusters and data interpretation [102,111]. Cloud computing platforms such as METASPACE and OpenMSI, have enabled the construction of imaging MS libraries that contribute to rapid and accurate metabolite identification by resolving ionization pathways and integrating all signals corresponding to a particular metabolite, and data analysis, thus increasing the widespread use of MSI [102]. The application of spatial metabolomics thus has the potential to further elucidate plant response mechanisms to abiotic stresses, as MSI enables rapid metabolite detection and identification, particularly of those lost or degraded in the extraction process, and the discovery of novel metabolites, that will provide insights of the metabolome changes in response to the abiotic stresses, within the organelles, cells, tissues or organs. Additional technologies for high throughput and automated analysis have been developed.

2.2.1.3. Lab-On-Chip and Microfluidic Devices

The emergence of microfluidic or lab-on-chip technology enables the integration of multiple analytical processes onto a single platform. The technology is based on the fabrication of microdevices with small channels and chambers that enables the manipulation of fluid volumes in the micrometer range and parallel workflows for automated and high throughput analysis [112,113]. The integrated processes in microfluidic devices include sample preparation, preconcentration, separation and delivery to analytical platforms [48,112]. Microfluidic platforms are divided into three categories: analog, droplet and digital microfluidics (DMF). Analog and droplet based microfluidic platforms utilize fluid shear stress between an oil phase and an aqueous phase, to generate continuous fluid streams and separate droplets within enclosed microchannels, using passive and active pumping mechanisms [114,115]. DMF platforms utilize electrodes, coated with a hydrophobic material layer, to move, mix, merge and separate samples when a voltage potential is applied [112,114]. The handling of small fluid volumes with low flow rates make microfluidic devices suitable for coupling with MS. Soft ionization techniques are usually employed for the ionization of delicate biomolecules, molecules that are sensitive to heat and are usually large polar organic molecules [78,116].

The coupling of microfluidic devices with electrospray ionization-mass spectrometry (ESI-MS) is dependent on emitters that are either integrated or external to the device [112,114]. Description of the emitters are well documented in [112]. MS analysis in droplet-based microfluidic platforms is performed by either extracting the droplets from the multiphase flow or by directly delivery of droplets to the MS without extraction. The coupling of DMF devices to ESI-MS requires the transfer of droplets from the device to the ESI emitters and the dissociation of the droplet manipulation and spray voltages. This is achieved by using sandwiched capillary emitters and pump based sampling capillaries connected to external emitters for closed and open systems, respectively [112,114]. Generally microfluidic platforms are coupled offline with MALDI-MS by either spotting samples onto MALDI plates or by direct MALDI analysis of the fabricated material of the microfluidic device. Online coupling has, however, been established [78,114].

Analog platforms utilize droplet ejectors such as embedded capillaries or integrated piezoelectric microdispenser for target spotting and they can be programmed for automated positioning. Droplet and DMF platforms can be coupled online as they enable contactless spotting of samples by exploiting their segmented flow. The precise handling of small fluids and the automation of sample preparation, concentration and separation provided by microfluidic platforms has improved MS sensitivity and reproducibility. Microfluidics can also increase separation times as the distance or steps between the separation and MS detector is reduced [48,114]. The microfluidic device, RootChip, has been applied for measuring changes in metabolite concentrations, such as calcium and phytohormones, and revealing the specific signal transduction responses in *Arabidopsis thaliana* roots treated

with biotic (flg22) and abiotic (sodium chloride (NaCl)) elicitors [117]. The RootChip allows the simultaneous stimulation of a single organ; in this case, the treatments were applied on one side of the root, while the cell responses are studied on the untreated side. The RootChip device can thus enable the testing of combined treatments without exposing the cells to the individual treatments. Microfluidics in plant–environment interaction studies can, therefore, enable the elucidation of environmental stress sensing, signaling, and intercellular and molecular mechanisms [117]. Microfluidics can deliver fluid droplets directly into MS analytical systems to increase throughput analysis and data reliability. The improvement of data reliability and accurate mass detection through the development of novel data acquisition workflows/algorithms is an active research area.

2.2.1.4. Virtual Metabolomics Mass Spectrometer

Emerging analytical workflows for MS and MS/MS spectral data acquisition are bound to improve accurate mass detection and thereby enhance metabolite identification. The coupling of chromatographic separation with MS/MS is the widely used and one of the most powerful approaches used for metabolite identification. MS/MS spectra are obtained by either data-dependent acquisition (DDA) or data-independent acquisition (DIA). In the DDA workflow, the MS instrument changes from a full scan MS to MS/MS when precursor ions overcome a predefined threshold of intensity or other criteria such as isotope pattern, mass defect, and the presence of diagnostic ion or characteristic neutral loss, thus resulting in poorer sampling of lower abundant precursor ions. DIA workflows, such as MS^E, involve alternating scans acquired at low or high collision energy to obtain full scan precursor ions, accurate mass fragments and neutral loss information. However, matching the precursor ions to product ions when there is substantial chromatographic co-elution or overlap is challenging. Algorithms are thus employed to intelligently select precursor ions, thus improving the efficiency of MS/MS spectral acquisition [29,118,119].

Recent software such as Virtual Metabolomics Mass Spectrometer (ViMMS) enables the prototyping, implementation, optimization and validation of new acquisition methods *in silico* [119,120]. These algorithms improve MS/MS acquisition by determining precursor cleanliness or by reducing irrelevant spectra acquisition and automatically adjusting to various chromatographic conditions to provide improved reproducibility [118,121]. The intelligent automation of precursor ion selection reduces the need for highly experienced or skilled analyst and makes the successive metabolite identification process less challenging [78,121]. Such an efficiency gain could also spark a renewed interest in LC-MSⁿ-based approaches where spectral trees are formed to gain deeper insight in the fragmentation pathways involved in the breaking apart of the metabolite structures, thereby facilitating metabolite annotation as shown for polyphenols standards and flavonoids in plant extracts [122,123]. In addition to the aforementioned technologies in MS-based platforms, advances in NMR for improved metabolite resolution and identification have been made [124].

2.2.2. Nuclear Magnetic Resonance (NMR)-Based Platforms

NMR is popular for its high reproducibility; however, in contrast to MS, it has poorer sensitivity, lower dynamic range and resolving power and, therefore, it typically results in a restricted metabolite coverage in plant metabolomic studies. NMR technology has recently been improved by the development of miniaturized radiofrequency coils, superconducting magnets, cryogenic probes and multidimensional NMR techniques [18,125]. These developments have thus enabled the improvement of resolution, acquisition time, multi-nuclei detection and performance of magnets at temperatures close to 4 Kelvin (K), due to using cryocoolers or closed cycle helium (He) refrigerators that automatically recycle He, thus eliminating the need for constant liquid helium (He) or liquid nitrogen (N₂) refills [125,126]. The study of metabolites by NMR was enabled by the ¹³C optimized cryogenic probes. The development of a ¹H-¹³C dual-optimized probes enables 2D NMR experiments at natural abundance [125].

The application of 2D NMR improves NMR sensitivity by reducing resonance overlap and provides information on chemical bonding between nuclei. However, 2D NMR experiments are time consuming and therefore challenging to use in large-scale metabolomic studies [29]. Acquisition time in 2D NMR experiments can be increased by various approaches. One approach is the use of diffusion and/or relaxation filters that reduce the interscan delay (i.e., relaxation time). Non-uniform sampling (NUS) is another approach that entails the recording of randomized data points instead of all data points. Ultrafast (UF) 2D NMR is another approach that enables the acquisition of a 2D spectrum in a single scan. These approaches are described in more detail in [34,125]. Metabolite fingerprinting of *Bougainvillea spectabilis* leaves exposed to high levels of air pollution was achieved with 2D NMR, and revealed the differential concentrations of amino acids, sugars, krebs cycle intermediates, phenylpropanoids, flavonoids, and production of putrescine, gamma-aminobutyric acid (GABA) and trigonelline metabolites in response to the air pollution stress [127]. Additional technologies have also been developed to enhance NMR sensitivity of which some are highlighted in the next paragraphs.

Dynamic nuclear polarization (DNP) is another advancement that enhances NMR sensitivity. DNP enhances the NMR signal by transferring the polarization of the electron spins to the nuclei prior to delivering the sample to the NMR spectrometer. DNP thus allows the detection of metabolites, particularly natural products, present in low concentration in complex mixtures [34,126]. DNP has been applied for the abundant detection of ^{13}C nuclei in the metabolomic analysis of tomato extracts [128]. DNP has the potential of increasing metabolite coverage to facilitate the use of NMR in untargeted studies of plant responses to environmental conditions.

Additionally, the detection of metabolites within intact tissue samples has been made possible by the advancements in high-resolution magic-angle spinning NMR spectroscopy (HRMAS) [126]. The development of microprobes for magic-angle spinning (μMAS) of sub-microgram specimens with high-resolution in HRMAS, allows site-specific metabolomic characterization of varying plant and tissue regions without the need for sample preparation. Therefore, HRMAS provides the opportunity to explore the specific tissue metabolism with more or less the same precision as MSI [126]. HRMAS has been applied to monitor plant responses to abiotic stresses. For instance, the study of soybean leaves grown with and without water-deficiency stress with HRMAS, revealed differences in the metabolite profiles. A total of 30 metabolites were identified, with the amino acid metabolites present in the metabolite profile of soybean leaves with water-deficient stress [129]. HRMAS can therefore aid in elucidating plant response mechanisms. The identification of novel metabolites will greatly contribute to the discovery of unidentified pathways and thus the elucidation of plant response mechanisms.

The de novo identification of novel metabolites from complex mixtures with NMR can be enhanced with on-line hyphenation of separation techniques. A flow cell matching the high performance liquid-chromatography (HPLC) module is used to connect the NMR, instead of a probe, thus resulting in HPLC-NMR. In HPLC-NMR, the separated analyte is eluted from the column to the flow cell and subsequently to the NMR spectrophotometer [34]. High performance liquid-chromatography-solid phase extraction-NMR (HPLC-SPE-NMR) is another hyphenated NMR technique that enriches the analyte by allowing the removal of the HPLC mobile phase, prior to NMR data acquisition. HPLC-SPE-NMR decreases the analyte chromatographic peak volume and, therefore, achieves greater sensitivity compared to HPLC-NMR. Additionally, HPLC-SPE-NMR has been enhanced with the integration of a photodiode array detector (PDA) and MS for authentication and the structural characterization of secondary metabolites from complex plant extracts [34,130]. For instance, HPLC-PDA-HRMS-SPE-NMR analysis of *Coleonema album* leaves enabled the structural identification of 23 coumarins [131]. An automated HPLC-MS-SPE-NMR approach was used to characterize flavonoid structures in crude tomato extracts [130]. Despite the accomplishments of these technologies in improving NMR sensitivity, resolution

and acquisition time, advances in software tools and databases are still required to aid in metabolite analysis [125].

Novel software tools consisting of machine and deep learning algorithms have been developed to provide accurate prediction of chemical shifts and automation to manual phasing, water removal, baseline correction, peak picking and spectral fitting methods, thereby improving accurate quantification of metabolites for reliable structural elucidation. These automatic deconvolution algorithms include Bayesian Automated Metabolite Analyser for NMR spectra (BATMAN) [132], Bayesil [125], AQUA [133], MagMet [125] and rDolphin [134], and they contribute to making NMR faster, consistent and user-friendly. The integration of this software with online websites/databases such as NMRShiftDB further enhances NMR accuracy and molecular coverage with increasing data, thus enhancing structural discoveries and elucidation [84,125,135]. Hence, recent efforts have been made to make NMR data publicly available through the development of software tools and databases for the discovery and structural elucidation of natural products [136].

Small Molecules Accurate Recognition Technology (SMART), a machine learning based tool, for example, enables the interpretation of 2D-NMR spectra and the acceleration of discovering and characterizing novel natural products [137]. Additional examples include the ImatraNMR [138] and SimpeleNMR [138] tools for quantitative NMR analysis. ImatraNMR and SimpeleNMR tools were applied in the 2D NMR analysis of 4 *Arabidopsis thaliana* strains under controlled and ozone exposure conditions. Metabolite differences were revealed, with GABA found to be produced in the plants responses to the ozone exposure (i.e., changes in oxidative stress) [138]. These 4IR technological and software developments in NMR will, therefore, enable rapid analysis for large datasets and potentially increase the widespread use of NMR-based plant metabolomics [139]. In this review, we mainly focus on LC-MS and LC-MS/MS platforms and their data processing and analysis workflows.

2.3. Machine Learning Methods for Metabolomic Data Mining and Interpretation

Axiomatically, metabolomics studies, particularly untargeted approaches, generate complex, with inherent covariance structure, and information-rich big data sets that are challenging to handle and interrogate. With technological advancements in analytical platforms (Section 2.2) and an increasing complexity in the study design, untargeted metabolomics data are increasingly becoming more heterogeneous, big data in terms of volume, velocity and variety [5,140]. Thus, the availability of big data in (plant) metabolomics reflects a new era of data-driven research employing powerful computational tools, involving AI and machine learning algorithms, to maximally mine and find novel knowledge in this (big) data [5,141]. The typical approach used in mining and interpreting untargeted metabolomics data is well-described in literature; it is a multistep workflow, employing dedicated mathematical modeling and chemometric algorithms (Figure 3) [26,142,143].

The philosophies of data science—the extracting of information from (big) data—have recently undergone an *aggiornamento*, with a resurgence in interests in linear and nonlinear machine learning (ML) methods [144–146]. Furthermore, historically, nonlinear ML methods have not been widely used due to difficulty in deriving statistical inference, and thus biological interpretation. However, with increasing appreciation of nonlinearity in biological (big) data, nonlinear ML algorithms are increasingly being explored and applied in interrogating data, in life sciences [102,147–150]. ML is a subdomain of AI that provides machines with the ability to learn directly from data and past experiences through computational algorithms to facilitate better informed decisions or actions, without the involvement of experts. ML methods are categorized as either supervised-, unsupervised-, or reinforcement learning algorithms [151]. The algorithms can be further categorized based on their learning techniques, which are classification, regression, clustering, and dimensionality reduction (Figure 3) [151,152].

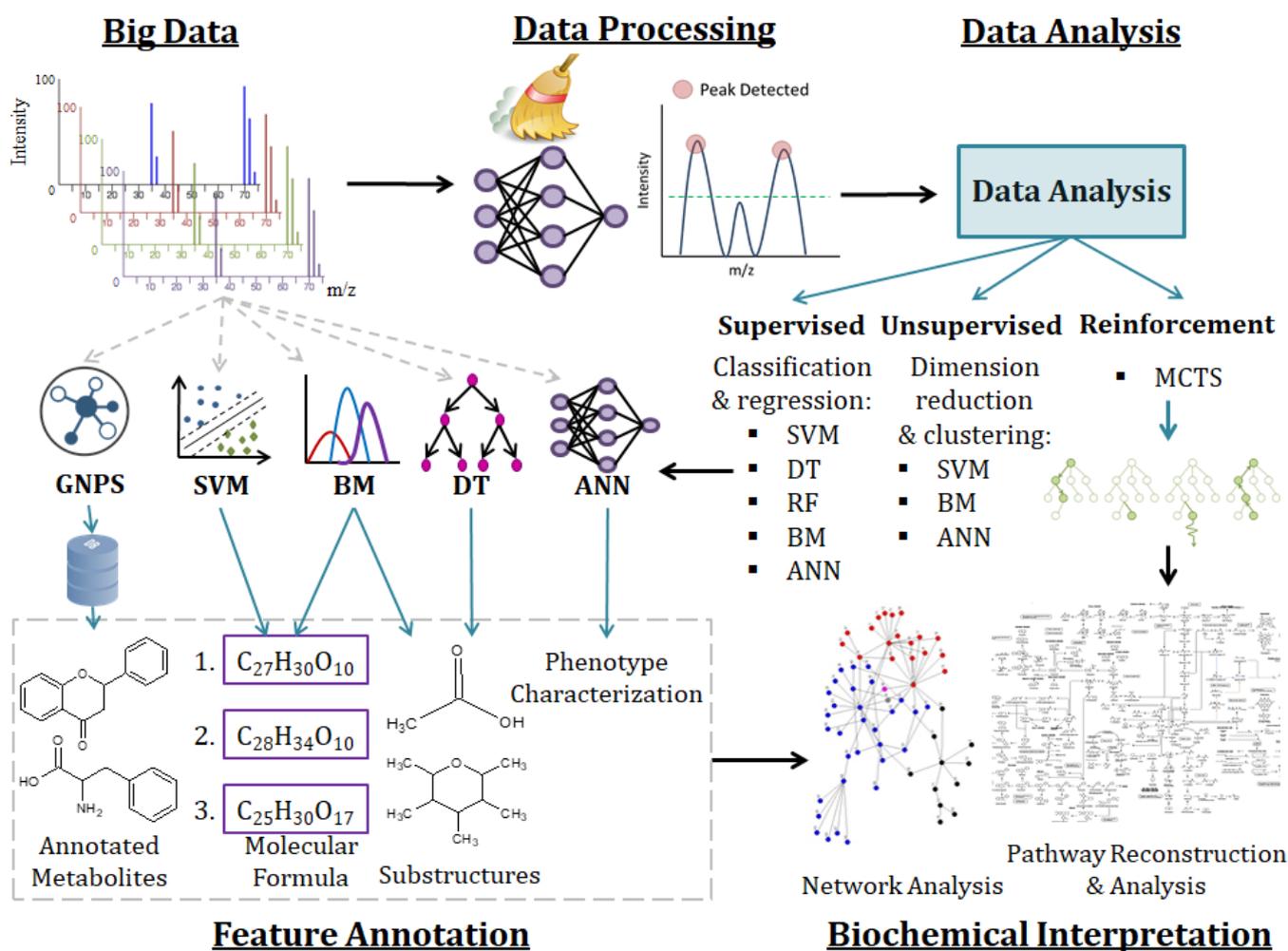


Figure 3. Workflow for metabolomics data analysis and biochemical interpretation with machine and deep learning algorithms. Various machine and deep learning algorithms are implemented all through the analysis and interpretation workflow. Artificial neural networks (ANNs) implemented in data processing remove noise and artifacts from raw data by performing peak detection and alignment. In data analysis, ANNs, support vector machines (SVMs) and Bayesian models (BMs) may be applied as either unsupervised or supervised methods and, in the former, they explore and reduce the dimensionality of the data. The supervised analysis with SVMs, decision trees (DTs), random forest (RF), BMs and ANNs, classify and extracts/selects relevant features from the dataset for feature annotation. Selected features from data analysis are used as input data for GNPS, SVMs, BMs, DTs and ANNs. Raw spectral data may also be used as input data. GNPS queries the data against its spectral library to annotate the selected features. SVMs and BMs predict and rank the features molecular formula. Additional BMs and DTs extract substructures from the data. ANNs characterize phenotypes from the data. The annotated features are integrated and applied for network analysis, pathway reconstruction and analysis. Abbreviations: ANN, Artificial neural network; SVM, Support vector machine; BM, Bayesian model; DT, Decision tree; RF, Random forest; MCTS, Monte Carlo tree search; GNPS, Global Natural Product Molecular Networking.

In plant metabolomics, several ML-based methods (Figure 3) are increasingly being explored to attain a comprehensive insight and understanding of the plant biological systems under different conditions (Table 2). Deep learning (DL) algorithms and artificial neural networks (ANNs) have been employed in biomarker discovery from natural products/herbal medicines [153,154]. Additionally, [155] explored ANN and DL algorithms in their studies to reveal the diversity of specialised metabolome in different *Camelina sativa* varieties. Deep convolutional neural networks (DCNN) have also been used as an alternative of ANNs in the mass spectrometry imaging-based detection of specific secondary metabolites in tomato [156]. Support vector machine (SVM) has been applied in plant metabolomics research for the classification, identification and predictive power of metabolite content in

exploring the medicinal properties of plants [148,157]. Random forests (RF) have also been used in optimization of metabolic fingerprinting and metabolite detection studies [158,159]. Other ML-based methods such as the nearest neighbour, Naive Bayes and decision trees have also been used in biological pathway predictions for a deeper understanding of metabolic pathways and networks in plants for crop improvement [160,161].

However, in metabolomics studies that investigate plant responses to abiotic stresses, the exploration and application of ML methods is still to be encouraged. The integration and use of these ML and deep learning (DL) algorithms in the metabolomics data analysis pipeline hold the promise of transforming the future of metabolomics research. The ML and DL methods offer the flexibility to effectively analyze and integrate a large volume of multi-omics data (e.g., metabolomics, transcriptomics, genomics, and proteomics) given a large enough sample size and enough biological context [162,163], thus providing better predictive models. This would advance the understanding of regulatory network rules and mechanistic events at cellular and chemical space of the plant, illuminating its responses to abiotic stresses. Thus, the paragraphs below dive into describing some of the ML methods, pointing out their applications in metabolomics studies to illustrate their potentials to support metabolomics analysis workflows and facilitate possible integration with other omics data types.

2.3.1. Support Vector Machines

Support vector machines (SVMs) are classification and regression ML algorithms that separate data into two classes [165,166]. SVM algorithms map samples as data points from various classes in a high-dimensional feature space. The SVM then constructs an optimal hyperplane that maximizes the distance to the nearest point of each class (i.e., margin) [146,167]. Maximizing the margin enables the SVMs to correctly classify new data points that lie within the margin on either side of the hyperplane [146]. SVMs have been reported to outperform PLS-discriminant analysis (PLS-DA) for feature extraction and classification accuracy of metabolomics data [168].

The support vector machine recursive feature elimination (SVM-RFE) algorithm has proven useful in identifying metabolic biomarkers in untargeted LC-MS datasets. This iterative algorithm ranks metabolites according to their ability in discriminating the two phenotypes/conditions (i.e., control and treated samples). Features most informative of discriminating phenotype or condition are ranked higher than those less informative, thus this ranking system is ideal for biomarker discovery as it ranks based on predictive accuracy [167]. The clustered support vector machine (CSVM), a clustering ML algorithm, divides input data into multiple clusters and trains a SVM within each cluster to separate the data in the clustered feature space, thus reducing the number of relevant features [169,170]. Furthermore, SVMs also play a role in metabolite annotation workflows, as in the SIRIUS 4, a tool for molecular structure identification, an SVM is used to aid in the determination of a molecular formula for a candidate molecule (Figure 3) [171].

SVMs can thus facilitate the visualization of relationships between plant samples, their classification and discovery of contributing markers or natural products that correlate to biological origin, such as abiotic stress, with strong confidence [148,172]. SVMs, as with the classical supervised statistical methods, are, however, widely used for the discrimination of two different classes and, due to their limitations, they are rarely used for multi-class analysis. Hence, ML algorithms such as decision trees (DTs) have been developed and applied for the analysis of plant complex multifactorial characteristics [167,173,174].

Table 2. Selected recent metabolomics studies showcasing 4IR technologies, particularly the use of ML methods in plant research.

Metabolomics Study	ML Method ¹	Reference
ML-modelling for prediction of metabolic pathways of plant enzymes.	SVM, ANN, NB, DTC	[160]
Identification of central and predictive molecular components of plant metabolic stress response.	SVMs, NNC, DTC	[157]
Untargeted metabolomics to reveal diversity of the metabolome in seeds of <i>Camelina</i> species.	DL, ANN	[155]
Characterisation of adaptive and signalling responses based on metabolite content under abiotic stresses.	SVM, DPCLus	[148]
Detection of aflatoxin metabolite in chilli pepper using machine vision.	SVM, RF	[158]
Detection and resolution on plant metabolites (<i>S.lycopersium</i>) using mass spectrometry imaging.	DCNN	[156]
Discovery of Q-markers from Jinqi Jiangtang for medicinal purposes.	ANN	[154]
Prediction of metabolic pathways in correlation networks in the pericarp of a tomato.	ML algorithms	[161]
Discovery and identification of biomarkers using ML algorithms in metabolomic studies.	ANN, DL	[153]
Enhancement of plant metabolite fingerprinting using ML methods.	SVM, RF	[164]

¹ SVM, Support Vector Machines; ANN, Artificial Neural Networks; NB, Naïve Bayes; NNC, Nearest Neighbour Classifiers; DTC, Decision Tree Classifiers; DL, Deep Learning; RF, Random Forest; DCNN, Deep Convolution Neural Networks.

2.3.2. Decision Trees

Decision trees (DTs) are ML algorithms that perform classification and/or regression of both categorical and continuous input and output variables in a tree-like structure composed of a root node, internal nodes and leaf nodes. A DT organizes or separates dataset (i.e., input at the root node) into smaller homogenous subsets or sub-populations (i.e., output at the leaf nodes) based on the most significant classifier among the independent variables [175,176]. The internal nodes represent the values of the attributes, whereas the leaf nodes represent the final decisions or predictions and the label of the class after following the path from the root to the leaf nodes [166].

In plant metabolomics, DTs have been applied in the extraction of metabolite features that discriminate different rice cultivars [173], in the identification of 11 anti-inflammatory biomarkers from 57 *Asteraceae* species leaf extracts [153] and in the prediction of the most frequent substructures based on the mass spectral features and retention index (Figure 3) [177]. A DT based on the fragmentation patterns of metabolites approach was used for the annotation of 28 unique avenanthramides in oat seedling extracts [178]. Additionally, DTs have also been used in predicting plant responses to drought, salinity, heat and cold stress [174]. DTs are, however, underexplored in plant metabolomics and their applications could provide the opportunity for multifactorial plant characteristic analysis, such as the investigation of plant response mechanisms to multiple stresses, and identification of discriminant metabolite biomarkers that correlate to the abiotic stress within the complex metabolomics data [179]. A general advantage of DTs is their interpretability as compared to many other classifiers such as random forest (RF) discussed in the next section. However, the comprehensibility of DTs is at the expense of lower predictive accuracy compared to other classifiers such as SVMs. Fortunately, the classification accuracy of DTs can be improved by ensemble learning (EL) methods [180].

2.3.3. Ensemble Learning

Ensemble learning (EL) methods are used to enhance the predictive performance of statistical learning and model fitting techniques by constructing a linear combination of a base learning algorithm. Common EL methods include random forest (RF) and bootstrap aggregating or bagging algorithms [175,176]. RF is a decision tree-based classification algorithm that is capable of handling high-dimensional data with many features, noise, unbalanced and missing values. RF aggregates multiple decision trees through bagging and bootstrapping methods and classifies new data based on the consensus of the classification trees [166,181]. RF, unlike DT algorithms, separates a sample of features for each separation instead of trying to separate all the features, thus reducing the variance and improve predictive accuracy. Bagging is an approach that repeatedly extracts samples from the same set of data, and with the use of bootstrapping methods, the predictability of this approach for classification is maximized through the direct log-likelihood optimization [166,176]. RFs, in addition to SVMs, are the widely applied ML algorithms in life sciences [182].

In omics sciences, RF has aided in extracting biological insights by detecting genotype from environment interactions [183]. RFs are, however, relatively new to metabolomics studies and are proving to be powerful classifiers due to their high classification accuracy, ability to determine the variable/feature importance (i.e., identifying variables that contribute to the prediction results) (Figure 3), avoidance of overfitting and tolerance to outliers and missing values [184–186]. Additionally, RFs have been reported to perform better on metabolomics data for phenotypic classification and feature extraction, as compared to other classification methods. For instance, RF with 100% accuracy outperformed the SVM, PLS and linear discriminant analysis (LDA) classifiers in the analysis of clinical metabolomic data of healthy and colorectal cancer diagnosed patients' urinary samples [185,187].

Considering clinical metabolomic data are relatively similar, in terms of complexity and size, to plant metabolomic data, the application of RFs in plant metabolomics, particularly for plant-environmental data, thus offers great potential to accurately extract features

that discriminate the plant phenotypic responses to environmental conditions such as abiotic stresses, and therefore improving data interpretation. RFs are, however, disadvantaged by their complex visualization and lack of statistical significance measurements such as a p-value. RFs thus provide a list of the most important metabolites without a cut-off value to select any “significant metabolites”. The feature extraction process for RFs is, however, not reliable in situations where variables differ in their scale of measurement or number of categories [168,188]. Hence, alternative approaches, such as Bayesian models (BMs), with enhanced feature extraction and high reproducibility of metabolomic data, have been considered for such situations [189,190].

2.3.4. Bayesian Models

Bayesian models (BMs) are a group of probabilistic models that conduct their analyses based on the Bayes’ theorem, which calculates the probability using the previously obtained probability and information of the data collected. BMs are employed for either classification or regression problems and clustering [175,176,191]. BMs modify probability distribution to identify possible concepts by assuming independence between the variables and calculating the conditional probability for each instance based on the assumed classes [166,176]. BMs are promising for metabolomic analysis as they allow the incorporation of prior knowledge with experimental data to facilitate better predictions, capture relationships in non-linear interactions between metabolites and phenotypes, reduce model overfitting, and identify significant individual metabolites [189,190].

BMs applied in metabolomics have resulted in confident/reliable compound annotations through Bayesian-based tools such as the Integrated Probabilistic Annotation (IPA) method. The IPA method incorporates multiple sources of information within the annotation process to increase the predictive power of assigning measured m/z values to putative formulas (Figure 3). Additionally, the confidence in annotations are quantified and re-evaluated when new information is provided, thus improving annotations, particularly for data obtained from similar biological samples using the same experimental procedure [192]. MS2LDA, a Bayesian Latent Dirichlet Allocation (LDA) model, extracts peak patterns in LC-MS/MS data that represent molecular substructures (Figure 3) to enable the grouping of molecules based on shared substructures regardless of overall spectral similarity, thus enhancing the extraction of plant metabolite building blocks that facilitate metabolite annotation and that could be linked to phenotypic differences through differential analysis [8]. SIRIUS 4.0, a computational tool for metabolite identification from MS/MS data, implements a Bayesian network scoring that enables it to identify the molecular formula and structure of a query compound with high accuracy (Figure 3), thus resulting in enhanced and more confident compound identification [171,193].

The identification of molecular formulas and structures for large compounds (i.e., >500 Da) is, however, still challenging, particularly for plant extracts that contain large conjugated compounds above 700 Da. Organic compound Determination by Integral Assignment of elemental Compositions (ZODIAC), a network-based algorithm for estimating molecular formula, is an alternative approach that combats this problem. ZODIAC uses Bayesian statistics to re-rank SIRIUS molecular formula candidates. The ZODIAC score increases the confidence in formula annotation, particularly for large compounds, and reduces the error rates. Analysis of tomato (*Solanum lycopersicum*) extracts with ZODIAC resulted in the decrease in error rates from 4.44% to 2.22% and the discovery of three novel molecular formulas not found in any structural databases [194]. These powerful Bayesian-based tools can, therefore, contribute to the elucidation of plant response mechanisms to abiotic stress by facilitating the accurate extraction of structural properties and identification of novel metabolites involved in the plants response to abiotic stress. BMs can, however, be complex, difficult to implement, and computationally quite expensive, in terms of the power requirements. Hence, alternative methods are also considered such as artificial neural networks (ANNs) [195,196].

2.3.5. Artificial Neural Networks

Despite the considerable efforts of the aforementioned ML methods to rapidly and accurately characterize non-linear complex samples and their metabolite properties, false positive signals, co-eluting metabolites and retention time shifts are still major bottlenecks that effect data analysis and interpretation of plant MS-based metabolomics. Artificial neural networks (ANNs) and deep learning (DL) methods are proposed to solve these issues and other bottlenecks involved in the mining of metabolomics data [154,197].

ANNs are interconnected information processing systems that resemble the human's nervous system. ANNs can be used for classification and/or regression of non-linear systems and are characterized by their architecture, patterns of interconnected processing units (i.e., neurons), method of determining the weights on the connections (i.e., learning algorithm) and their activation function [176,198]. The general ANN architecture consists of an input layer, where the data is introduced into the system, hidden layer(s), where learning occurs and the information obtained is linked to the output layer that provides the resultant decision/prediction. The connections are based on the weight values defined during the training process and, therefore, the output values will be very close to those defined in the training model [166,175]. ANNs can be divided into two categories: conventional ANNs and deep ANNs, also referred to as DL or deep neural networks (DNNs). DNNs have multiple processing layers (i.e., multiple hidden layers) that enable them to learn and fit raw data through representation at multiple levels of hidden layers. Hence, more and improved representation of observed patterns in the upper or upstream layers is achieved [162,175]. DNNs are advantageous in their ability to extract features automatically, thus eliminating potential bias, and to function as either a supervised or unsupervised method. The commonly used DNN is the convolution neural network (CNN). Alternative DNNs include unsupervised pretrained networks, recurrent neural network (RNN) and recursive neural network [175,197,199]. DNN and ANN models have been widely applied in the data processing and data interpretation steps of metabolomic pipelines (Figure 3) [197,200].

An optimized back-propagation ANN (BP-ANN) was applied in the classification and discovery of quality markers with bioactivity from the Jinqi Jiangtang (JQJT), a Chinese herbal medicine, metabolomic dataset. The BP-ANN outperformed the PLS in terms of accuracy and low error rate [154]. DNNs reported in metabolomics studies include a CNN model, peakonly [201], for peak detection and integration. The peakonly algorithm classifies LC-MS data into regions of noise, chemical peaks and uncertain peaks, thereby eliminating false noise peaks and determines peak boundaries for integration. DeepSpectra [202], a CNN-based model, enables phenotype characterization directly from raw spectral data. DeepSpectra has been applied in the prediction of biomass and protein content in environmental datasets. ChromAlignNet, a RNN-based model, for peak alignment of GC-MS data enables the alignment of complex data without the requirement of additional parameter selection and reference chromatograms [197,200]. Furthermore, DNNs have also been used to predict collision cross section (CSS) values, properties of ion-mobility MS that can be used to eliminate the time-consuming search of unknown metabolite identification by matching the CSS values with metabolites in spectral databases, which will also increase confidence in identification [197]. ANNs are showing great promise in detecting plant diseases and discriminating between healthy and diseased plants through hyperspectral imaging data. The advancement of these models in hyperspectral imaging could translate to the development of ANNs for mass spectrometry imaging (MSI) [203].

Altogether, plant metabolomic studies are likely to benefit from such versatile ANN and DL models, due to their ability to encode and model complex non-linear datasets. For instance, in the case of investigating plant response mechanisms to abiotic stresses, ANN and DL models will facilitate classification and extraction of relevant features with high accuracy from raw data, due to their ability to eliminate noise and false signals, and separate co-eluting metabolites. Additionally, their capability in predicting the presence/absence of substructures with high accuracy based on the spectral information and molecular formulas,

will facilitate enhanced metabolite identification and discovery of novel metabolites that will aid in elucidating plant response mechanisms to abiotic stress and the modeling of active pathways [197].

2.3.6. Machine Learning for Pathway Modeling

The identification and understanding of metabolic pathways is important in elucidating the underlying biological processes/mechanisms, such as the causes of metabolite diversity, of an organism. Substantial knowledge on metabolic pathways, particularly plant pathways, have been elucidated and curated in various databases such as Kyoto Gene and Genome Encyclopedia (KEGG); however, there still remain unidentified metabolic pathways with their associated gene–enzyme–metabolite relationships [204,205]. The reconstruction of metabolic pathways assists in the identification of unidentified pathways. This process is complex and widely based on hypothesis-driven models (i.e., constraint-based and kinetic models) that use gene annotation and ontology, prerequisite knowledge that include stoichiometry, thermodynamic and evolutionary assumptions, and enzyme kinetics [161,206,207].

The reconstruction of metabolic pathways is, however, challenged by the vast number of genes belonging to large gene families, which encode metabolic enzymes, and the expensive, time-consuming biochemical and genetic experimental approaches [205,208]. Additionally, the advancements of data acquisition technologies have resulted in increased large and complex omics datasets, which makes it challenging for hypothesis-driven models to incorporate and extract meaningful information from the data for pathway modeling [206,209]. For instance, constraint-based models (CBMs) are limited by their steady-state assumption and insufficiency to accurately capture cellular dynamics such as metabolic responses, whereas kinetic models are limited by the unknown mechanistic kinetic rate law for most enzymes involved in specific reactions and the computational challenges associated with parameter estimation and model expansion [207,210,211]. Hence, alternative mathematical and computational models with the capability of predicting organism-specific metabolic pathways directly from omics data are required to address these challenges. Data-driven models have been reported to predict biological systems' behavior, without prerequisite knowledge, from captured data [206,207,212].

Data-driven models are based on ML principles that enable them to predict pathway dynamics from experimental data without any knowledge on the organism's metabolism. This systematic approach increases the prediction accuracy with increasing data [210]. However, data-driven models cannot provide mechanistic explanations of phenomena, thus they are combined with hypothesis-driven models to provide mechanistic insights in a sample-specific manner [209,211]. In such hybrid-models, experimental data is integrated with the hypothesis-driven models to extract mechanistic information that is provided as input data for unsupervised or supervised ML algorithms that result in output data as relevant features that improve prediction accuracy and genotype to phenotype predictions [36,213]. ML algorithms optimize CBMs and kinetic models using known pathway instances curated in databases as training data to facilitate predictions based on experimental data and not exclusively on prior knowledge [207], to extract features in existing pathway and models [36], and to map multi-omic data onto a metabolic model that contains different integrated omics and model data through different ML algorithms [214]. Furthermore, the integration of ML algorithms with hypothesis-driven models, enables the models to accurately extract complex patterns from non-linear biological samples, improve predictability and provide cause–effect information about the metabolites in the pathway [209,214,215]. Several tools have been developed based on the aforementioned ML algorithms with the aim of modeling metabolic pathways.

Probabilistic modeling for Untargeted Metabolomics Analysis (PUMA), a Bayesian inference-based probabilistic approach, is an example of a model that predicts pathways that are likely to be active based on the metabolomics measurements, followed by probabilistic annotations, which assign chemical identities to the measurements. A pathway

is predicted to be active if the measurements generated from a particular path is greater than the user-defined threshold [216]. The recently developed Lilikoi v2.0 R software, implements a DL neural network classifier in addition to six ML algorithms including RF and SVM, for the prediction and visualization of metabolic pathways. Lilikoi requires metabolomics data matrix with a column of categorical variables (i.e., treated/control) for classification as input data, which is analyzed prior to the construction of the pathway [217]. RetroPath Reinforcement Learning-based metabolic space exploration (RetroPath RL) implements the Monte Carlo Tree Search (MCTS) reinforcement learning method (Figure 3) to discover and suggest experimentally relevant pathways. Due to its focus on exploring the research space, RetroPath RL can propose multiple pathways for the same compound. This feature will greatly aid in the elucidation of plant response mechanisms to stresses, as plants consist of metabolites involved in multiple pathways. RetroPath RL, however, has its limits, particularly in pathway ranking [218,219].

Ranking pathways with accuracy has been achieved by the adoption of a SVM in the multi-label classifier, iMPT-FRAKEL, which uses compound fingerprints to identify metabolic pathways of compounds [204]. Pathway Activity Level Scoring (PALS), available as a Python library, command line tool, and web application, analyzes metabolites grouped as metabolic pathways by ranking changing pathways or intensities in sets of metabolites over different experimental conditions [220]. The aforementioned ML-based tools or models hold great potential in the identification and discovery of pathways from poorly or lack of characterized experimental data and highlight information on signaling pathways, gene expression regulation and epigenetics within the system studied. The extraction of such information in plant-environmental metabolomic studies will contribute to the elucidation of plant response mechanisms to stress. Furthermore, the differences in plant response mechanisms to specific stresses can also be highlighted.

2.4. Large-Scale Metabolite Annotation

The developments in metabolomics platforms enable large-scale MS/MS experiments containing thousands of distinct spectra from a single sample. These detected features and their MS/MS spectra can be used for matching to spectral libraries and can include indicative fragmentation patterns that provide information on the often yet unknown chemical structures that were fragmented [221,222]. However, this process is still cumbersome and laborious for large-scale data, particularly for plant metabolomics data which is vastly diverse in physiochemical properties [222,223]. Hence, we are currently witnessing the development of tools that can analyze such large-scale, complex metabolomic data using computational networking approaches such as molecular networking (MN), which provides visualization of all the detected features and their chemical relationships by grouping structurally similar features into a network to enable mining of the metabolome (Figure 4) [224–226].

2.4.1. Spectral Similarity and Substructure Based Annotation

GNPS, a web-based ecosystem for the sharing, storage and analysis of raw, processed or annotated MS spectral data, enables the generation and analyses of molecular networks (Figure 4) [6,227]. MN through GNPS also enables rapid dereplication, the process of identifying known metabolites, as well as detection of analogs and the discovery of novel metabolites from the acquired MS spectral data [224,228]. Dereplication is achieved by automatically matching the thousands of MS/MS spectra acquired to public MS/MS spectral libraries within GNPS [224,229]. The large-scale spectral matching of MS spectral data is one approach that GNPS uses to facilitate the analysis of vast number of samples, as exemplified by the analysis of 15 batches of *Microctis Folium* (*Microcos paniculata* L.) leaf extracts that resulted in the identification of 168 metabolites, of which 165 metabolites were identified for the first time in *Microctis Folium*, and approximately 500 analogues could also be annotated [227]. In addition to GNPS, MS2LDA is another metabolome mining tool that uses an unsupervised Latent Dirichlet Allocation (LDA) algorithm to

group molecules based on shared molecular substructures (i.e., as conserved fragments and neutral loss features), referred to as Mass2Motifs, regardless of their overall spectral similarity (Figure 4) [8,230].

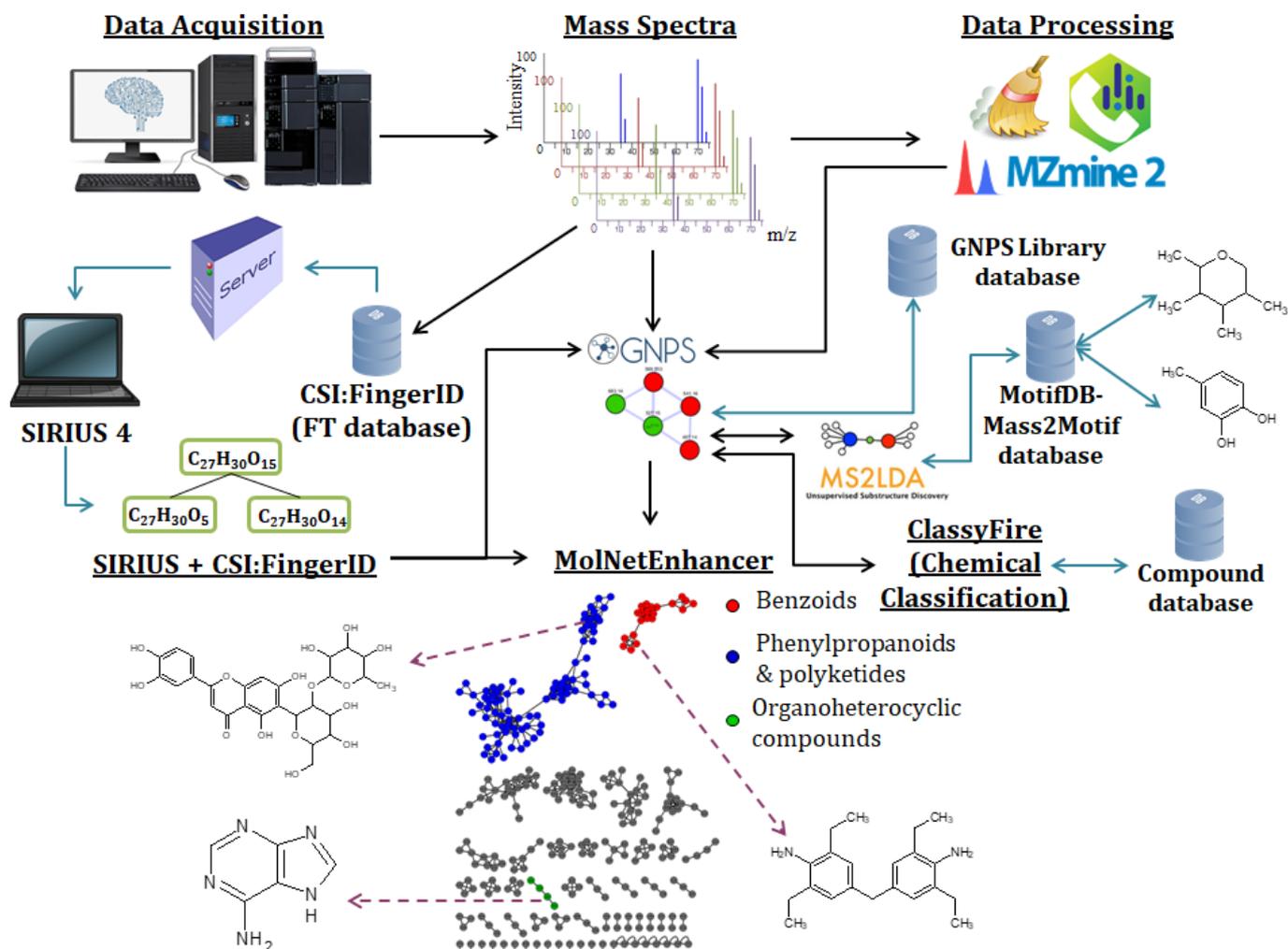


Figure 4. A schematic overview of 4IR technologies applied in metabolomics for large-scale automatic metabolite annotation and exploration by molecular networking. 4IR technologies include analytical intelligence (AI) that ensures high reliable/quality data acquisition, cloud computing and machine learning algorithms that facilitate rapid searching in library databases for spectral and structural similarities (i.e., GNPS, and MotifDB derived from MS2LDA), prediction of molecular formula and structure (SIRIUS + CSI:FingerID), and chemical class annotations (ClassyFire). MolNetEnhancer incorporates GNPS molecular networking, MS2LDA and the in silico annotation tools (ClassyFire, SIRIUS + CSI:FingerID, etc.) outputs, to provide detailed structural and chemical molecular information for large-scale MS data exploration and annotation.

The ms2lda.org web application allows us to inspect the extracted Mass2Motifs and annotate them by expert knowledge, comparison of mass fragments and neutral losses to those in MS/MS libraries, or by either comparison to Mass2Motifs extracted from earlier experiments [8,230]. This process is, however, laborious, particularly for large-scale data, and thus MS2LDA has been improved for automating the matching process to include the analysis of multiple samples with MotifSets of structurally characterized Mass2motifs from MotifDB together with “free Mass2Motifs” to capture the yet “unknown” chemistry not captured yet in MotifDB [231]. MS2LDA was applied to MS/MS data of 70 *Rhamnaceae* plant species and resulted in the extraction of 200 Mass2Motifs, of which 25 were annotated with substructures. Characterized Mass2Motifs enabled the probing of substructural diversity within the plant family. For instance, only the Rhamnoid clade was found to

develop diversity in flavonoid glycoside, whereas the Ziziphoid clade developed more variety in the triterpenoid pathway [9]. Thus, MS2LDA enables the extraction of sub-structural diversity within each class of metabolites from complex datasets and provides meaningful biochemical interpretation. Furthermore, an increase in structurally characterized Mass2Motifs will accelerate the identification of metabolites within plants dark matter with high confidence and thereby enhance understanding of plant mechanisms [9,231]. Despite these advances, structural elucidation of entire (plant) metabolites remains challenging in untargeted metabolomics workflows [171]. In the following sections, a number of computational metabolomics tools are highlighted that further enhance the annotation of either complete structures, substructures, or chemical compound class information to (plant) metabolomics profiles.

2.4.2. Structure-Based Annotation

Compound Structure Identification (CSI): FingerID is a web-service tool for searching in molecular structure databases using MS/MS data (Figure 4) [171,232]. CSI:FingerID converts a spectrum into a fragmentation tree that is searched against a database of known trees to predict a molecular fingerprint of the unknown compound. Although CSI:FingerID achieves increased identification rates compared to other related methods, it searches fragmentation trees individually [171,233]. Integration of CSI:FingerID with SIRIUS 4, a software tool for the computational annotation of MS/MS data, enables the full LC-MS/MS dataset to be processed instead of an individual compound [171,234].

The SIRIUS 4 tool offers molecular formula annotation and, when integrated with CSI:FingerID, structure database search and ranked annotations (Figure 4). This is achieved by first analyzing isotope patterns and fragmentation trees to determine the molecular formula of the query compound, followed by the use of CSI:FingerID to predict molecular fingerprint of the resultant spectrum and fragmentation tree. The predicted fingerprint is searched against the structural database to identify the most likely candidate [171,235]. SIRIUS 4 is capable of deducing information of unknown compounds, through Zodiac, now including molecules with higher masses (>500 Da) that are particularly of interest to plant metabolomics researchers, that are not structurally annotated and thus facilitate the mining of the metabolome [194]. Similarity score functions are crucial for the accurate matching of experimentally acquired spectra to library spectra for structural elucidation and for the assessment of spectral pair similarities [236].

2.4.3. Spectral Similarity Scoring for Library Matching and Correlation of Spectra

Spec2Vec, inspired by a natural language processing based algorithm, is a novel spectral similarity score that learns the relations between fragments to enable accurate spectral library matches. Spec2Vec computationally calculates similarity scores more efficiently than cosine-based scores, correlates more closely with structural similarity and identifies relationships between different spectra that are chemically related, especially for larger molecules with two or more subtle local modifications. Spec2Vec thus enables the rapid querying spectra of unknown molecules against all spectra in large databases and the selection of potential candidates for further exploration [236]. Additionally, Spec2Vec will enhance interpretation as it facilitates the correct assignment of metabolites, whose chemical class can be annotated and used to supplement metabolite annotation [237]. Despite its remarkable performance, Spec2Vec was not trained for the task of returning a higher similarity score for spectral pairs of structurally more closely related metabolites; therefore, very recently, MS2DeepScore was introduced that uses a Siamese neural network to predict the structural similarity between two chemical structures solely based on their MS/MS fragmentation spectra [238]. MS2DeepScore outperforms classical spectral similarity measures as well as Spec2Vec in retrieving chemically related compound pairs from large mass spectral datasets, thereby illustrating its potential for spectral library matching.

2.4.4. Chemical Compound Class-Based Annotation

ClassyFire, a web-based program, enables systematic chemical classification into a formal chemical ontology (Figure 4). ClassyFire automatically assigns all known chemical compounds to the predefined taxonomy ChemOnt based on their chemical structures and structural features [239,240]. The ClassyFire server and database consists of more than 77 million compounds that vary from drugs, toxins, phytochemicals, natural and synthetic molecules. Access to this database via the web server enables rapid, large-scale and automated chemical classification that can aid in the annotation or enrichment of known and unknown compounds [237,239,241]. MolNetEnhancer is a software package that incorporates the outputs of GNPS MN, MS2LDA, in silico annotation tools such as SIRIUS+CSI:FingerID, and the automated chemical classification through ClassyFire, into a single workflow (Figure 4) [237].

MolNetEnhancer reveals detailed information on molecular families and the subtle structural differences between them, thus resulting in enhanced molecular networks that facilitate rapid metabolite exploration of large complex datasets by providing a global visual of chemical diversity [226,237]. Another chemical classification tool is NPClassifier, a DL tool for automated structural classification of natural products (NP), that uses chemical fingerprints as input encoding and classifies NP at three levels; 7 Pathways, 70 Superclasses and 653 Classes. NPClassifier has the potential to accelerate NP discovery by enabling large-scale metabolome mining [242,243].

Most recently, the chemical class assignment and ontology prediction using mass spectrometry (CANOPUS) DNN tool for systematic compound class annotation was introduced. CANOPUS predicts compound classes from fragmentation spectra, including biologically relevant classes. CANOPUS specifically targets compounds that have no spectral or structural reference data available and predicts classes for which no MS/MS training data available [241,244]. CANOPUS assigns compound classes to all MS/MS features for which fragmentation trees can be computed in a LC-MS/MS run and automatically provides structural insights of novel compounds from crude samples, thus accelerating the structural elucidation process or allowing prioritization of a subset of metabolite features to further examine [244]. CANOPUS and the aforementioned computational tools and technological advancements are increasingly being applied in metabolomics studies. Whilst there is a lot of promise in these kinds of approaches, exactly how they benefit and improve plant metabolomics workflows is yet to be witnessed. We refer the interested reader to a recent review that further explains the workings of substructure-based, chemical compound class-based, and network-based strategies for metabolite annotation [245].

2.4.5. Large-Scale and Repository-Wide Metabolomics Analyses

The 4IR technologies will improve plant metabolomics workflows and facilitate complex design studies by, for instance, investigating plant response mechanisms to multiple stresses simultaneously. Automation in sample preparation and in analytical platforms, equipped with analytical intelligence will enable large metabolite coverage, rapid, automatic and potentially remote data acquisition from complex samples such as plants. Development and implementation of machine and deep learning algorithms will aid in data mining and pathway modelling by ensuring acquisition of reliable data with minimal noise and artifacts, which relevant features associated to the particular stress are extracted from and confidently annotated through spectral library matches, thus achieving rapid large-scale annotations of metabolites that contribute to elucidating the metabolome studied, in this case the plant metabolome in response to abiotic stresses. The expected transition from experiment-based analysis toward repository-wide analysis has recently gained traction by the introduction of the MASST [246] and ReDU [247] frameworks that allow researchers to locate similar MS/MS spectra across all public datasets, i.e., akin to doing a BLAST search with a genetic sequence. An increasing number of these public datasets are annotated with consistent metadata through a controlled vocabulary (i.e., ReDU) or ontologies (i.e., MetaboLights [7]). Thus, if a researcher finds hits across different

public datasets, the metadata (sample information) can provide valuable information on the possible source of that fragmented component in their sample. Moreover, plant researchers can easily select subsets of publicly available datasets of the plant-based origins relevant to their experiment to compare their data to and facilitate the annotation of plant metabolites.

3. Metabolomics and Plant Responses to Abiotic Stresses—Current Frameworks

Abiotic stresses are changes in the environmental conditions that negatively affect plants' growth and development, metabolism and physiology. The main abiotic stresses include extreme temperatures, drought, waterlogging, light, radiation, salinity, heavy metals and chilling [14,18]. Abiotic stresses activate complex cellular responses that are increasingly being elucidated by progresses made in exploring and understanding plant responses to abiotic stress factors at the whole-plant, physiological, cellular, biochemical and molecular levels. As briefly mentioned in Section 1.2, plant responses to abiotic stress involve the activation of multi-layered events comprising complex gene interactions and crosstalk with different molecular pathways and networks. Current knowledge (and understanding) of plant responses to abiotic stresses (Section 1.2) is still the tip of an iceberg. Comprehensive and accurate metabolic descriptions (models) are imperatively needed and essential for devising a roadmap for the next generation of crops resilient to environmental deterioration.

As mentioned in Section 1.1, metabolomics is increasingly enabling the decoding of the language of cells at molecular level, advancing the understanding of regulatory network rules and mechanistic events at cellular and chemical space of the plant under consideration. The application of this multidisciplinary *omics* science in studying plant responses to abiotic stresses has gained a momentum, generating insights into molecular events that govern plant 'defensome'. To illustrate such applications, in this review, we focus on drought and salinity stress conditions. Drought and salinity are the two most impactful abiotic stresses, affecting 45% and 19.5% of the world's agricultural lands, respectively. These two (globally) major abiotic stresses (drought and salinity) lead to an annual economic loss above USD 20 billion in developing countries [248]. While drought stress is caused by water deficit, salinity is the result of accumulation of ions (sodium and chloride) in the rhizosphere [248–250]. Both stresses affect morphological, physiological and biochemical processes in plants, resulting in growth suppression and, failing stress alleviation, eventually in cell death. The onset of drought and salt stresses in plants prompts stress-responsive mechanisms including the signal perception and transduction via several signalling pathways including second messenger signalling (ROS, Ca²⁺), the kinase-signalling cascades (MAPK) and hormone signalling [251–253]. Stress signal perception triggers several stress-responsive processes encompassing genetic reprogramming, activation of ion channels and synthesis of osmolytes.

Several metabolic studies on different plants have indicated altered primary and secondary metabolite profiles in response to salinity and drought stresses (Table 3). Sugars, amino acids, and organic acids and their derivatives are the most altered primary metabolites in osmotic stressed plants [254–256]. Additionally, the perturbed metabolism of secondary (specialized) metabolites (phenolics acids, flavonoids, phytohormones) has also been detected in several metabolic studies in response to salinity and drought stresses [257–259]. Comprehensive characterisation of stress-responsive metabolic profiles and fluxes is essential for the elucidation of the molecular mechanisms used by plants to combat abiotic stress conditions. Such knowledge can be translated into suitable biotechnological tools for the development of stress-tolerant crop plants for sustainable food security.

Metabolomics was applied to elucidate the integrative biochemical networks of two spring-wheat cultivars (*Bahar*—drought-susceptible; *Kavir*—drought-tolerant) to drought stress. Metabolome changes of wheat leaves exposed to 7 days of drought stress was investigated with LC-MS. Three hundred peaks were detected per sample, with 165 and 146 identified metabolites for *Bahar* and *Kavir*, respectively. The main metabolites changed due to drought stress were amino acids, organic acids and sugars. In *Bahar*, proline,

methionine, arginine, lysine, aromatic and branched chain amino acids were increased. In contrast, only the purine metabolic pathway was significantly affected by the drought stress in *Kavir*. Metabolomics in this study thus provided better understanding in the wheat plant response mechanisms to drought stress [256].

Additionally, metabolomics has also been used to investigate two foxtail millet (*Setaria italica*) cultivars of Yugu2 and An04 subjected to salinity stress. A total of 720 metabolites were identified from the LC-MS data, with 150 metabolites involved in the response to salt stress, ranging from organic acids, sugars, phenolics, amino acids and others. The flavonoids were significantly up-regulated in the roots, with the majority of them and other secondary metabolites being observed in Yungu2. This metabolomics analysis revealed the flavonoid pathway, starch and sucrose metabolism, glutathione metabolism, glycerophospholipid metabolism, ascorbate and aldarate metabolism, phenylpropanoid and shikimate pathways to be involved in the responses to salinity stress. The authors also noted metabolites such as lysophospholipids, which play a role in response to salinity stress. Therefore, they suggest that there are diverse existing or new response mechanisms in foxtail millet to cope with salinity stress [260].

Table 3. An overview of metabolism alterations in response to drought and salinity stresses, and their corresponding roles in plant stress responses.

Metabolite Group	Stress-Responsive Roles	Plant Species	References
Amino acids	ROS scavenging (proline), protein stabilisation and synthesis, redox control	<i>Dianthus superbus</i> , <i>Lens esculenta</i>	[261,262]
Polyols	Protection of photosynthesis systems, ROS scavenging, protein stabilisation	Rice, apple leaves, <i>Fraxinus excelsior</i> , <i>Zea mays</i>	[263–266]
Organic acids	Energy production, signalling molecules, antioxidant activities	<i>Oryza sativa</i> , Wheat	[249,267]
Sugars	Signalling molecules, carbon energy reserve, maintenance of redox homeostasis, osmoprotectants	<i>Solanum lycopersicum</i> , <i>Triticum aestivum</i>	[268,269]
Polyamines	Activation of antioxidant enzymes, regulation of ion channels activity, protein and membrane stabilisation	Tobacco, <i>Triticum aestivum</i>	[270,271]
Phenolics	Hormonal regulation, antioxidant activity, photosynthetic activity	Patagonian shrublands, <i>Amaranthus tricolor</i>	[272,273]

4. Conclusions and Perspectives

Metabolomics has been crucial in elucidating cellular and molecular mechanisms of plant-environmental interactions. There are still some bottlenecks that limit or hamper the comprehensiveness of biological insights derived from this multidisciplinary omics science, metabolomics. Some of these challenges include the complexity, non-linearity and volume of the generated (metabolomics) datasets, the lack of accurate metabolite annotation, the difficulty in assigning precursor ions to product ions for overlapping chromatographic peaks, and the lack of large-scale analysis tools and reconstruction of metabolic pathways. These challenges can be addressed by the implementation or a revival of 4IR technologies in plant metabolomics workflows. For example, the automation of sample preparation methods has reduced the sample preparation time with minimum variation and human error, thereby resulting in large-scale metabolite extraction, increased metabolome coverage and the potential to discover novel metabolites that will illuminate pathways involved in plant response molecular and cellular events, thereby elucidating plant response mechanisms. Furthermore, the advancement and incorporation of analytical intelligence in instruments facilitates accurate quantification of metabolites through automatic optimization of chromatographic conditions to enhance peak picking and resolution, and reproducibility of the data for accurate and confident metabolite annotation. Machine and deep learning algorithms can handle non-linear data, thus accurately identifying

complex relationships within plant data, and extracting relevant metabolic features that correlate to biological origin (i.e., abiotic stress) without bias or approximation. Machine learning and computational tools with predictive capabilities and rapid search of unknown metabolites in databases aid in metabolic pathway and molecular formula predictions, and rapid metabolite annotation and/or structural elucidation at various levels from chemical compound class, via substructures, to complete metabolite structures, thus facilitating comprehensive large-scale mining of (plant) metabolomes.

We expect that future developments of 4IR technologies in plant metabolomics will further increase automation and connectivity within workflows. For instance, robotic systems may link prepared samples from the automated sample preparation method instruments to the analytically intelligent instruments with a custom software that utilizes the internet of things and wireless telecommunication technologies to automatically upload the resultant data onto a cloud server for download or sharing by the user, thus enabling remote data acquisition and analysis. Additionally, portable analytical devices that facilitate metabolite data acquisition directly from the plant tissue without extraction, will be developed and may be mounted on to drones for field purposes. An increase in novel computational tools for processing and analysis of spectral imaging data is anticipated. Development of computational tools based on machine and deep learning algorithms for predicting and grouping metabolite functional activity will enable pathway constructions and elucidate/predict the synergistic behavior of metabolites/pathways in response to the changing environment, thus further elucidating plant response mechanisms. The integration of 4IR technologies will also facilitate rapid data analysis. For instance, the integration of analytical intelligent instruments with cloud servers, computational tools and algorithms, will enable simultaneous (i.e., real time) metabolite acquisition and metabolite annotation, thus enhancing rapidity of plant metabolomic studies and rapid elucidation of plant response mechanisms to stress. The first examples of tools that are the building blocks of such large-scale analyses were highlighted in this review. From our review, it becomes clear that in the last two decades, metabolomics workflows have matured and that we are at the dawn of a new era where computational metabolomics analysis workflows in combination with more automated data sampling and acquisition will decrease the barrier to implement metabolomics studies in research programs in academia and industry. We believe that this will increase the rate of development for next generation of crops that are highly productive and resilient to climate change.

Author Contributions: Conceptualization, F.T. and J.J.J.v.d.H.; Writing—Original draft preparation, M.M.T. and K.H.C.; writing—review and editing, F.T. and J.J.J.v.d.H.; supervision, F.T. and J.J.J.v.d.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The University of Johannesburg (UJ) is thanked for providing a scholarship to M.M.T.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, M.; David, J.M.; Kim, S.H. The fourth industrial revolution: Opportunities and challenges. *Int. J. Financ. Res.* **2018**, *9*, 90–95. [[CrossRef](#)]
2. Ndung'u, N.; Signé, L. The Fourth Industrial Revolution and digitization will transform Africa into a global powerhouse. *Foresight Afr.* **2020**, *2020*, 60–73.
3. Koh, L.; Orzes, G.; Jia, F. The fourth industrial revolution (Industry 4.0): Technologies disruption on operations and supply chain management. *Int. J. Oper. Prod. Manag.* **2019**, *39*, 817–828. [[CrossRef](#)]
4. Mendez, K.M.; Pritchard, L.; Reinke, S.N.; Broadhurst, D.I. Toward collaborative open data science in metabolomics using Jupyter Notebooks and cloud computing. *Metabolomics* **2019**, *15*, s11306–s113019. [[CrossRef](#)]
5. Tugizimana, F.; Engel, J.; Salek, R.; Dubery, I.; Piater, L.; Burgess, K. *The Disruptive 4IR in the Life Sciences: Metabolomics*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 674, ISBN 9783030482305.

6. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kaponov, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [[CrossRef](#)]
7. Kale, N.S.; Haug, K.; Conesa, P.; Jayseelan, K.; Moreno, P.; Rocca-Serra, P.; Nainala, V.C.; Spicer, R.A.; Williams, M.; Li, X.; et al. MetaboLights: An open-access database repository for metabolomics data. *Curr. Protoc. Bioinform.* **2016**, *2016*, 14.13.1–14.13.18. [[CrossRef](#)]
8. Van Der Hooft, J.J.J.; Wandy, J.; Barrett, M.P.; Burgess, K.E.V.; Rogers, S. Topic modeling for untargeted substructure exploration in metabolomics. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 13738–13743. [[CrossRef](#)]
9. Kang, K.B.; Ernst, M.; van der Hooft, J.J.J.; da Silva, R.R.; Park, J.; Medema, M.H.; Sung, S.H.; Dorrestein, P.C. Comprehensive mass spectrometry-guided phenotyping of plant specialized metabolites reveals metabolic diversity in the cosmopolitan plant family Rhamnaceae. *Plant J.* **2019**, *98*, 1134–1144. [[CrossRef](#)] [[PubMed](#)]
10. Kell, D.B.; Brown, M.; Davey, H.M.; Dunn, W.B.; Spasic, I.; Oliver, S.G. Metabolic footprinting and systems biology: The medium is the message. *Nat. Rev. Microbiol.* **2005**, *3*, 557–565. [[CrossRef](#)]
11. Rai, A.; Yamazaki, M.; Saito, K. A new era in plant functional genomics. *Curr. Opin. Syst. Biol.* **2019**, *15*, 58–67. [[CrossRef](#)]
12. Fernie, A.R.; Bachem, C.W.B.; Helariutta, Y.; Neuhaus, H.E.; Prat, S.; Ruan, Y.L.; Stitt, M.; Sweetlove, L.J.; Tegeder, M.; Wahl, V.; et al. Synchronization of developmental, molecular and metabolic aspects of source–sink interactions. *Nat. Plants* **2020**, *6*, 55–66. [[CrossRef](#)]
13. Drobek, M.; Fraç, M.; Cybulska, J. Plant biostimulants: Importance of the quality and yield of horticultural crops and the improvement of plant tolerance to abiotic stress—A review. *Agronomy* **2019**, *9*, 335. [[CrossRef](#)]
14. Llorens, E.; González-Hernández, A.I.; Scalschi, L.; Fernández-Crespo, E.; Camañes, G.; Vicedo, B.; García-Agustín, P. *Priming Mediated Stress and Cross-Stress Tolerance in Plants: Concepts and Opportunities*; Elsevier: Amsterdam, The Netherlands, 2020; ISBN 9780128178935.
15. Janni, M.; Gulli, M.; Maestri, E.; Marmiroli, M.; Valliyodan, B.; Nguyen, H.T.; Marmiroli, N.; Foyer, C. Molecular and genetic bases of heat stress responses in crop plants and breeding for increased resilience and productivity. *J. Exp. Bot.* **2020**, *71*, 3780–3802. [[CrossRef](#)]
16. Ramegowda, V.; Da Costa, M.V.J.; Harihar, S.; Karaba, N.N.; Sreeman, S.M. *Abiotic and Biotic Stress Interactions in Plants: A Cross-Tolerance Perspective*; Elsevier: Amsterdam, The Netherlands, 2020; ISBN 9780128178935.
17. Nobori, T.; Tsuda, K. The plant immune system in heterogeneous environments. *Curr. Opin. Plant Biol.* **2019**, *50*, 58–66. [[CrossRef](#)] [[PubMed](#)]
18. Razzaq, A.; Sadia, B.; Raza, A.; Hameed, M.K.; Saleem, F. Metabolomics: A way forward for crop improvement. *Metabolites* **2019**, *9*, 303. [[CrossRef](#)] [[PubMed](#)]
19. El Boukhari, M.E.M.; Barakate, M.; Bouhia, Y.; Lyamlouli, K. Trends in seaweed extract based biostimulants: Manufacturing process and beneficial effect on soil-plant systems. *Plants* **2020**, *9*, 359. [[CrossRef](#)]
20. Jones, J.D.G.; Dangl, J.L. The plant immune system. *Nature* **2006**, *444*, 323–329. [[CrossRef](#)]
21. Redondo-Gómez, S. Abiotic and Biotic Stress Tolerance in Plants. In *Molecular Stress Physiology of Plants*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 1–440, ISBN 9788132208075.
22. Obata, T.; Fernie, A.R. The use of metabolomics to dissect plant responses to abiotic stresses. *Cell. Mol. Life Sci.* **2012**, *69*, 3225–3243. [[CrossRef](#)]
23. Zhu, J. Review Abiotic Stress Signaling and Responses in Plants. *Cell* **2016**, *167*, 313–324. [[CrossRef](#)]
24. Saud, S.; Li, X.; Chen, Y.; Zhang, L.; Fahad, S.; Hussain, S.; Sadiq, A.; Chen, Y. Silicon application increases drought tolerance of Kentucky bluegrass by improving plant water relations and morphophysiological functions. *Sci. World J.* **2014**, *2014*. [[CrossRef](#)]
25. Tátrai, Z.A.; Sanoubar, R.; Pluhár, Z.; Mancarella, S.; Orsini, F.; Gianquinto, G. Morphological and Physiological Plant Responses to Drought Stress in *Thymus citriodorus*. *Int. J. Agron.* **2016**, *2016*. [[CrossRef](#)]
26. Tugizimana, F.; Piater, L.; Dubery, I. Plant metabolomics: A new frontier in phytochemical analysis. *S. Afr. J. Sci.* **2013**, *109*, 1–11. [[CrossRef](#)]
27. Castro-moretti, F.R.; Gentzel, I.N.; Mackey, D.; Alonso, A.P. Metabolomics as an emerging tool for the study of plant–pathogen interactions. *Metabolites* **2020**, *10*, 52. [[CrossRef](#)] [[PubMed](#)]
28. Dawid, C.; Hille, K. Functional metabolomics—a useful tool to characterize stress-induced metabolome alterations opening new avenues towards tailoring food crop quality. *Agronomy* **2018**, *8*, 138. [[CrossRef](#)]
29. Monge, M.E.; Dodds, J.N.; Baker, E.S.; Edison, A.S.; Fernacutendez, F.M. Challenges in Identifying the Dark Molecules of Life. *Annu. Rev. Anal. Chem.* **2019**, *12*, 177–199. [[CrossRef](#)] [[PubMed](#)]
30. Wolfender, J.L.; Nuzillard, J.M.; Van Der Hooft, J.J.J.; Renault, J.H.; Bertrand, S. Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography–High-Resolution Tandem Mass Spectrometry and NMR Profiling, in Silico Databases, and Chemometrics. *Anal. Chem.* **2019**, *91*, 704–742. [[CrossRef](#)] [[PubMed](#)]
31. Rinschen, M.M.; Ivanisevic, J.; Giera, M.; Siuzdak, G. Identification of bioactive metabolites using activity metabolomics. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 353–367. [[CrossRef](#)]
32. Beisken, S.; Eiden, M.; Salek, R.M. Getting the right answers: Understanding metabolomics challenges. *Expert Rev. Mol. Diagn.* **2015**, *15*, 97–109. [[CrossRef](#)] [[PubMed](#)]

33. Tugizimana, F.; Mhlongo, M.I.; Piater, L.A.; Dubery, I.A. Metabolomics in plant priming research: The way forward? *Int. J. Mol. Sci.* **2018**, *19*, 1759. [[CrossRef](#)]
34. Salem, M.A.; De Souza, L.P.; Serag, A.; Fernie, A.R.; Farag, M.A.; Ezzat, S.M.; Alseekh, S. Metabolomics in the context of plant natural products research: From sample preparation to metabolite analysis. *Metabolites* **2020**, *10*, 37. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, X.W.; Li, Q.H.; Xu, Z.D.; Dou, J.J. Mass spectrometry-based metabolomics in health and medical science: A systematic review. *RSC Adv.* **2020**, *10*, 3092–3104. [[CrossRef](#)]
36. Damiani, C.; Gaglio, D.; Sacco, E.; Alberghina, L.; Vanoni, M. Systems metabolomics: From metabolomic snapshots to design principles. *Curr. Opin. Biotechnol.* **2020**, *63*, 190–199. [[CrossRef](#)]
37. Kell, D.B.; Oliver, S.G. The metabolome 18 years on: A concept comes of age. *Metabolomics* **2016**, *12*, 1–8. [[CrossRef](#)] [[PubMed](#)]
38. Tugizimana, F.; Djami-Tchatchou, A.T.; Steenkamp, P.A.; Piater, L.A.; Dubery, I.A. Metabolomic analysis of defense-related reprogramming in sorghum bicolor in response to Colletotrichum sublineolum infection reveals a functional metabolic web of phenylpropanoid and flavonoid pathways. *Front. Plant Sci.* **2019**, *9*, 1–20. [[CrossRef](#)]
39. Filla, L.A.; Sanders, K.L.; Filla, R.T.; Edwards, J.L. Automated sample preparation in a microfluidic culture device for cellular metabolomics. *Analyst* **2016**, *141*, 3858–3865. [[CrossRef](#)] [[PubMed](#)]
40. 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](#)]
41. Joo, M.; Park, J.M.; Duong, V.A.; Kwon, D.; Jeon, J.; Han, M.; Cho, B.K.; Choi, H.K.; Lee, C.G.; Kang, H.G.; et al. An automated high-throughput sample preparation method using double-filtration for serum metabolite LC-MS analysis. *Anal. Methods* **2019**, *11*, 4060–4065. [[CrossRef](#)]
42. Roopashree, K.M.; Naik, D. Advanced method of secondary metabolite extraction and quality analysis. *J. Pharmacogn. Phytochem.* **2019**, *8*, 1829–1842.
43. Rigobello-Masini, M.; Pereira, E.A.O.; Abate, G.; Masini, J.C. Solid-Phase Extraction of Glyphosate in the Analyses of Environmental, Plant, and Food Samples. *Chromatographia* **2019**, *82*, 1121–1138. [[CrossRef](#)]
44. Van Der Hooft, J.J.J.; Akermi, M.; Unlu, F.Y.; Mihaleva, V.; Roldan, V.G.; Bino, R.J.; De Vos, R.C.H.; Vervoort, J. Structural annotation and elucidation of conjugated phenolic compounds in black, green, and white tea extracts. *J. Agric. Food Chem.* **2012**, *60*, 8841–8850. [[CrossRef](#)]
45. Raks, V.; Al-Suod, H.; Buszewski, B. Isolation, Separation, and Preconcentration of Biologically Active Compounds from Plant Matrices by Extraction Techniques. *Chromatographia* **2018**, *81*, 189–202. [[CrossRef](#)]
46. Bladergroen, M.R.; van der Burgt, Y.E.M. Solid-Phase Extraction Strategies to Surmount Body Fluid Sample Complexity in High-Throughput Mass Spectrometry-Based Proteomics. *J. Anal. Methods Chem.* **2015**, *2015*. [[CrossRef](#)] [[PubMed](#)]
47. Sitnikov, D.G.; Monnin, C.S.; Vuckovic, D. Systematic Assessment of Seven Solvent and Solid-Phase Extraction Methods for Metabolomics Analysis of Human Plasma by LC-MS. *Sci. Rep.* **2016**, *6*, 1–11. [[CrossRef](#)] [[PubMed](#)]
48. Miggiels, P.; Wouters, B.; van Westen, G.J.P.; Dubbelman, A.C.; Hankemeier, T. Novel technologies for metabolomics: More for less. *TrAC Trends Anal. Chem.* **2019**, *120*, 115323. [[CrossRef](#)]
49. Reyes-Garcés, N.; Gionfriddo, E. Recent developments and applications of solid phase microextraction as a sample preparation approach for mass-spectrometry-based metabolomics and lipidomics. *TrAC Trends Anal. Chem.* **2019**, *113*, 172–181. [[CrossRef](#)]
50. Mousavi, F.; Bojko, B.; Pawliszyn, J. High-Throughput Solid-Phase Microextraction–Liquid Chromatography–Mass Spectrometry for Microbial Untargeted Metabolomics. In *Microbial Metabolomics: Methods in Molecular Biology*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 1859, pp. 133–152, ISBN 9781493987573.
51. Dugheri, S.; Mucci, N.; Bonari, A.; Marrubini, G.; Cappelli, G.; Ubiali, D.; Campagna, M.; Montalti, M.; Arcangeli, G. Liquid Phase Microextraction Techniques Combined with Chromatography Analysis: A Review. *Acta Chromatogr.* **2020**, *32*, 69–79. [[CrossRef](#)]
52. He, Y.; Concheiro-Guisan, M. Microextraction sample preparation techniques in forensic analytical toxicology. *Biomed. Chromatogr.* **2019**, *33*. [[CrossRef](#)]
53. Drouin, N.; Rudaz, S.; Schappler, J. Sample preparation for polar metabolites in bioanalysis. *Analyst* **2018**, *143*, 16–20. [[CrossRef](#)]
54. Sramkova, I.H.; Horstkotte, B.; Fikarova, K.; Sklenarova, H.; Solich, P. Direct-immersion single-drop microextraction and in-drop stirring microextraction for the determination of nanomolar concentrations of lead using automated Lab-In-Syringe technique. *Talanta* **2018**, *184*, 162–172. [[CrossRef](#)]
55. Kellogg, J.J.; Wallace, E.D.; Graf, T.N.; Oberlies, N.H.; Cech, N.B. Conventional and accelerated-solvent extractions of green tea (*Camellia sinensis*) for metabolomics-based chemometrics. *J. Pharm. Biomed. Anal.* **2017**, *145*, 604–610. [[CrossRef](#)]
56. Pinu, F.R.; Villas-Boas, S.G.; Aggio, R. Analysis of intracellular metabolites from microorganisms: Quenching and extraction protocols. *Metabolites* **2017**, *7*, 53. [[CrossRef](#)] [[PubMed](#)]
57. Njila, M.I.N.; Mahdi, E.; Lembe, D.; Nde, Z.; Doriane, N. Review on Extraction and Isolation of Plant Secondary Metabolites. In Proceedings of the 7th International Conference on Agricultural, Chemical, Biological and Environmental Sciences (ACBES-2017), Kuala Lumpur, Malaysia, 22–24 May 2017. [[CrossRef](#)]
58. Sanchez-Camargo, A.d.P.; Parada-Alonso, F.; Ibanez, E.; Cifuentes, A. Recent applications of on-line supercritical fluid extraction coupled to advanced analytical techniques for compounds extraction and identification. *J. Sep. Sci.* **2019**, *42*, 243–257. [[CrossRef](#)]
59. Akhtar, I.; Javad, S.; Yousaf, Z.; Iqbal, S.; Jabeen, K. Microwave assisted extraction of phytochemicals an efficient and modern approach for botanicals and pharmaceuticals. *Pak. J. Pharm. Sci.* **2019**, *32*, 223–230.

60. Lozano-Grande, M.A.; Dávila-Ortiz, G.; García-Dávila, J.; Ríos-Cortés, G.; Espitia-Rangel, E.; Martínez-Ayala, A.L. Optimisation of Microwave-Assisted Extraction of Squalene from *Amaranthus* spp. Seeds. *J. Microw. Power Electromagn. Energy* **2019**, *53*, 243–258. [[CrossRef](#)]
61. Melgar, B.; Dias, M.I.; Barros, L.; Ferreira, I.C.F.R.; Rodriguez-Lopez, A.D.; Garcia-Castello, E.M. Ultrasound and Microwave Assisted Extraction of *Opuntia* Fruit Peels Biocompounds: Optimization and Comparison Using RSM-CCD. *Molecules* **2019**, *24*, 3618. [[CrossRef](#)] [[PubMed](#)]
62. Ramasamy, J.; Kandasamy, R.; Palanisamy, S.; Nadesan, S. Optimization of Ultrasonic-Assisted Extraction of Flavonoids and Anti-oxidant Capacity from the Whole Plant of *Andrographis echinoides* (L.) Nees by Response Surface Methodology and Chemical Composition Analysis. *Pharmacogn. Mag.* **2019**, *15*, 547–556. [[CrossRef](#)]
63. Alcantara, C.; Zugic, T.; Abdelkebir, R.; Garcaa-Perez, J.V.; Jambrak, A.R.; Lorenzo, J.M.; Collado, M.C.; Granato, D.; Barba, F.J. Effects of ultrasound-assisted extraction and solvent on the phenolic profile, bacterial growth, and anti-inflammatory/antioxidant activities of mediterranean olive and fig leaves extracts. *Molecules* **2020**, *25*, 1718. [[CrossRef](#)] [[PubMed](#)]
64. Dugheri, S.; Mucci, N.; Bonari, A.; Marrubini, G.; Cappelli, G.; Ubiali, D.; Campagna, M.; Montalti, M.; Arcangeli, G. Solid phase microextraction techniques used for gas chromatography: A review. *Acta Chromatogr.* **2020**, *32*, 1–9. [[CrossRef](#)]
65. Khaled, A.; Gionfriddo, E.; Acquaro, V.; Singh, V.; Pawliszyn, J. Development and validation of a fully automated solid phase microextraction high throughput method for quantitative analysis of multiresidue veterinary drugs in chicken tissue. *Anal. Chim. Acta* **2019**, *1056*, 34–46. [[CrossRef](#)]
66. Medina, D.A.V.; Santos-Neto, A.J.; Cerda, V.; Maya, F. Automated dispersive liquid-liquid microextraction based on the solidification of the organic phase. *Talanta* **2018**, *189*, 241–248. [[CrossRef](#)] [[PubMed](#)]
67. Ekezie, F.G.C.; Sun, D.W.; Cheng, J.H. Acceleration of microwave-assisted extraction processes of food components by integrating technologies and applying emerging solvents: A review of latest developments. *Trends Food Sci. Technol.* **2017**, *67*, 160–172. [[CrossRef](#)]
68. Llompert, M.; Celeiro, M.; Dagnac, T. Microwave-assisted extraction of pharmaceuticals, personal care products and industrial contaminants in the environment. *TrAC Trends Anal. Chem.* **2019**, *116*, 136–150. [[CrossRef](#)]
69. Chemat, F.; Rombaut, N.; Sicaire, A.G.; Meullemiestre, A.; Fabiano-Tixier, A.S.; Abert-Vian, M. Ultrasound assisted extraction of food and natural products. Mechanisms, techniques, combinations, protocols and applications. A review. *Ultrason. Sonochem.* **2017**, *34*, 540–560. [[CrossRef](#)]
70. Panzella, L.; Moccia, F.; Nasti, R.; Marzorati, S.; Verotta, L.; Napolitano, A. Bioactive Phenolic Compounds from Agri-Food Wastes: An Update on Green and Sustainable Extraction Methodologies. *Front. Nutr.* **2020**, *7*, 1–27. [[CrossRef](#)] [[PubMed](#)]
71. Risticvic, S.; Souza-Silva, E.A.; Gionfriddo, E.; DeEll, J.R.; Cochran, J.; Hopkins, W.S.; Pawliszyn, J. Application of in vivo solid phase microextraction (SPME) in capturing metabolome of apple (*Malus × domestica* Borkh.) fruit. *Sci. Rep.* **2020**, *10*, 1–11. [[CrossRef](#)] [[PubMed](#)]
72. Kabir, A.; Locatelli, M.; Ulusoy, H.I. Recent trends in microextraction techniques employed in analytical and bioanalytical sample preparation. *Separations* **2017**, *4*, 36. [[CrossRef](#)]
73. Suh, J.H.; Han, S.B.; Wang, Y. Development of an improved sample preparation platform for acidic endogenous hormones in plant tissues using electromembrane extraction. *J. Chromatogr. A* **2018**, *1535*, 1–8. [[CrossRef](#)] [[PubMed](#)]
74. Gallo-Molina, A.C.; Castro-Vargas, H.I.; Garzón-Méndez, W.F.; Martínez Ramírez, J.A.; Rivera Monroy, Z.J.; King, J.W.; Parada-Alfonso, F. Extraction, isolation and purification of tetrahydrocannabinol from the *Cannabis sativa* L. plant using supercritical fluid extraction and solid phase extraction. *J. Supercrit. Fluids* **2019**, *146*, 208–216. [[CrossRef](#)]
75. Belwal, T.; Pandey, A.; Bhatt, I.D.; Rawal, R.S. Optimized microwave assisted extraction (MAE) of alkaloids and polyphenols from *Berberis* roots using multiple-component analysis. *Sci. Rep.* **2020**, *10*, 1–10. [[CrossRef](#)] [[PubMed](#)]
76. Tsiaka, T.; Fotakis, C.; Lantzouraki, D.Z.; Tsiantas, K.; Siapi, E.; Sinanoglou, V.J.; Zoumpoulakis, P. Expanding the Role of Sub-Exploited DOE-High Energy Extraction and Metabolomic Profiling towards Agro-Byproduct Valorization: The Case of Carotenoid-Rich Apricot Pulp. *Molecules* **2020**, *25*, 2702. [[CrossRef](#)]
77. Fleischer, H.; Drews, R.R.; Janson, J.; Chinna Patlolla, B.R.; Chu, X.; Klos, M.; Thurow, K. Application of a Dual-Arm Robot in Complex Sample Preparation and Measurement Processes. *J. Lab. Autom.* **2016**, *21*, 671–681. [[CrossRef](#)] [[PubMed](#)]
78. Elpa, D.P.; Prabhu, G.R.D.; Wu, S.P.; Tay, K.S.; Urban, P.L. Automation of mass spectrometric detection of analytes and related workflows: A review. *Talanta* **2020**, *207*, 120304. [[CrossRef](#)]
79. Joshi, S.; Chu, X.; Fleischer, H.; Roddelkopf, T.; Klos, M.; Thurow, K. Analysis of measurement process design for a dual-arm robot using graphical user interface. *IEEE Int. Instrum. Meas. Technol. Conf.* **2019**, 1–6. [[CrossRef](#)]
80. Alseekh, S.; Fernie, A.R. Metabolomics 20 years on: What have we learned and what hurdles remain? *Plant J.* **2018**, *94*, 933–942. [[CrossRef](#)] [[PubMed](#)]
81. Nandania, J.; Peddinti, G.; Pessia, A.; Kokkonen, M.; Velagapudi, V. Validation and automation of a high-throughput multitargeted method for semiquantification of endogenous metabolites from different biological matrices using tandem mass spectrometry. *Metabolites* **2018**, *8*, 44. [[CrossRef](#)] [[PubMed](#)]
82. Spivia, W.R.; Raedschelders, K.; Gomez, M.; Van Eyk, J.E. Automated Metabolite Extraction for Plasma Using the Agilent Bravo Platform. 2019. Available online: <https://www.agilent.com/cs/library/technicaloverviews/public/technicaloverview-metabolomics-sample-prep-bravo-5994-0685en-agilent.pdf> (accessed on 27 February 2020).

83. Akita, S.; Watanabe, K. *New Analytical Intelligence Concept—Support for Automating Analytical Operations*; TR C190-E2; Shimadzu Corp.: Kyoto, Japan, 2019; pp. 1–4.
84. Chen, D.; Wang, Z.; Guo, D.; Orekhov, V.; Qu, X. Review and Prospect: Deep Learning in Nuclear Magnetic Resonance Spectroscopy. *Chem. A Eur. J.* **2020**, *26*, 10391–10401. [[CrossRef](#)] [[PubMed](#)]
85. Rodrigues, A.M.; Ribeiro-Barros, A.I.; Antonio, C. Experimental design and sample preparation in forest tree metabolomics. *Metabolites* **2019**, *9*, 285. [[CrossRef](#)] [[PubMed](#)]
86. Lu, W.; Su, X.; Klein, M.S.; Lewis, I.A.; Fiehn, O.; Rabinowitz, J.D. Metabolite Measurement: Pitfalls to Avoid and Practices to Follow. *Annu. Rev. Biochem.* **2017**, *86*, 277–304. [[CrossRef](#)]
87. Stoll, D.R.; Carr, P.W. Two-Dimensional Liquid Chromatography: A State of the Art Tutorial. *Anal. Chem.* **2017**, *89*, 519–531. [[CrossRef](#)]
88. Keppler, E.A.H.; Jenkins, C.L.; Davis, T.J.; Bean, H.D. Advances in the application of comprehensive two-dimensional gas chromatography in metabolomics. *TrAC Trends Anal. Chem.* **2018**, *109*, 275–286. [[CrossRef](#)]
89. Fouque, K.J.D.; Fernandez-Lima, F. Recent advances in biological separations using trapped ion mobility spectrometry—Mass spectrometry. *TrAC Trends Anal. Chem.* **2019**, *116*, 308–315. [[CrossRef](#)]
90. De Raad, M.; Fischer, C.R.; Northen, T.R. High-throughput platforms for metabolomics. *Curr. Opin. Chem. Biol.* **2016**, *30*, 7–13. [[CrossRef](#)]
91. Dodds, J.N.; Baker, E.S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2185–2195. [[CrossRef](#)] [[PubMed](#)]
92. Luo, M.D.; Zhou, Z.W.; Zhu, Z.J. The Application of Ion Mobility-Mass Spectrometry in Untargeted Metabolomics: From Separation to Identification. *J. Anal. Test.* **2020**, *4*, 163–174. [[CrossRef](#)]
93. Purves, R.W. *Enhancing Biological LC-MS Analyses Using Ion Mobility Spectrometry*; Elsevier Ltd.: Amsterdam, The Netherlands, 2018; Volume 79.
94. Yang, X.; Wei, S.; Liu, B.; Guo, D.; Zheng, B.; Feng, L.; Liu, Y.; Tomas-Barberan, F.A.; Luo, L.; Huang, D. A novel integrated non-targeted metabolomic analysis reveals significant metabolite variations between different lettuce (*Lactuca sativa*. L) varieties. *Hortic. Res.* **2018**, *5*, 1–14. [[CrossRef](#)]
95. McCullagh, M.; Pereira, C.A.M.; Yariwake, J.H. Use of ion mobility mass spectrometry to enhance cumulative analytical specificity and separation to profile 6-C/8-C-glycosylflavone critical isomer pairs and known–unknowns in medicinal plants. *Phytochem. Anal.* **2019**, *30*, 424–436. [[CrossRef](#)]
96. Metz, T.O.; Baker, E.S.; Schymanski, E.L.; Renslow, R.S.; Thomas, D.G.; Causon, T.J.; Webb, I.K.; Hann, S.; Smith, R.D.; Teeguarden, J.G. Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: What can it add and how far can it go? *Bioanalysis* **2017**, *9*, 81–98. [[CrossRef](#)]
97. Lanucara, F.; Holman, S.W.; Gray, C.J.; Evers, C.E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **2014**, *6*, 281–294. [[CrossRef](#)] [[PubMed](#)]
98. Mu, Y.; Schulz, B.L.; Ferro, V. Applications of ion mobility-mass spectrometry in carbohydrate chemistry and glycobiology. *Molecules* **2018**, *23*, 2557. [[CrossRef](#)]
99. Boughton, B.A.; Thinakaran, D.; Sarabia, D.; Bacic, A.; Roessner, U. Mass spectrometry imaging for plant biology: A review. *Phytochem. Rev.* **2016**, *15*, 445–488. [[CrossRef](#)]
100. Qin, L.; Zhang, Y.; Liu, Y.; He, H.; Han, M.; Li, Y.; Zeng, M.; Wang, X. Recent advances in matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI) for in situ analysis of endogenous molecules in plants. *Phytochem. Anal.* **2018**, *29*, 351–364. [[CrossRef](#)]
101. Holzlechner, M.; Eugenin, E.; Prideaux, B. Mass spectrometry imaging to detect lipid biomarkers and disease signatures in cancer. *Cancer Rep.* **2019**, *2*. [[CrossRef](#)]
102. Alexandrov, T. Spatial Metabolomics and Imaging Mass Spectrometry in the Age of Artificial Intelligence. *Annu. Rev. Biomed. Data Sci.* **2020**, *3*, 61–87. [[CrossRef](#)] [[PubMed](#)]
103. Swales, J.G.; Hamm, G.; Clench, M.R.; Goodwin, R.J.A. Mass spectrometry imaging and its application in pharmaceutical research and development: A concise review. *Int. J. Mass Spectrom.* **2019**, *437*, 99–112. [[CrossRef](#)]
104. Snel, M.F. *Ion Mobility Separation Mass Spectrometry Imaging*, 1st ed.; Elsevier B.V.: Amsterdam, The Netherlands, 2019; Volume 83.
105. Fujii, T.; Matsuda, S.; Tejedor, M.L.; Esaki, T.; Sakane, I.; Mizuno, H.; Tsuyama, N.; Masujima, T. Direct metabolomics for plant cells by live single-cell mass spectrometry. *Nat. Protoc.* **2015**, *10*, 1445–1456. [[CrossRef](#)] [[PubMed](#)]
106. Foll, M.C.; Moritz, L.; Wollmann, T.; Stillger, M.N.; Vockert, N.; Werner, M.; Bronsert, P.; Rohr, K.; Grüning, B.A.; Schilling, O. Accessible and reproducible mass spectrometry imaging data analysis in Galaxy. *Gigascience* **2019**, *8*, 1–12. [[CrossRef](#)]
107. Dueñas, M.E.; Larson, E.A.; Lee, Y.J. Toward mass spectrometry imaging in the metabolomics scale: Increasing metabolic coverage through multiple on-tissue chemical modifications. *Front. Plant Sci.* **2019**, *10*, 1–11. [[CrossRef](#)]
108. Silva, D.B.; Turatti, I.C.C.; Gouveia, D.R.; Ernst, M.; Teixeira, S.P.; Lopes, N.P. Mass Spectrometry of flavonoid vicenin-2, based sunlight barriers in *Lychnophora* species. *Sci. Rep.* **2014**, *4*, 1–8. [[CrossRef](#)] [[PubMed](#)]
109. Ernst, M.; Nothias, L.F.; van der Hooff, J.J.J.; Silva, R.R.; Saslis-Lagoudakis, C.H.; Grace, O.M.; Martinez-Swatson, K.; Hassemer, G.; Funez, L.A.; Simonsen, H.T.; et al. Assessing specialized metabolite diversity in the cosmopolitan plant genus *Euphorbia* l. *Front. Plant Sci.* **2019**, *10*, 1–15. [[CrossRef](#)] [[PubMed](#)]

110. Bokhart, M.T.; Nazari, M.; Garrard, K.P.; Muddiman, D.C. MSiReader v1.0: Evolving Open-Source Mass Spectrometry Imaging Software for Targeted and Untargeted Analyses. *J. Am. Soc. Mass Spectrom.* **2018**, *28*, 8–16. [[CrossRef](#)] [[PubMed](#)]
111. Ganesh, S.; Hu, T.; Woods, E.; Allam, M.; Cai, S.; Henderson, W.; Coskun, A.F. Spatially resolved 3D metabolomic profiling in tissues. *Sci. Adv.* **2021**, *7*, 1–17. [[CrossRef](#)]
112. Wang, H.; Chen, L.; Sun, L. Digital microfluidics: A promising technique for biochemical applications. *Front. Mech. Eng.* **2017**, *12*, 510–525. [[CrossRef](#)]
113. Damiani, S.; Kompella, U.B.; Damiani, S.A.; Kodzius, R. Microfluidic devices for drug delivery systems and drug screening. *Genes* **2018**, *9*, 103. [[CrossRef](#)] [[PubMed](#)]
114. Pedde, R.D.; Li, H.; Borchers, C.H.; Akbari, M. Microfluidic-Mass Spectrometry Interfaces for Translational Proteomics. *Trends Biotechnol.* **2017**, *35*, 954–970. [[CrossRef](#)]
115. Feng, D.; Xu, T.; Li, H.; Shi, X.; Xu, G. Single-cell Metabolomics Analysis by Microfluidics and Mass Spectrometry: Recent New Advances. *J. Anal. Test.* **2020**, *4*, 198–209. [[CrossRef](#)]
116. Steckel, A.; Schlosser, G. An organic chemist's guide to electrospray mass spectrometric structure elucidation. *Molecules* **2019**, *24*, 611. [[CrossRef](#)]
117. Stanley, C.E.; Shrivastava, J.; Brugman, R.; Heinzlmann, E.; van Swaay, D.; Grossmann, G. Dual-flow-RootChip reveals local adaptations of roots towards environmental asymmetry at the physiological and genetic levels. *New Phytol.* **2018**, *217*, 1357–1369. [[CrossRef](#)]
118. Fenaille, F.; Barbier Saint-Hilaire, P.; Rousseau, K.; Junot, C. Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? *J. Chromatogr. A* **2017**, *1526*, 1–12. [[CrossRef](#)]
119. Davies, V.; Wandy, J.; Weidt, S.; van der Hooft, J.J.J.; Miller, A.; Daly, R.; Rogers, S. Rapid Development of Improved Data-Dependent Acquisition Strategies. *Anal. Chem.* **2021**, *93*, 5676–5683. [[CrossRef](#)]
120. Wandy, J.; Davies, V.; Van Der Hooft, J.J.J.; Weidt, S.; Daly, R.; Rogers, S. In silico optimization of mass spectrometry fragmentation strategies in metabolomics. *Metabolites* **2019**, *9*, 219. [[CrossRef](#)]
121. Bailey, D.J.; McDevitt, M.T.; Westphall, M.S.; Pagliarini, D.J.; Coon, J.J. Intelligent data acquisition blends targeted and discovery methods. *J. Proteome Res.* **2014**, *13*, 2152–2161. [[CrossRef](#)]
122. Van Der Hooft, J.J.J.; Vervoort, J.; Bino, R.J.; Beekwilder, J.; De Vos, R.C.H. Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation. *Anal. Chem.* **2011**, *83*, 409–416. [[CrossRef](#)]
123. Van der Hooft, J.J.J.; Vervoort, J.; Bino, R.J.; de Vos, R.C.H. Spectral trees as a robust annotation tool in LC-MS based metabolomics. *Metabolomics* **2012**, *8*, 691–703. [[CrossRef](#)]
124. Van der Hooft, J.J.J.; Rankin, N. *Metabolite Identification in Complex Mixtures Using Nuclear Magnetic Resonance Spectroscopy*; Springer: Berlin/Heidelberg, Germany, 2016; ISBN 9783319283883.
125. Wishart, D.S. NMR metabolomics: A look ahead. *J. Magn. Reson.* **2019**, *306*, 155–161. [[CrossRef](#)]
126. Emwas, A.H.; Roy, R.; McKay, R.T.; Tenori, L.; Saccenti, E.; Nagana Gowda, G.A.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; et al. Nmr spectroscopy for metabolomics research. *Metabolites* **2019**, *9*, 123. [[CrossRef](#)]
127. Mishra, S.; Gogna, N.; Dorai, K. NMR-based investigation of the altered metabolic response of *Bougainvillea spectabilis* leaves exposed to air pollution stress during the circadian cycle. *Environ. Exp. Bot.* **2019**, *164*, 58–70. [[CrossRef](#)]
128. Bornet, A.; Maucourt, M.; Deborde, C.; Jacob, D.; Milani, J.; Vuichoud, B.; Ji, X.; Dumez, J.N.; Moing, A.; Bodenhausen, G.; et al. Highly Repeatable Dissolution Dynamic Nuclear Polarization for Heteronuclear NMR Metabolomics. *Anal. Chem.* **2016**, *88*, 6179–6183. [[CrossRef](#)]
129. Coutinho, I.D.; Moraes, T.B.; Mertz-Henning, L.M.; Nepomuceno, A.L.; Giordani, W.; Marcolino-Gomes, J.; Santagneli, S.; Colnago, L.A. Integrating High-Resolution and Solid-State Magic Angle Spinning NMR Spectroscopy and a Transcriptomic Analysis of Soybean Tissues in Response to Water Deficiency. *Phytochem. Anal.* **2017**, *28*, 529–540. [[CrossRef](#)]
130. Van der Hooft, J.J.J.; Mihaleva, V.; De Vos, R.C.H.; Bino, R.J.; Vervoort, J. A strategy for fast structural elucidation of metabolites in small volume plant extracts using automated MS-guided LC-MS-SPE-NMR. *Magn. Reson. Chem.* **2011**, *49*, S55–S60. [[CrossRef](#)]
131. Lima, R.D.C.L.; Gramsbergen, S.M.; Van Staden, J.; Jager, A.K.; Kongstad, K.T.; Staerk, D. Advancing HPLC-PDA-HRMS-SPE-NMR Analysis of Coumarins in *Coleonema album* by Use of Orthogonal Reversed-Phase C18 and Pentafluorophenyl Separations. *J. Nat. Prod.* **2017**, *80*, 1020–1027. [[CrossRef](#)]
132. Hao, J.; Liebeke, M.; Astle, W.; De Iorio, M.; Bundy, J.G.; Ebbels, T.M.D. Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nat. Protoc.* **2014**, *9*, 1416–1427. [[CrossRef](#)]
133. Röhnisch, H.E.; Eriksson, J.; Müllner, E.; Agback, P.; Sandström, C.; Moazzami, A.A. AQuA: An Automated Quantification Algorithm for High-Throughput NMR-Based Metabolomics and Its Application in Human Plasma. *Anal. Chem.* **2018**, *90*, 2095–2102. [[CrossRef](#)]
134. Cañueto, D.; Gómez, J.; Salek, R.M.; Correig, X.; Cañellas, N. rDolphin: A GUI R package for proficient automatic profiling of 1D ¹H-NMR spectra of study datasets. *Metabolomics* **2018**, *14*, 1–5. [[CrossRef](#)]
135. Jonas, E.; Kuhn, S. Rapid prediction of NMR spectral properties with quantified uncertainty. *J. Cheminform.* **2019**, *11*, 1–7. [[CrossRef](#)]
136. Nugroho, A.E.; Morita, H. Computationally-assisted discovery and structure elucidation of natural products. *J. Nat. Med.* **2019**, *73*, 687–695. [[CrossRef](#)]

137. Reher, R.; Kim, H.W.; Zhang, C.; Mao, H.H.; Wang, M.; Nothias, L.F.; Caraballo-Rodriguez, A.M.; Glukhov, E.; Teke, B.; Leao, T.; et al. A Convolutional Neural Network-Based Approach for the Rapid Annotation of Molecularly Diverse Natural Products. *J. Am. Chem. Soc.* **2020**, *142*, 4114–4120. [[CrossRef](#)]
138. Mäkelä, V.; Vaahtera, L.; Helminen, J.; Koskela, H.; Brosché, M.; Kilpeläinen, I.; Heikkinen, S. Automated processing and statistical analysis of nmr spectra obtained from *Arabidopsis thaliana* extracts. *bioRxiv* **2019**. [[CrossRef](#)]
139. Peng, W.K. Clustering Nuclear Magnetic Resonance: Machine learning assistive rapid two-dimensional relaxometry mapping. *Eng. Reports* **2021**, 1–11. [[CrossRef](#)]
140. Goeddel, L.C.; Patti, G.J. Maximizing the value of metabolomic data. *Bioanalysis* **2012**, *4*, 2199–2201. [[CrossRef](#)]
141. Biesecker, L.G. Hypothesis-generating research and predictive medicine. *Genome Res.* **2013**, *23*, 1051–1053. [[CrossRef](#)]
142. Goodacre, R.; Vaidyanathan, S.; Dunn, W.B.; Harrigan, G.G.; Kell, D.B. Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends Biotechnol.* **2004**, *22*, 245–252. [[CrossRef](#)]
143. Boccard, J.; Rudaz, S. Harnessing the complexity of metabolomic data with chemometrics. *J. Chemom.* **2014**, *28*, 1–9. [[CrossRef](#)]
144. Kell, D.B. Metabolomics, machine learning and modelling: Towards an understanding of the language of cells. *Biochem. Soc. Trans.* **2005**, *33*, 520–524. [[CrossRef](#)]
145. Ma, L.; Konter, J.; Herndon, E.; Jin, L.; Steinhofel, G.; Sanchez, D.; Brantley, S. Quantifying an early signature of the industrial revolution from lead concentrations and isotopes in soils of Pennsylvania, USA. *Anthropocene* **2014**, *7*, 16–29. [[CrossRef](#)]
146. Mendez, K.M.; Reinke, S.N.; Broadhurst, D.I. A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification. *Metabolomics* **2019**, *15*. [[CrossRef](#)]
147. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [[CrossRef](#)]
148. Liu, K.; Abdullah, A.A.; Huang, M.; Nishioka, T.; Altaf-Ul-Amin, M.; Kanaya, S. Novel Approach to Classify Plants Based on Metabolite-Content Similarity. *Biomed. Res. Int.* **2017**, *2017*. [[CrossRef](#)]
149. Willett, D.S.; Rering, C.C.; Ardura, D.A.; Beck, J.J. *Application of Mathematical Models and Computation in Plant Metabolomics*; Elsevier Inc.: Amsterdam, The Netherlands, 2018; ISBN 9780128123645.
150. Mendez, K.M.; Broadhurst, D.I.; Reinke, S.N. The application of artificial neural networks in metabolomics: A historical perspective. *Metabolomics* **2019**, *15*, 1–14. [[CrossRef](#)]
151. Heinemann, J. *Machine Learning in Untargeted Metabolomics Experiments*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 1859, ISBN 9781493987573.
152. Nguyen, L.H.; Holmes, S. Ten quick tips for effective dimensionality reduction. *PLoS Comput. Biol.* **2019**, *15*, 1–19. [[CrossRef](#)]
153. Chagas-Paula, D.A.; Oliveira, T.B.; Zhang, T.; Edrada-Ebel, R.; Da Costa, F.B. Prediction of anti-inflammatory plants and discovery of their biomarkers by machine learning algorithms and metabolomic studies. *Planta Med.* **2015**, *81*, 450–458. [[CrossRef](#)]
154. Yang, L.; Xue, Y.; Wei, J.; Dai, Q.; Li, P. Integrating metabolomic data with machine learning approach for discovery of Q - markers from Jinqi Jiangtang preparation against type 2 diabetes. *Chin. Med.* **2021**, *16*, 1–12. [[CrossRef](#)]
155. Boutet, S.; Barreda, L.; Perreau, F.; Mouille, G.; Delannoy, E.; Magniette, M.-L.; Monti, A.; Lepiniec, L.; Zanetti, F.; Corso, M. Untargeted metabolomic analyses reveal the diversity and plasticity of the specialized metabolome in seeds of different *Camelina sativa* varieties. *bioRxiv* **2021**. [[CrossRef](#)]
156. Etalo, D.; de Vos, R.C.; Joosten, M.H.A.J.; Hall, R. Spatially-resolved plant metabolomics: Some potentials and limitations of Laser-Ablation Electrospray Ionization (LAESI) Mass Spectrometry metabolite imaging. *Plant Physiol.* **2015**, *169*, 1424–1435. [[CrossRef](#)]
157. Fürtauer, L.; Pschenitschnigg, A.; Scharhosi, H.; Weckwerth, W.; Nägele, T. Combined multivariate analysis and machine learning reveals a predictive module of metabolic stress response in *Arabidopsis thaliana*. *Mol. Omics* **2018**, *14*, 437–449. [[CrossRef](#)]
158. Ataş, M.; Yardimci, Y.; Temizel, A. A new approach to aflatoxin detection in chili pepper by machine vision. *Comput. Electron. Agric.* **2012**, *87*, 129–141. [[CrossRef](#)]
159. Scott, I.M.; Vermeer, C.P.; Liakata, M.; Corol, D.I.; Ward, J.L.; Lin, W.; Johnson, H.E.; Whitehead, L.; Kular, B.; Baker, J.M.; et al. Enhancement of plant metabolite fingerprinting by machine learning. *Plant Physiol.* **2010**, *153*, 1506–1520. [[CrossRef](#)]
160. de Oliveira Almeida, R.; Valente, G.T. Predicting metabolic pathways of plant enzymes without using sequence similarity: Models from machine learning. *Plant Genome* **2020**. [[CrossRef](#)]
161. Toubiana, D.; Puzis, R.; Wen, L.; Sikron, N.; Kurmanbayeva, A.; Soltabayeva, A.; del Mar Rubio Wilhelmi, M.; Sade, N.; Fait, A.; Sagi, M.; et al. Combined network analysis and machine learning allows the prediction of metabolic pathways from tomato metabolomics data. *Commun. Biol.* **2019**, *2*. [[CrossRef](#)]
162. Grapov, D.; Fahrman, J.; Wanichthanarak, K.; Khoomrung, S. Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. *Omics A J. Integr. Biol.* **2018**, *22*, 630–636. [[CrossRef](#)]
163. Ching, T.; Himmelstein, D.S.; Beaulieu-Jones, B.K.; Kalinin, A.A.; Do, B.T.; Way, G.P.; Ferrero, E.; Agapow, P.M.; Zietz, M.; Hoffman, M.M.; et al. Opportunities and Obstacles for Deep Learning in Biology and Medicine. *J. R. Soc. Interface* **2018**, *15*. [[CrossRef](#)]
164. Bueno, P.C.P.; Lopes, N.P. Metabolomics to Characterize Adaptive and Signaling Responses in Legume Crops under Abiotic Stresses. *ACS Omega* **2020**, *5*, 1752–1763. [[CrossRef](#)]

165. Jarrin, E.P.; Cordeiro, F.B.; Medranda, W.C.; Barrett, M.; Zambrano, M.; Regato, M. A Machine Learning-Based algorithm for the assessment of clinical metabolomic fingerprints in Zika virus disease. In Proceedings of the 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Guayaquil, Ecuador, 11–15 November 2019; pp. 8–13. [\[CrossRef\]](#)
166. Silva, J.C.F.; Teixeira, R.M.; Silva, F.F.; Brommonschenkel, S.H.; Fontes, E.P.B. Machine learning approaches and their current application in plant molecular biology: A systematic review. *Plant Sci.* **2019**, *284*, 37–47. [\[CrossRef\]](#)
167. Heinemann, J.; Mazurie, A.; Tokmina-Lukaszewska, M.; Beilman, G.J.; Bothner, B. Application of support vector machines to metabolomics experiments with limited replicates. *Metabolomics* **2014**, *10*, 1121–1128. [\[CrossRef\]](#)
168. Gromski, P.S.; Muhamadali, H.; Ellis, D.I.; Xu, Y.; Correa, E.; Turner, M.L.; Goodacre, R. A tutorial review: Metabolomics and partial least squares-discriminant analysis—A marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **2015**, *879*, 10–23. [\[CrossRef\]](#)
169. Gu, Q.; Han, J. Clustered support vector machines. *J. Mach. Learn. Res.* **2013**, *31*, 307–315.
170. Carrizosa, E.; Nogales-Gómez, A.; Romero Morales, D. Clustering categories in support vector machines. *Omega* **2017**, *66*, 28–37. [\[CrossRef\]](#)
171. Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A.A.; Melnik, A.V.; Meusel, M.; Dorrestein, P.C.; Rousu, J.; Böcker, S. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **2019**, *16*, 299–302. [\[CrossRef\]](#)
172. Pan, H.; Yao, C.; Yao, S.; Yang, W.; Wu, W.; Guo, D. A metabolomics strategy for authentication of plant medicines with multiple botanical origins, a case study of *Uncariae Rammulus Cum Uncis*. *J. Sep. Sci.* **2020**, *43*, 1043–1050. [\[CrossRef\]](#)
173. Du, H.; Huo, Y.; Liu, H.; Kamal, G.M.; Yang, J.; Zeng, Y.; Zhao, S.; Liu, Y. Fast nutritional characterization of different pigmented rice grains using a combination of NMR and decision tree analysis. *J. Food* **2019**, *17*, 128–136. [\[CrossRef\]](#)
174. Niazian, M.; Niedbała, G. Machine learning for plant breeding and biotechnology. *Agriculture* **2020**, *10*, 436. [\[CrossRef\]](#)
175. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [\[CrossRef\]](#) [\[PubMed\]](#)
176. Mekonnen, Y.; Namuduri, S.; Burton, L.; Sarwat, A.; Bhansali, S. Review—Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture. *J. Electrochem. Soc.* **2020**, *167*, 037522. [\[CrossRef\]](#)
177. Hummel, J.; Strehmel, N.; Selbig, J.; Walther, D.; Kopka, J. Decision tree supported substructure prediction of metabolites from GC-MS profiles. *Metabolomics* **2010**, *6*, 322–333. [\[CrossRef\]](#)
178. De Bruijn, W.J.C.; van Dinteren, S.; Gruppen, H.; Vincken, J.P. Mass spectrometric characterisation of avenanthramides and enhancing their production by germination of oat (*Avena sativa*). *Food Chem.* **2019**, *277*, 682–690. [\[CrossRef\]](#)
179. Dos Santos, F.A.; Sousa, I.P.; Furtado, N.A.J.C.; Da Costa, F.B. Combined OPLS-DA and decision tree as a strategy to identify antimicrobial biomarkers of volatile oils analyzed by gas chromatography–mass spectrometry. *Rev. Bras. Farmacogn.* **2018**, *28*, 647–653. [\[CrossRef\]](#)
180. Sok, H.K.; Ooi, M.P.L.; Kuang, Y.C.; Demidenko, S. Multivariate alternating decision trees. *Pattern Recognit.* **2016**, *50*, 195–209. [\[CrossRef\]](#)
181. Alakwaa, F.M.; Chaudhary, K.; Garmire, L.X. Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. *J. Proteome Res.* **2018**, *17*, 337–347. [\[CrossRef\]](#) [\[PubMed\]](#)
182. Touw, W.G.; Bayjanov, J.R.; Overmars, L.; Backus, L.; Boekhorst, J.; Wels, M.; Sacha van Hijum, A.F.T. Data mining in the life science with random forest: A walk in the park or lost in the jungle? *Brief. Bioinform.* **2013**, *14*, 315–326. [\[CrossRef\]](#)
183. Ansarifard, J.; Akhavadegan, F.; Wang, L. Performance prediction of crosses in plant breeding through genotype by environment interactions. *Sci. Rep.* **2020**, *10*, 1–11. [\[CrossRef\]](#) [\[PubMed\]](#)
184. Lim, D.K.; Mo, C.; Lee, J.H.; Long, N.P.; Dong, Z.; Li, J.; Lim, J.; Kwon, S.W. The integration of multi-platform MS-based metabolomics and multivariate analysis for the geographical origin discrimination of *Oryza sativa* L. *J. Food Drug Anal.* **2018**, *26*, 769–777. [\[CrossRef\]](#)
185. Oza, V.H.; Aicher, J.K.; Reed, L.K. Random forest analysis of untargeted metabolomics data suggests increased use of omega fatty acid oxidation pathway in drosophila melanogaster larvae fed a medium chain fatty acid rich high-fat diet. *Metabolites* **2019**, *9*, 5. [\[CrossRef\]](#)
186. Lima, E.D.O.; Navarro, L.C.; Morishita, K.N.; Kamikawa, C.M.; Rodrigues, R.G.M.; Dabaja, M.Z.; de Oliveira, D.N.; Delafiori, J.; Dias-Audibert, F.L.; Ribeiro, M.d.S.; et al. Metabolomics and Machine Learning Approaches Combined in Pursuit for More Accurate Paracoccidioidomycosis Diagnoses. *mSystems* **2020**, *5*, 1–12. [\[CrossRef\]](#)
187. Chen, T.; Cao, Y.; Zhang, Y.; Liu, J.; Bao, Y.; Wang, C.; Jia, W.; Zhao, A. Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection. Evidence-based Complement. *Altern. Med.* **2013**. [\[CrossRef\]](#)
188. Antonelli, J.; Claggett, B.L.; Henglin, M.; Kim, A.; Ovsak, G.; Kim, N.; Deng, K.; Rao, K.; Tyagi, O.; Watrous, J.D.; et al. Statistical workflow for feature selection in human metabolomics data. *Metabolites* **2019**, *9*, 143. [\[CrossRef\]](#)
189. Kelly, R.S.; McGeachie, M.J.; Lee-Sarwar, K.A.; Kachroo, P.; Chu, S.H.; Virkud, Y.V.; Huang, M.; Litonjua, A.A.; Weiss, S.T.; Lasky-Su, J. Partial least squares discriminant analysis and bayesian networks for metabolomic prediction of childhood asthma. *Metabolites* **2018**, *8*, 68. [\[CrossRef\]](#)
190. Gillies, C.E.; Jennaro, T.S.; Puskarich, M.A.; Sharma, R.; Ward, K.R.; Fan, X.; Jones, A.E.; Stringer, K.A. A multilevel bayesian approach to improve effect size estimation in regression modeling of metabolomics data utilizing imputation with uncertainty. *Metabolites* **2020**, *10*, 319. [\[CrossRef\]](#)

191. Teklehaymanot, F.K.; Muma, M.; Zoubir, A.M. Bayesian Cluster Enumeration Criterion for Unsupervised Learning. *IEEE Trans. Signal Process.* **2018**, *66*, 5392–5406. [[CrossRef](#)]
192. Del Carratore, F.; Schmidt, K.; Vinaixa, M.; Hollywood, K.A.; Greenland-Bews, C.; Takano, E.; Rogers, S.; Breitling, R. Integrated Probabilistic Annotation: A Bayesian-Based Annotation Method for Metabolomic Profiles Integrating Biochemical Connections, Isotope Patterns, and Adduct Relationships. *Anal. Chem.* **2019**, *91*, 12799–12807. [[CrossRef](#)]
193. Ludwig, M.; Dührkop, K.; Böcker, S. Bayesian networks for mass spectrometric metabolite identification via molecular fingerprints. *Bioinformatics* **2018**, *34*, i333–i340. [[CrossRef](#)]
194. Ludwig, M.; Nothias, L.F.; Dührkop, K.; Koester, I.; Fleischauer, M.; Hoffmann, M.A.; Petras, D.; Vargas, F.; Morsy, M.; Aluwihare, L.; et al. Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nat. Mach. Intell.* **2020**, *2*, 629–641. [[CrossRef](#)]
195. O'Brien, K.A.; Atkinson, R.A.; Richardson, L.; Kou, A.; Murray, A.J.; Harr, S.D.R.; Martin, D.S.; Levett, D.Z.H.; Mitchell, K.; Mythen, M.G.; et al. Metabolomic and lipidomic plasma profile changes in human participants ascending to Everest Base Camp. *Sci. Rep.* **2019**, *9*, 1–12. [[CrossRef](#)]
196. McGeachie, M.J.; Chang, H.H.; Weiss, S.T. CGBayesNets: Conditional Gaussian Bayesian Network Learning and Inference with Mixed Discrete and Continuous Data. *PLoS Comput. Biol.* **2014**, *10*. [[CrossRef](#)]
197. Pomyen, Y.; Wanichthanarak, K.; Pongsombat, P.; Fahrman, J.; Grapov, D.; Khoomrung, S. Deep metabolome: Applications of deep learning in metabolomics. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 2818–2825. [[CrossRef](#)]
198. O'Shea, K.; Cameron, S.J.S.; Lewis, K.E.; Lu, C.; Mur, L.A.J. Metabolomic-based biomarker discovery for non-invasive lung cancer screening: A case study. *Biochim. Biophys. Acta Gen. Subj.* **2016**, *1860*, 2682–2687. [[CrossRef](#)]
199. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [[CrossRef](#)]
200. Liebal, U.W.; Phan, A.N.T.; Sudhakar, M.; Raman, K.; Blank, L.M. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites* **2020**, *10*, 243. [[CrossRef](#)]
201. Melnikov, A.D.; Tsentelovich, Y.P.; Yanshole, V.V. Deep Learning for the Precise Peak Detection in High-Resolution LC-MS Data. *Anal. Chem.* **2020**, *92*, 588–592. [[CrossRef](#)]
202. Zhang, X.; Lin, T.; Xu, J.; Luo, X.; Ying, Y. DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis. *Anal. Chim. Acta* **2019**, *1058*, 48–57. [[CrossRef](#)]
203. Golhani, K.; Balasundram, S.K.; Vadamalai, G.; Pradhan, B. A review of neural networks in plant disease detection using hyperspectral data. *Inf. Process. Agric.* **2018**, *5*, 354–371. [[CrossRef](#)]
204. Jia, Y.; Zhao, R.; Chen, L. Similarity-Based Machine Learning Model for Predicting the Metabolic Pathways of Compounds. *IEEE Access* **2020**, *8*, 130687–130696. [[CrossRef](#)]
205. Wang, P.; Moore, B.M.; Uygun, S.; Lehti-Shiu, M.D.; Barry, C.; Shiu, S.-H. Optimizing the use of gene expression data to predict metabolic pathway memberships with unsupervised and supervised machine learning. *bioRxiv* **2020**. [[CrossRef](#)]
206. Presnell, K.V.; Alper, H.S. Systems Metabolic Engineering Meets Machine Learning: A New Era for Data-Driven Metabolic Engineering. *Biotechnol. J.* **2019**, *14*. [[CrossRef](#)]
207. Volk, M.J.; Lourentzou, I.; Mishra, S.; Vo, L.T.; Zhai, C.; Zhao, H. Biosystems Design by Machine Learning. *ACS Synth. Biol.* **2020**, *9*, 1514–1533. [[CrossRef](#)]
208. Nagaraja, A.A.; Fontaine, N.; Delsaut, M.; Charton, P.; Damour, C.; Offmann, B.; Grondin-Perez, B.; Cadet, F. Flux prediction using artificial neural network (ANN) for the upper part of glycolysis. *PLoS ONE* **2019**, *14*, 10–12. [[CrossRef](#)]
209. Zampieri, G.; Vijayakumar, S.; Yaneske, E.; Angione, C. Machine and deep learning meet genome-scale metabolic modeling. *PLoS Comput. Biol.* **2019**, *15*, 1–24. [[CrossRef](#)]
210. Costello, Z.; Martin, H.G. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ Syst. Biol. Appl.* **2018**, *4*, 1–14. [[CrossRef](#)] [[PubMed](#)]
211. Volkova, S.; Matos, M.R.A.; Mattanovich, M.; de Mas, I.M. Metabolic modelling as a framework for metabolomics data integration and analysis. *Metabolites* **2020**, *10*, 303. [[CrossRef](#)]
212. Baranwal, M.; Magner, A.; Elvati, P.; Saldinger, J.; Violi, A.; Violi, A.; Hero, A.O. A deep learning architecture for metabolic pathway prediction. *Bioinformatics* **2020**, *36*, 2547–2553. [[CrossRef](#)] [[PubMed](#)]
213. Yang, J.H.; Wright, S.N.; Hamblin, M.; McCloskey, D.; Alcantar, M.A.; Schrübbers, L.; Lopatkin, A.J.; Satish, S.; Nili, A.; Palsson, B.O.; et al. A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell* **2019**, *177*, 1649–1661. [[CrossRef](#)]
214. Cuperlovic-Culf, M. Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites* **2018**, *8*, 4. [[CrossRef](#)]
215. Rana, P.; Berry, C.; Ghosh, P.; Fong, S.S. Recent advances on constraint-based models by integrating machine learning. *Curr. Opin. Biotechnol.* **2020**, *64*, 85–91. [[CrossRef](#)] [[PubMed](#)]
216. Hosseini, R.; Hassanpour, N.; Liu, L.P.; Hassoun, S. Pathway activity analysis and metabolite annotation for untargeted metabolomics using probabilistic modeling. *Metabolites* **2020**, *10*, 183. [[CrossRef](#)] [[PubMed](#)]
217. Fang, X.; Liu, Y.; Ren, Z.; Du, Y.; Huang, Q.; Garmire, L.X. LiliKoi V2.0: A deep learning-enabled, personalized pathway-based R package for diagnosis and prognosis predictions using metabolomics data. *Gigascience* **2021**, *10*, 1–11. [[CrossRef](#)]

218. Wang, S.; Alseekh, S.; Fernie, A.R.; Luo, J. The Structure and Function of Major Plant Metabolite Modifications. *Mol. Plant* **2019**, *12*, 899–919. [[CrossRef](#)]
219. Koch, M.; Duigou, T.; Faulon, J.L. Reinforcement learning for bioretrosynthesis. *ACS Synth. Biol.* **2020**, *9*, 157–168. [[CrossRef](#)]
220. McLuskey, K.; Wandy, J.; Vincent, I.; van der Hooft, J.J.J.; Rogers, S.; Burgess, K.; Daly, R. Ranking Metabolite Sets by Their Activity Levels. *Metabolites* **2021**, *11*, 103. [[CrossRef](#)]
221. Scheubert, K.; Hufsky, F.; Petras, D.; Wang, M.; Nothias, L.F.; Dührkop, K.; Bandeira, N.; Dorrestein, P.C.; Böcker, S. Significance estimation for large scale metabolomics annotations by spectral matching. *Nat. Commun.* **2017**, *8*. [[CrossRef](#)]
222. Rawlinson, C.; Jones, D.; Rakshit, S.; Meka, S.; Moffat, C.S.; Moolhuijzen, P. Hierarchical clustering of MS/MS spectra from the firefly metabolome identifies new lucibufagin compounds. *Sci. Rep.* **2020**, *10*, 1–9. [[CrossRef](#)]
223. 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](#)]
224. Quinn, R.A.; Nothias, L.F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P.C. Molecular Networking as a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends Pharmacol. Sci.* **2017**, *38*, 143–154. [[CrossRef](#)]
225. De Souza, L.P.; Alseekh, S.; Brotman, Y.; Fernie, A.R. Network-based strategies in metabolomics data analysis and interpretation: From molecular networking to biological interpretation. *Expert Rev. Proteomics* **2020**, *17*, 243–255. [[CrossRef](#)] [[PubMed](#)]
226. Van Der Hooft, J.J.J.; Mohimani, H.; Bauermeister, A.; Dorrestein, P.C.; Duncan, K.R.; Medema, M.H. Linking genomics and metabolomics to chart specialized metabolic diversity. *Chem. Soc. Rev.* **2020**, *49*, 3297–3314. [[CrossRef](#)] [[PubMed](#)]
227. Bai, Y.; Jia, Q.; Su, W.; Yan, Z.; Situ, W.; He, X.; Peng, W.; Yao, H. Integration of molecular networking and fingerprint analysis for studying constituents in *Microctis Folium*. *PLoS ONE* **2020**, *15*, 1–14. [[CrossRef](#)] [[PubMed](#)]
228. Demarque, D.P.; Dusi, R.G.; de Sousa, F.D.M.; Grossi, S.M.; Silvério, M.R.S.; Lopes, N.P.; Espindola, L.S. Mass spectrometry-based metabolomics approach in the isolation of bioactive natural products. *Sci. Rep.* **2020**, *10*, 1–9. [[CrossRef](#)] [[PubMed](#)]
229. Aron, A.T.; Gentry, E.C.; McPhail, K.L.; Nothias, L.F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J.M.; Sikora, N.; Vargas, F.; et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* **2020**, *15*, 1954–1991. [[CrossRef](#)] [[PubMed](#)]
230. Wandy, J.; Zhu, Y.; Van Der Hooft, J.J.J.; Daly, R.; Barrett, M.P.; Rogers, S. Ms2lda.org: Web-based topic modelling for substructure discovery in mass spectrometry. *Bioinformatics* **2018**, *34*, 317–318. [[CrossRef](#)]
231. Rogers, S.; Ong, C.W.; Wandy, J.; Ernst, M.; Ridder, L.; Van Der Hooft, J.J.J. Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. *Faraday Discuss.* **2019**, *218*, 284–302. [[CrossRef](#)]
232. Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites* **2018**, *8*, 31. [[CrossRef](#)]
233. Li, Y.; Kuhn, M.; Gavin, A.C.; Bork, P. Identification of metabolites from tandem mass spectra with a machine learning approach utilizing structural features. *Bioinformatics* **2020**, *36*, 1213–1218. [[CrossRef](#)]
234. Nothias, L.F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; et al. Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **2020**, *17*, 905–908. [[CrossRef](#)]
235. Ludwig, M.; Fleischauer, M.; Dührkop, K.; Hoffmann, M.A.; Böcker, S. De Novo Molecular Formula Annotation and Structure Elucidation Using SIRIUS 4. *Methods Mol. Biol.* **2020**, *2104*, 185–207. [[CrossRef](#)] [[PubMed](#)]
236. Huber, F.; Ridder, L.; Rogers, S.; van der Hooft, J.J.J. Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships. *PLoS Comput. Biol.* **2021**, *17*. [[CrossRef](#)]
237. Ernst, M.; Bin Kang, K.; Carballo-Rodríguez, A.M.; Nothias, L.F.; Wandy, J.; Wang, M.; Rogers, S.; Medema, M.H.; Dorrestein, P.C.; van der Hooft, J.J.J. MolNetEnhancer: Enhanced molecular networks by integrating metabolome mining and annotation tools. *Metabolites* **2019**, *9*, 144. [[CrossRef](#)]
238. Huber, F.; van der Burg, S.; van der Hooft, J.J.J.; Ridder, L. MS2DeepScore—A novel deep learning similarity measure for mass fragmentation spectrum comparisons. *bioRxiv* **2021**. [[CrossRef](#)]
239. Feunang, Y.D.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; et al. ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **2016**, *8*, 1–20. [[CrossRef](#)]
240. Sha, B.; Schymanski, E.L.; Ruttkies, C.; Cousins, I.T.; Wang, Z. Exploring open cheminformatics approaches for categorizing per- and polyfluoroalkyl substances (PFASs). *Environ. Sci. Process. Impacts* **2019**, *21*, 1835–1851. [[CrossRef](#)]
241. Dührkop, K.; Nothias, L.F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M.A.; Petras, D.; Gerwick, W.H.; Rousu, J.; Dorrestein, P.C.; et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **2020**. [[CrossRef](#)]
242. Kim, H.W.; Wang, M.; Leber, C.A.; Nothias, L. NPClassifier: Deep neural structural classification tool for natural products. *ChemRxiv* **2020**. [[CrossRef](#)]
243. Hastings, J.; Glauer, M.; Memariani, A.; Neuhaus, F.; Mossakowski, T. Learning Chemistry: Exploring the suitability of machine learning for the task of structure-based chemical ontology classification. *J. Cheminform.* **2021**, *13*, 1–20. [[CrossRef](#)]
244. Dührkop, K.; Nothias, L.F.; Fleischauer, M.; Ludwig, M.; Hoffmann, M.A.; Rousu, J.; Dorrestein, P.C.; Böcker, S. Classes for the masses: Systematic classification of unknowns using fragmentation spectra. *bioRxiv* **2020**, 1–42. [[CrossRef](#)]

245. Beniddir, M.A.; Kang, K.B.; Genta-Jouve, G.; Huber, F.; Rogers, S.; van der Hooft, J.J.J. Advances in decomposing complex metabolite mixtures using substructure- and network-based computational metabolomics approaches. *Nat. Prod. Rep.* **2021**. [[CrossRef](#)]
246. Wang, M.; Jarmusch, A.K.; Vargas, F.; Aksenov, A.A.; Gauglitz, J.M.; Weldon, K.; Petras, D.; da Silva, R.; Quinn, R.; Melnik, A.V.; et al. Mass spectrometry searches using MASST. *Nat. Biotechnol.* **2020**, *38*, 23–26. [[CrossRef](#)]
247. Jarmusch, A.K.; Wang, M.; Aceves, C.M.; Advani, R.S.; Aguirre, S.; Aksenov, A.A.; Aleti, G.; Aron, A.T.; Bauermeister, A.; Bolleddu, S.; et al. ReDU: A framework to find and reanalyze public mass spectrometry data. *Nat. Methods* **2020**, *17*, 901–904. [[CrossRef](#)]
248. Abdelraheem, A.; Esmaeili, N.; O’Connell, M.; Zhang, J. Progress and perspective on drought and salt stress tolerance in cotton. *Ind. Crops Prod.* **2019**, *130*, 118–129. [[CrossRef](#)]
249. Ullah, N.; Yuce, M.; Neslihan Ozturk Gokce, Z.; Budak, H. Comparative metabolite profiling of drought stress in roots and leaves of seven Triticeae species. *BMC Genomics* **2017**, *18*, 1–12. [[CrossRef](#)]
250. Wang, X.; Cai, X.; Xu, C.; Wang, Q.; Dai, S. Drought-responsive mechanisms in plant leaves revealed by proteomics. *Int. J. Mol. Sci.* **2016**, *17*, 1706. [[CrossRef](#)]
251. Rasmussen, M.W.; Roux, M.; Petersen, M.; Mundy, J. MAP Kinase Cascades in Arabidopsis Innate Immunity. *Front. Plant Sci.* **2012**, *3*, 1–6. [[CrossRef](#)]
252. Pandey, G.K. Emergence of a novel calcium signaling pathway in plants: CBL-CIPK signaling network. *Physiol. Mol. Biol. Plants* **2008**, *14*, 51–68. [[CrossRef](#)]
253. Mühlenbock, P.; Szechyńska-Hebda, M.; Płaszczycza, M.; Baudo, M.; Mullineaux, P.M.; Parker, J.E.; Karpińska, B.; Karpińskie, S. Chloroplast signaling and lesion simulating disease1 regulate crosstalk between light acclimation and immunity in Arabidopsis. *Plant Cell* **2008**, *20*, 2339–2356. [[CrossRef](#)]
254. Ahmad, P.; Prasad, M.N.V. *Environmental Adaptations and Stress Tolerance of Plants in the Era of Climate Change*; Springer: New York, NY, USA, 2012; ISBN 978-1-4614-0814-7.
255. Poltronieri, P.; Bonsegna, S.; de Domenico, S.; Santino, A. Molecular Mechanisms in Plant Abiotic Stress Response. *Ratarstvo i Povrtarstvo Field Veg. Crop Res.* **2011**, *48*, 15–24. [[CrossRef](#)]
256. Michaletti, A.; Naghavi, M.R.; Toorchi, M.; Zolla, L.; Rinalducci, S. Metabolomics and proteomics reveal drought-stress responses of leaf tissues from spring-wheat. *Sci. Rep.* **2018**, *8*, 1–18. [[CrossRef](#)]
257. Gupta, P.; De, B. Metabolomics analysis of rice responses to salinity stress revealed elevation of serotonin, and gentisic acid levels in leaves of tolerant varieties. *Plant Signal. Behav.* **2017**, *12*, 1–11. [[CrossRef](#)]
258. Negrão, S.; Schmöckel, S.M.; Tester, M. Evaluating physiological responses of plants to salinity stress. *Ann. Bot.* **2017**, *119*, 1–11. [[CrossRef](#)]
259. Guo, Y.; Halfter, U.; Ishitani, M.; Zhu, J.-K. Molecular Characterization of Functional Domains in the Protein Kinase SOS2 That Is Required for Plant Salt Tolerance. *Plant Cell* **2007**, *13*, 1383. [[CrossRef](#)]
260. Pan, J.; Li, Z.; Dai, S.; Ding, H.; Wang, Q.; Li, X.; Ding, G.; Wang, P.; Guan, Y.; Liu, W. Integrative analyses of transcriptomics and metabolomics upon seed germination of foxtail millet in response to salinity. *Sci. Rep.* **2020**, *10*, 1–16. [[CrossRef](#)]
261. Ma, X.; Zheng, J.; Zhang, X.; Hu, Q.; Qian, R. Salicylic acid alleviates the adverse effects of salt stress on dianthus superbus (Caryophyllaceae) by activating photosynthesis, protecting morphological structure, and enhancing the antioxidant system. *Front. Plant Sci.* **2017**, *8*, 1–13. [[CrossRef](#)] [[PubMed](#)]
262. Misra, N.; Saxena, P. Effect of salicylic acid on proline metabolism in lentil grown under salinity stress. *Plant Sci.* **2009**, *177*, 181–189. [[CrossRef](#)]
263. Conde, A.; Silva, P.; Agasse, A.; Conde, C.; Gerós, H. Mannitol transport and mannitol dehydrogenase activities are coordinated in *Olea europaea* under salt and osmotic stresses. *Plant Cell Physiol.* **2011**, *52*, 1766–1775. [[CrossRef](#)]
264. Kimura, S.; Kaya, H.; Kawarazaki, T.; Hiraoka, G.; Senzaki, E.; Michikawa, M.; Kuchitsu, K. Protein phosphorylation is a prerequisite for the Ca²⁺-dependent activation of Arabidopsis NADPH oxidases and may function as a trigger for the positive feedback regulation of Ca²⁺ and reactive oxygen species. *Biochim. Biophys. Acta Mol. Cell Res.* **2012**, *1823*, 398–405. [[CrossRef](#)]
265. Nounjan, N.; Nghia, P.T.; Theerakulpisut, P. Exogenous proline and trehalose promote recovery of rice seedlings from salt-stress and differentially modulate antioxidant enzymes and expression of related genes. *J. Plant Physiol.* **2012**, *169*, 596–604. [[CrossRef](#)] [[PubMed](#)]
266. Kaya, C.; Sonmez, O.; Aydemir, S.; Ashraf, M.; Dikilitas, M. Exogenous application of mannitol and thiourea regulates plant growth and oxidative stress responses in salt-stressed maize (*Zea mays* L.). *J. Plant Interact.* **2013**, *8*, 234–241. [[CrossRef](#)]
267. Siahpoosh, M.R.; Sanchez, D.H.; Schlereth, A.; Scofield, G.N.; Furbank, R.T.; van Dongen, J.T.; Kopka, J. Modification of OsSUT1 gene expression modulates the salt response of rice *Oryza sativa* cv. Taipei 309. *Plant Sci.* **2012**, *182*, 101–111. [[CrossRef](#)]
268. Aldesuquy, H.S.; Abbas, M.A.; Hamed, S.A.A.; Elhakem, A.H.; Alsokari, S.S. Glycine betaine and salicylic acid induced modification in productivity of two different cultivars of wheat grown under water stress. *J. Stress Physiol. Biochem.* **2012**, *8*, 72–89.
269. Manaa, A.; Gharbi, E.; Mimouni, H.; Wasti, S.; Aschi-Smiti, S.; Lutts, S.; Ben Ahmed, H. Simultaneous application of salicylic acid and calcium improves salt tolerance in two contrasting tomato (*Solanum lycopersicum*) cultivars. *S. Afr. J. Bot.* **2014**, *95*, 32–39. [[CrossRef](#)]
270. Cvikrová, M.; Gemperlová, L.; Martincová, O.; Vanková, R. Effect of drought and combined drought and heat stress on polyamine metabolism in proline-over-producing tobacco plants. *Plant Physiol. Biochem.* **2013**, *73*, 7–15. [[CrossRef](#)]

-
271. Sarafraz-Ardakani, M.-R.; Khavari-Nejad, R.-A.; Moradi, F.; Najafi, F. Abscisic Acid and Cytokinin-Induced Osmotic and Antioxidant Regulation in Two Drought-Tolerant and Drought-Sensitive Cultivars of Wheat During Grain Filling Under Water Deficit in Field Conditions. *Not. Sci. Biol.* **2014**, *6*, 354–362. [[CrossRef](#)]
 272. Sarker, U.; Oba, S. Drought stress enhances nutritional and bioactive compounds, phenolic acids and antioxidant capacity of Amaranthus leafy vegetable. *BMC Plant Biol.* **2018**, *18*, 258. [[CrossRef](#)]
 273. Varela, M.C.; Arslan, I.; Reginato, M.A.; Cenzano, A.M.; Luna, M.V. Phenolic compounds as indicators of drought resistance in shrubs from Patagonian shrublands (Argentina). *Plant Physiol. Biochem.* **2016**, *104*, 81–91. [[CrossRef](#)]