www.jpnim.com Open Access eISSN: 2281-0692
Journal of Pediatric and Neonatal Individualized Medicine 2024;13(1):e130105
doi: 10.7363/130105
Received: 2024 Jan 03; revised: 2024 Feb 04; accepted: 2024 Mar 06; published online: 2024 Mar 11

Review

Unlocking the secrets of metabolomics with Artificial Intelligence: a comprehensive literature review

Karolina Krystyna Kopeć[1], Federico Cannas[1], Cristina Piras[2], Martina Spada[2], Antonio Noto[2], Luigi Atzori[2], Vassilios Fanos[3]

[1]Department of Mechanical, Chemical and Materials Engineering, University of Cagliari, Cagliari, Italy
[2]Department of Biomedical Sciences, University of Cagliari, Cagliari, Italy
[3]Department of Surgical Sciences, University of Cagliari, Cagliari, Italy

Abstract

This comprehensive review synthesizes the wealth of scientific literature pertaining to the application of Artificial Intelligence (AI) in the field of metabolomics. Over the past decade, AI has played an increasingly pivotal role in deciphering the complexities of metabolomic data, offering novel insights into the molecular underpinnings of biological systems. Through an extensive examination of relevant research papers, we provide a comprehensive overview of the diverse AI techniques and methodologies, from data preprocessing and feature selection to predictive modeling and pathway analysis, employed in metabolomics studies. The review dissects key trends and advancements in AI-driven metabolomics, shedding light on its pivotal role in biomarker discovery, disease diagnosis, and personalized medicine. In addition to highlighting the significant contributions of AI to metabolomics, emerging frontiers will be explored, such as the

incorporation of multi-omics data integration and the growing importance of explainable AI in biological research. Ultimately, this review underscores the transformative impact of AI on metabolomics, emphasizing its potential to reshape our understanding of metabolic pathways, disease mechanisms, and therapeutic interventions. The combination of AI and metabolomics stands as a powerful paradigm shift with far-reaching implications for advancing both fundamental scientific knowledge and practical applications across diverse domains.

Keywords

Metabolomics, Artificial Intelligence, Machine Learning, biomarkers.

Corresponding author

Karolina Krystyna Kopeć, Department of Mechanical, Chemical and Materials Engineering, University of Cagliari, Cagliari, Italy; email: k.kopec@studenti.unica.it.

How to cite

Kopeć KK, Cannas F, Piras C, Spada M, Noto A, Atzori L, Fanos V. Unlocking the secrets of metabolomics with Artificial Intelligence: a comprehensive literature review. J Pediatr Neonat Individual Med. 2024;13(1):e130105. doi: 10.7363/130105.

Introduction

Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) stands at the forefront of modern technological advancements, driven by the aspiration to replicate human-like intelligence within computer systems. At its core, Machine Learning (ML) serves as the pivotal toolset underpinning the quest to achieve this ambitious goal. ML encompasses a diverse array of algorithms designed to facilitate the emulation of cognitive processes [1]. These algorithms possess the remarkable ability to autonomously collect insights from data, recognize complex underlying patterns, and process unstructured data types that stand up to conventional statistical techniques. Such unstructured data, including free-text documents, images, videos, and audio recordings, has the potential to significantly enhance the quantity and possibly the quality of available information, thereby reinforcing predictive capabilities. Thanks to ML algorithms in the AI environment, systems have been developed that demonstrate exceptional performance for specific tasks [2]. ML, as a field, comprises two fundamental paradigms: supervised learning and unsupervised learning. In supervised learning, the algorithm operates with prior knowledge of the classes or labels during training, enabling it to learn and make predictions accordingly. In contrast, unsupervised learning, exemplified by techniques like hierarchical clustering, endeavors to unveil the inherent structure within datasets, often leading to the discovery of latent classes or groupings [3]. Various ML techniques exist to facilitate the mapping of objects to classes and the creation of predictive models. Notable methodologies include Logistic Regression (LR), Random Forest (RF), the naive Bayes classifier, Support Vector Machine (SVM), and Artificial Neural Networks (ANN), including the growing rapidly domain of Deep Neural Networks (DNN) [4].

Metabolomics

Metabolism, the cornerstone of all biological systems, serves as the vital process responsible for providing essential energy, constructing the requisite building blocks for cellular growth and adaptation, and functioning as a central regulatory hub governing a myriad of biological functions. In recent years, the field of metabolomics has garnered increasing recognition for its unparalleled capacity to offer immediate and comprehensive insights into physiological processes [5]. Metabolites, the end products of metabolism, are widely acknowledged as providing a uniquely representative description of a phenotype, as they closely mirror the dynamic reactions unfolding within a biological system [2]. The field of metabolomics offers a wealth of data that has the potential to bring about transformative advancements in various domains, including clinical, environmental, and biological sciences. This field plays an instrumental role in diagnosing diseases, conducting rigorous toxicological investigations, and monitoring the progress or enhancement of treatments [5]. Due to its ability to decipher the intricate molecular signatures that underlie a multitude of biological states, metabolomics emerges as a potent instrument ready to expose the inner mechanisms of living systems and catalyze innovation across a broad spectrum of scientific research [6].

Machine Learning in metabolomics

In recent years, ML has risen to the forefront of metabolomics research, revolutionizing the way we analyze and interpret metabolomic data. The synergy between metabolomics and ML holds great promise [7]. ML algorithms can discern subtle variations in metabolite profiles associated with specific biological conditions, unravel complex metabolic networks, and aid in biomarker discovery. Moreover, ML-driven metabolomics enables the development of predictive models for disease diagnosis, treatment optimization, and personalized medicine [8]. Many applications of ML in the metabolomics field have predominantly centered on a critical step in the untargeted analysis pipeline: feature selection. In this phase, ML algorithms play a crucial role in distilling vast datasets containing thousands of features down to a more manageable set, typically in the order of tens. These selected features hold predictive value for various health outcomes or phenotypes. It is worth noting that current ML applications in metabolomics are primarily constrained to individual 'omics' datasets. However, recent advancements have extended ML's capabilities to integrate data across different 'omics' levels, adopting a systems biology approach. This progressive development promises to unearth additional and combined biomarkers, enhancing both specificity and ability to unravel the complex web of factors associated with disease initiation and progression [9].

Materials and methods

The statistical processing in this study was conducted using the R software version 4.2.2 and R package known as Biblioshiny (Bibliometrix, K-Synth S.r.l., Naples, Italy). The entire process was organized into two distinct phases: firstly, the collection of data, and secondly, the subsequent bibliometric analysis. Notably, Biblioshiny stands out as a specialized software dedicated to bibliometric analysis. Designed to facilitate user-friendly and interactive exploration of bibliographic data, Biblioshiny offers a comprehensive suite of functions tailored for this purpose. The tool is specifically crafted for the extraction and visualization of bibliometric data derived from scholarly databases like Web of Science™ (Clarivate™, St. Helier, Jersey) [10].

Implemented in the widely utilized R programming language, Biblioshiny provides researchers with a robust platform to filter and extract relevant articles from expansive bibliographic databases based on specified criteria. These criteria may include language, publication period, and keyword inclusion. This functionality allows researchers to streamline their dataset, focusing on articles most pertinent to their research interests [11].

One standout feature of Biblioshiny is its integration of interactive visualization tools. These tools empower users to generate informative graphs, charts, and network analyses. These visualizations play a pivotal role in enabling researchers to gain deeper insights into the bibliographic data [12]. They aid in identifying trends, patterns, and relationships among articles, authors, or keywords. Such insights hold significant value for shaping research strategies, pinpointing knowledge gaps, and situating findings within the broader scientific landscape [13, 14].

In our study, Biblioshiny played a central role in the analysis of extracted data, enabling us to report on key information such as publication volume, participating institutions and countries, prominent keywords, emerging trends, and influential sources. Additionally, a co-occurrence network of the most frequently encountered terms was generated, providing further insights into the interconnectedness of concepts within the analyzed literature.

A systematic literature search (**Fig. 1**) was performed within the Web of Science™ Core Collection, an extensive repository encompassing a vast array of records and documents. The ultimate retrieval strategies employed in this search were: subject

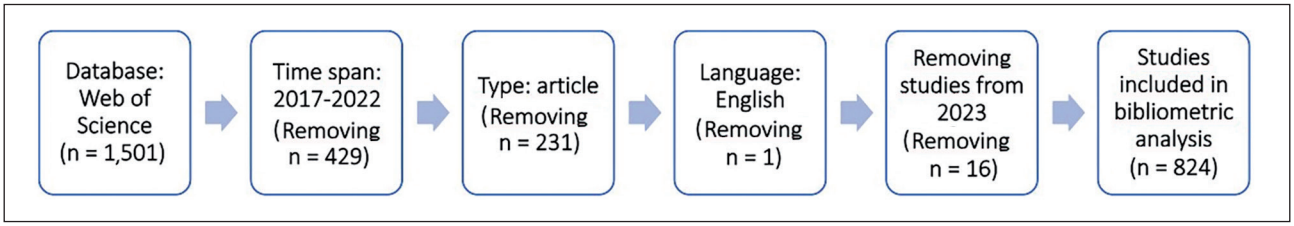

Figure 1. Data collection flow diagram.

words: (metabolomics) AND (Artificial Intelligence OR Machine Learning OR Deep Learning OR Neural Networks); literature type: article; language: English; timespan: 2017-2022. Records and references of 824 studies were downloaded. The results were subsequently imported into bibliometric analysis tools for analysis. The final selection of 88 articles that underwent extensive analysis was primarily driven by the inclusion of experimental studies. Experimental works were prioritized for their ability to provide robust empirical evidence relevant to the study objectives.

Results

The main information of the papers are shown in **Tab. 1**.

Annual analysis

Between 2017 and 2022, a grand total of 824 articles exploring the application of AI in metabolomics were disseminated in the Web of Science™ database, as illustrated in **Fig. 2**.

Notably, the corpus of publications related to the integration of AI within the field of metabolomics has exhibited a remarkable and consistent upward trajectory since 2017.

The annual distribution of these publications is summarized in **Tab. 2**, providing a clear view of the growing interest and engagement in this dynamic intersection of scientific domains.

Table 1. Main information regarding the articles.

Description	Result
Timespan	2017-2022
Sources (journals, books, etc.)	386
Documents	824
Annual growth rate %	48.89
Keywords Plus® (ID) [a]	2,381
Author Keywords (DE) [a]	1,950
Authors	6,232
Authors of single-authored docs	11
Single-authored docs	12
Co-authors per doc	9.08
International co-authorships %	34.95

[a] In the Web of Science™ Core Collection, Keywords Plus® (or "identifiers" [ID]) are keywords generated algorithmically from words or phrases that frequently appear in the titles of an article's references but do not appear in the title of the article itself, while Author Keywords (or "descriptors" [DE]) are the keywords provided by the authors.

Table 2. Number of publications per year.

Year	Articles
2017	41
2018	47
2019	85
2020	128
2021	223
2022	300

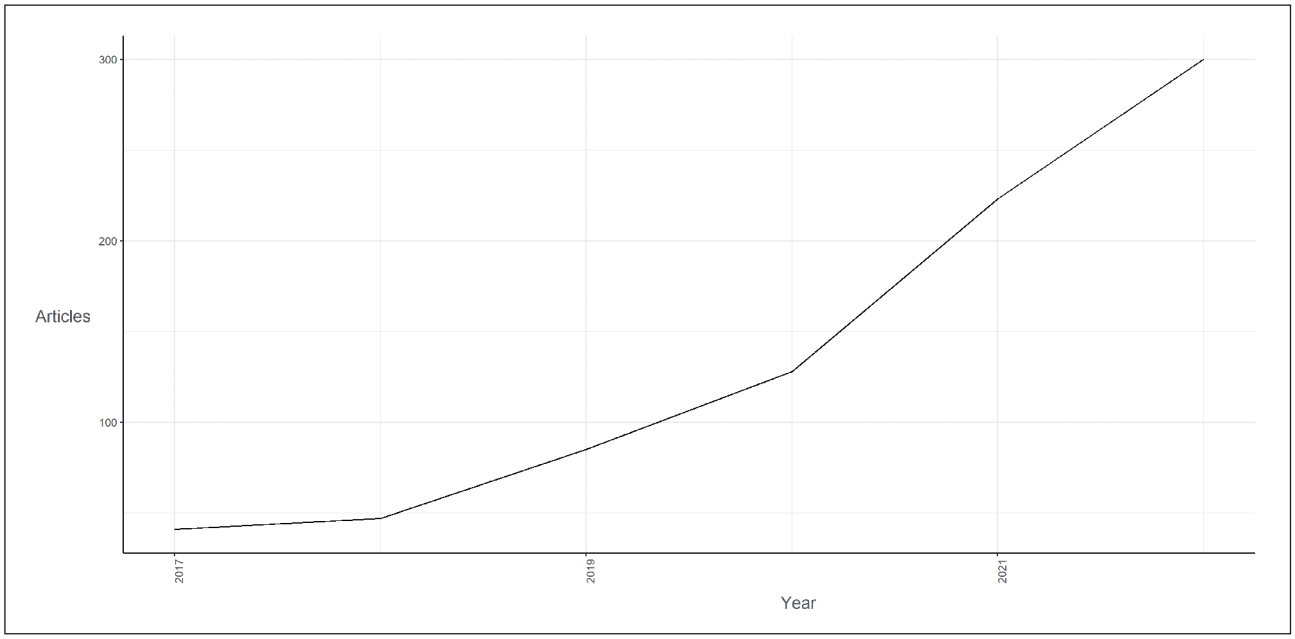

Figure 2. Number of publications per year.

Countries/regions analysis

In the period spanning from 2017 to 2022, the United States emerged as the leading contributor, publishing the highest volume of articles pertaining to the application of AI in metabolomics. Following the United States, other prominent contributors included respectively China, Germany, the United Kingdom, Japan, Canada, Italy, France, Spain, and the Netherlands. The contributions of these countries are visually depicted in **Fig. 3**, generated through the utilization of the Biblioshiny package in the R programming environment, offering a comprehensive representation of their respective research output in this specialized domain.

As depicted in **Fig. 4**, the publication output in the United States, China, Germany, the United Kingdom and Japan experienced a substantial increase from 2017 to 2022, reflecting a consistent and positive trend across all nations. This demonstrate the growing significance of utilizing AI methods in the field of metabolomics research worldwide.

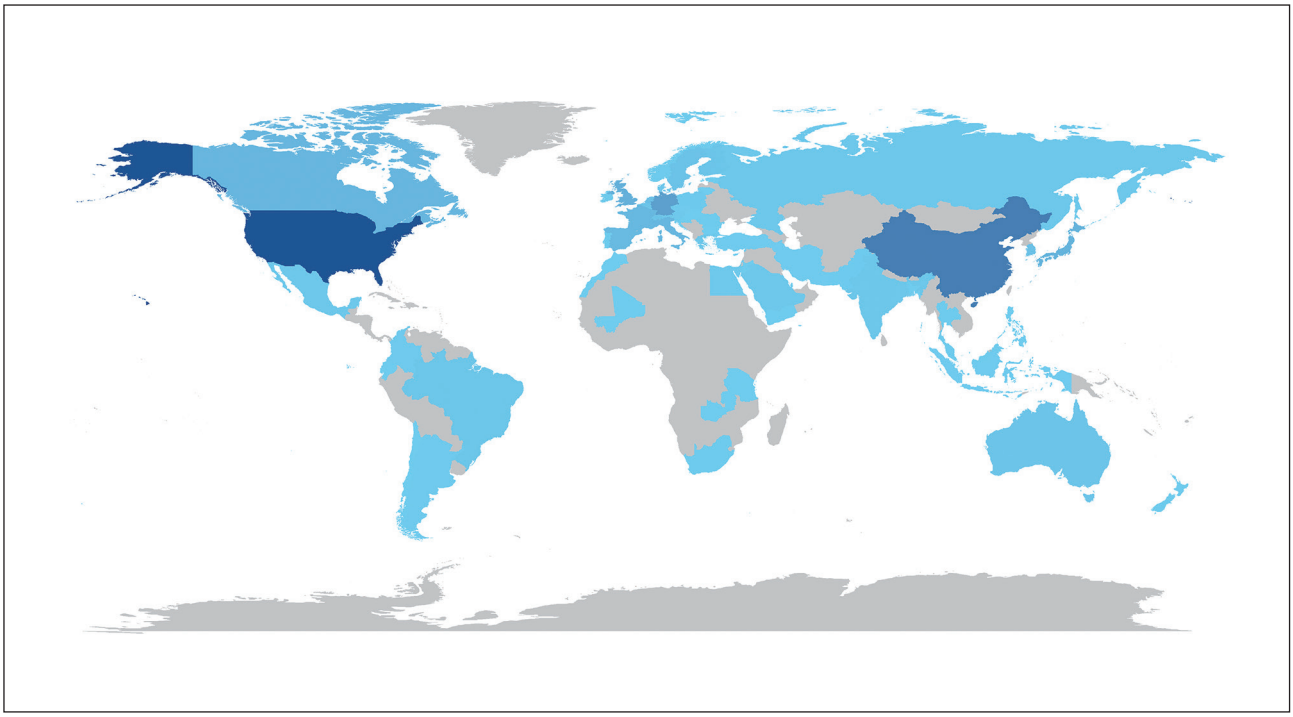

Figure 3. Countries' scientific production.
Darker blue represents greater productivity. Grey indicates countries not producing articles related to the application of Artificial Intelligence (AI) in metabolomics.

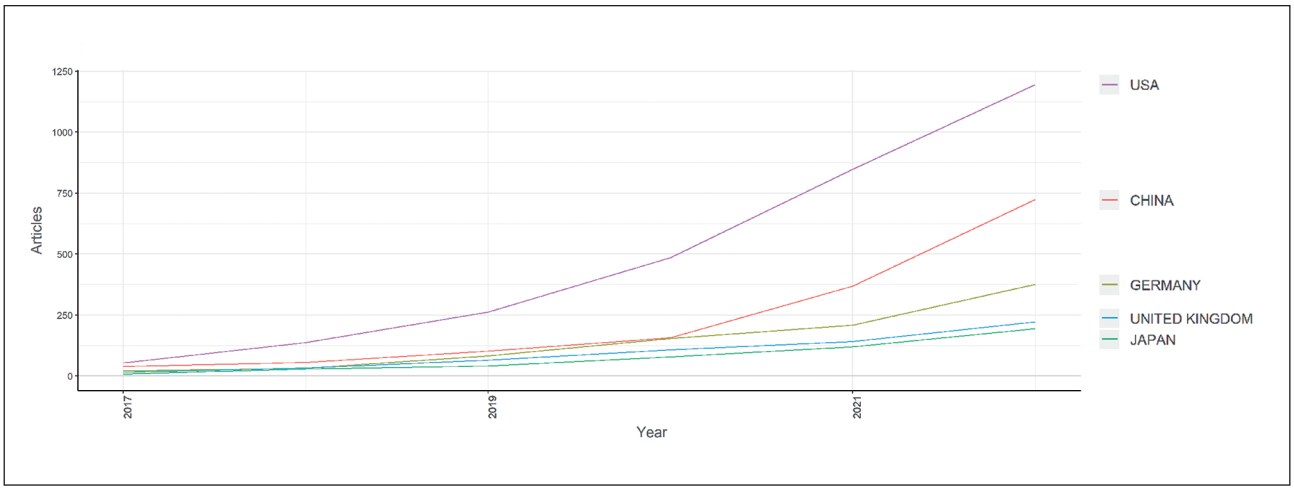

Figure 4. The top 5 countries' scientific production over time.

Institutions analysis

Over the span of 2017 to 2022, a total of 1,654 institutions have actively contributed to research in the realm of AI applied to metabolomics. Remarkably, the top 10 institutions collectively authored 521 articles, representing a substantial 63.2% share of the total publication output. Notably, the University of California System emerged as the foremost contributor to this field, with 94 publications attributed to their research efforts. Following closely, the University of California San Diego accounted for 75 publications, while Harvard University and the Udice-French Research Universities made significant contributions with 73 and 66 publications, respectively, as detailed in **Tab. 3** and **Fig. 5**.

As evident in **Fig. 6**, beginning in 2018, all institutions within the top 5 have undergone a notable and continuous increase in their research output, indicative of a persistent and favorable trend.

Table 3. The top 10 contributing institutions.

Affiliation	Articles
University of California System	94
University of California San Diego	75
Harvard University	73
Udice-French Research Universities	66
Helmholtz Association	41
University of Michigan	37
Chinese Academy of Sciences	36
Institut National de la Santé et de la Recherche Médicale (Inserm)	36
Harvard Medical School	32
Université Paris Cité	31

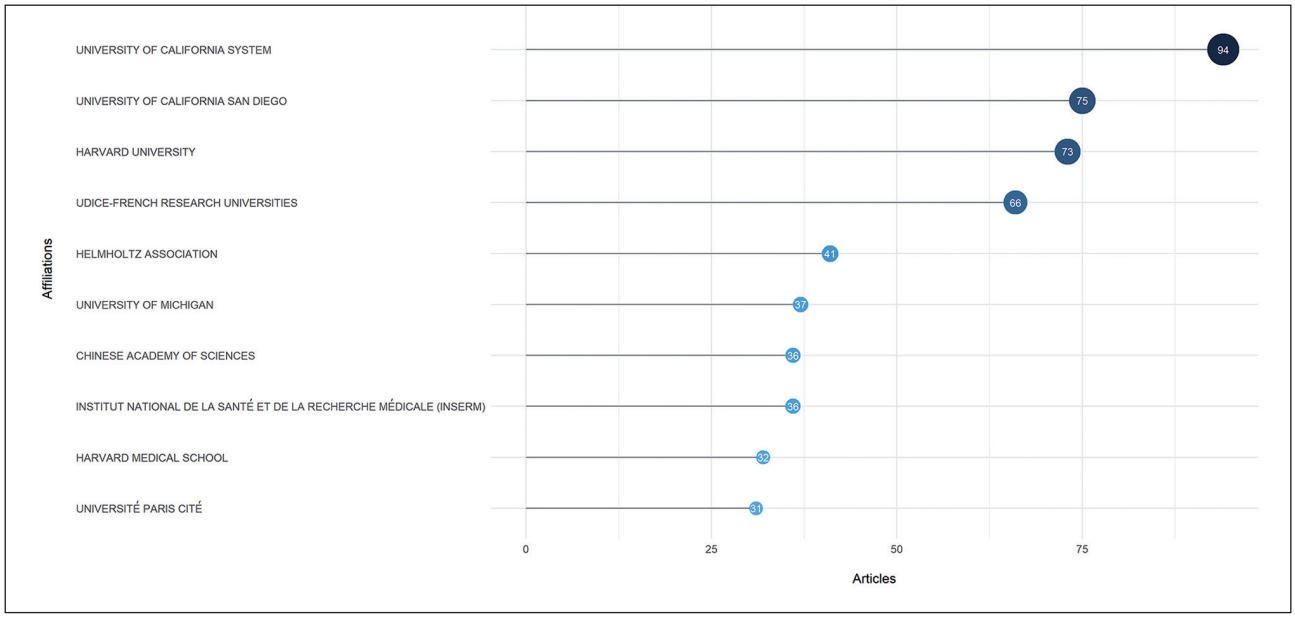

Figure 5. The top 10 contributing institutions.

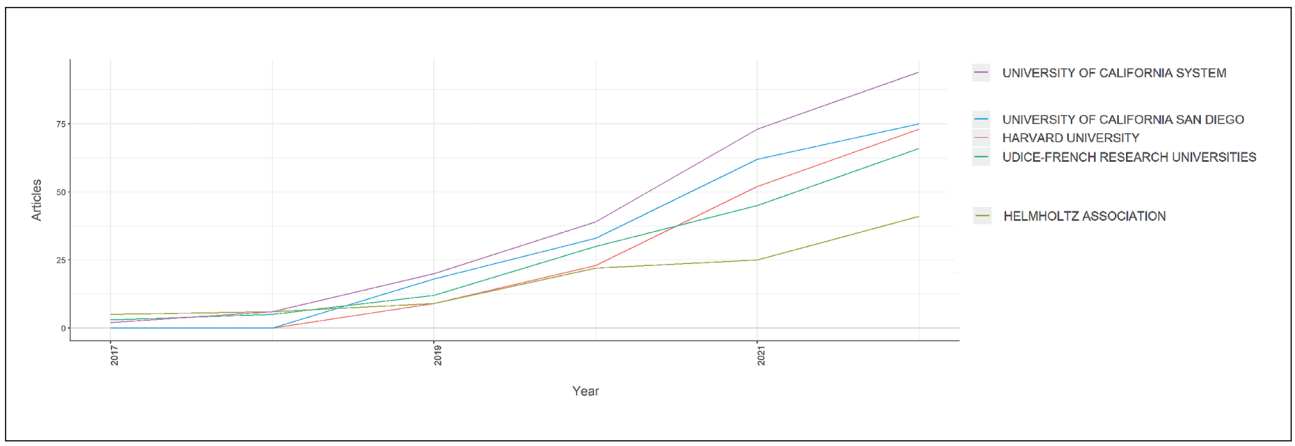

Figure 6. The top 5 contributing institutions' scientific production over time.

Journals analysis

Since 2017, a total of 386 journals have contributed to the dissemination of articles exploring the application of AI in metabolomics.

Our analysis has identified the top 10 contributing journals, as detailed in **Tab. 4** and **Fig. 7**. Collectively, these select journals have published 220 articles, constituting 26.7% of the entire body of publications within the domain of AI applied to metabolomics. Consequently, delving into the articles featured in these prominent journals provides a comprehensive overview of the prevailing research frontiers in this evolving field.

Notably, journals such as *Metabolites*, *Analytical Chemistry*, *Scientific Reports*, and *Metabolomics* not only boast a significant volume of published papers but also carry substantial impact factors within this specific realm of research (**Tab. 5**, **Fig. 8**). These journals emerge as pivotal sources of knowledge in the realm of AI applied to metabolomics, underscoring their importance in shaping the discourse and advancing the field.

The analysis of the source co-citation network (**Fig. 9**) revealed the presence of 4 distinct clusters.

Table 4. The top 10 contributing journals.

Journals	Articles
Metabolites	56
Analytical Chemistry	42
Scientific Reports	34
Metabolomics	26
BMC Bioinformatics	12
Journal of Proteome Research	12
PLoS One	11
International Journal of Molecular Sciences	10
Cancers	9
Analytica Chimica Acta	8

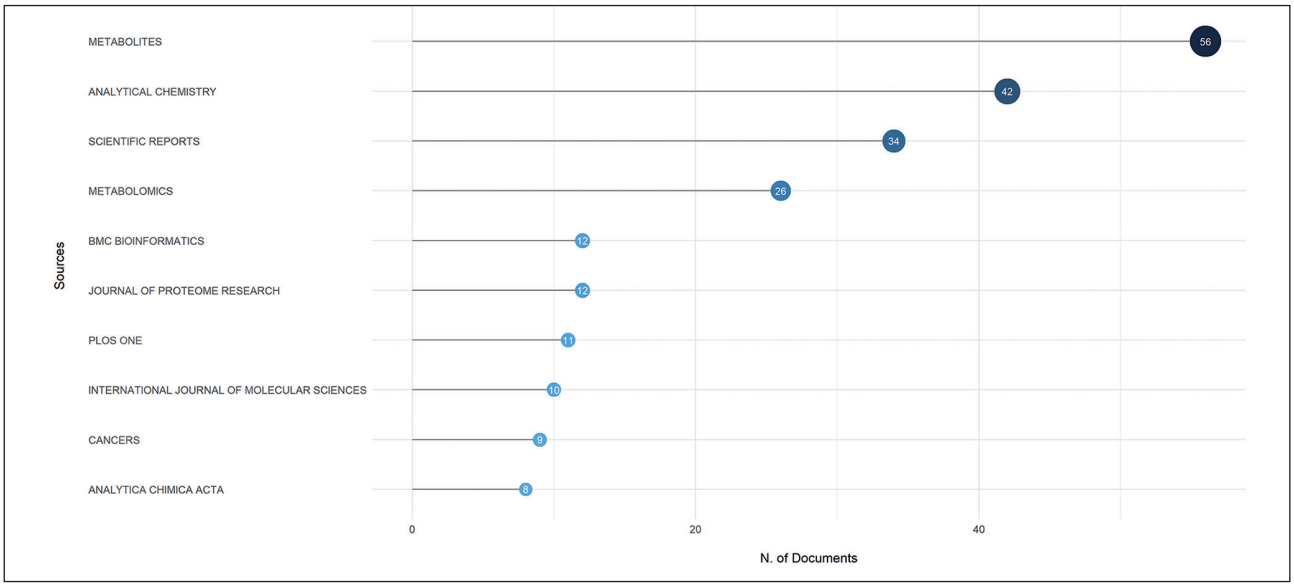

Figure 7. The top 10 contributing journals.

Table 5. The top 10 H-index journals.

Journals	H-index	G-index	M-index	TC	NP	PY_start
Analytical Chemistry	19	35	2.71428571428571	1,285	42	2017
Metabolomics	13	18	1.85714285714286	371	26	2017
Scientific Reports	13	21	1.85714285714286	498	34	2017
Metabolites	11	15	1.57142857142857	324	56	2017
Journal of Proteome Research	9	12	1.28571428571429	303	12	2017
Analytica Chimica Acta	7	8	1.16666666666667	189	8	2018
Nature Communications	7	8	1.75	200	8	2020
Food Chemistry	6	7	0.857142857142857	199	7	2017
Proceedings of the National Academy of Sciences of the United States of America	6	7	1.2	196	7	2019
BMC Bioinformatics	5	9	0.714285714285714	99	12	2017

H-index: it is an author-level metric that measures both the productivity and citation impact of the publications; G-index: it is a variant of the H-index that, in its calculation, gives credit for the most highly cited papers in a data set; M-index: it is another variant of the H-index that displays H-index per year since first publication; TC: total citations; NP: number of publications; PY_start: start of publishing (year).

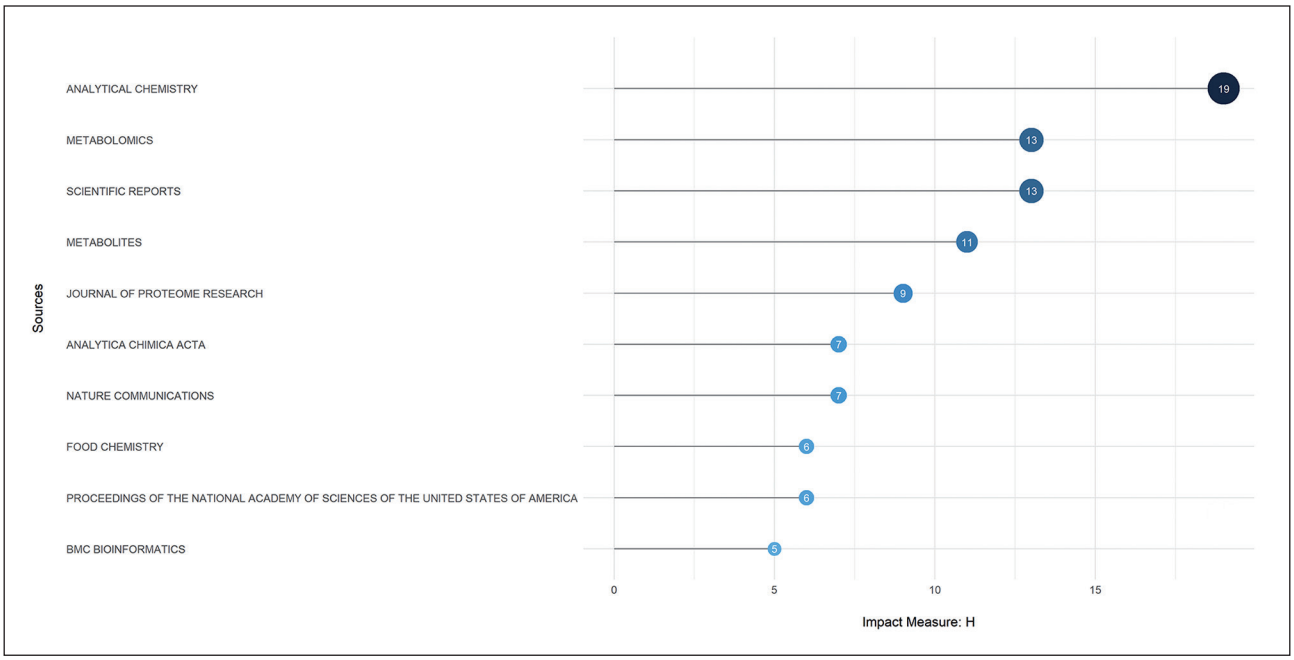

Figure 8. The top 10 H-index journals.

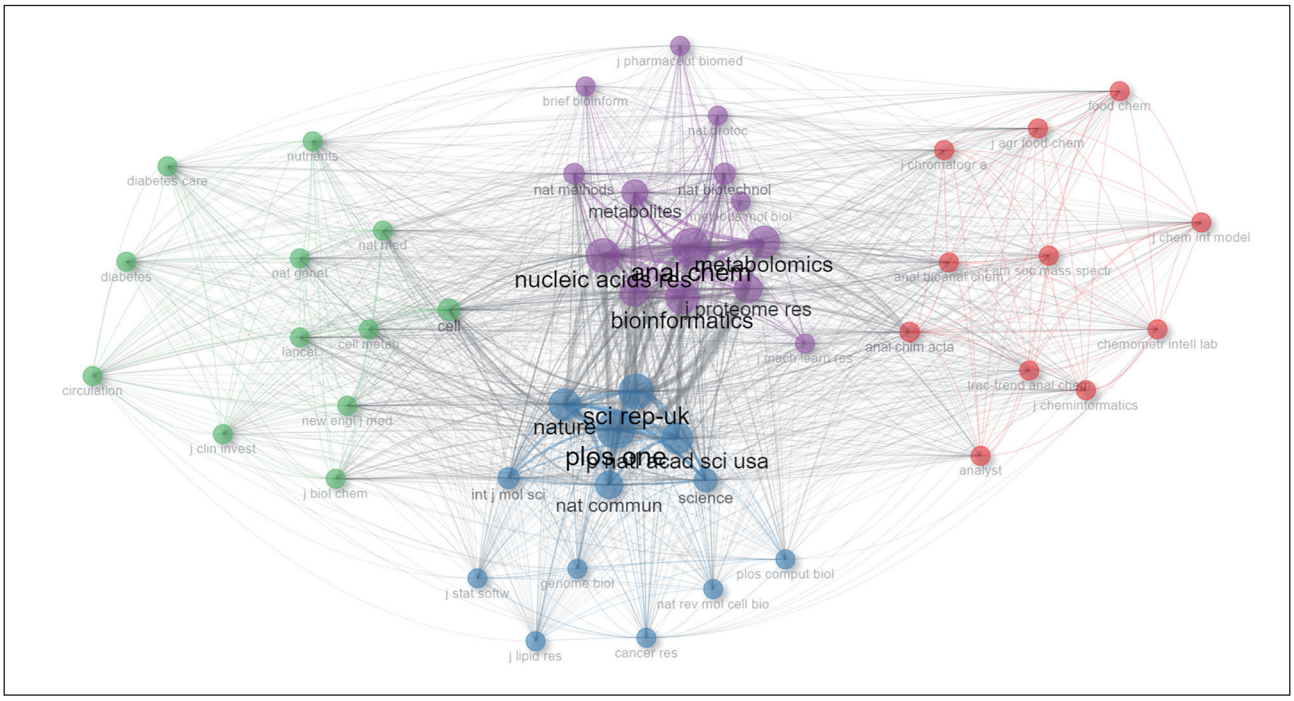

Figure 9. Co-citation network of sources.
Green cluster: cluster 1; purple cluster: cluster 2; red cluster: cluster 3; blue cluster: cluster 4.

Generally, the journals positioned at the outermost edges of each cluster exhibit weaker relationships with journals from other clusters. The size of the nodes within the network denotes the strength of these connections. Notably, clusters 2 and 4 emerge as pivotal components within the network, displaying robust connections not only with each other but also serving as connectors between the remaining clusters (1 and 3). This interconnectedness between clusters 2 and 4 is further underscored by their advantageous positions, both in terms of betweenness centrality and the number of publications they represent. Remarkably, the analysis highlights that more than half of the top 10 productive journals are situated within clusters 2 and 4, reinforcing their significance within the co-citation network and their central role in shaping the scholarly discourse in this domain.

Authors analysis

Author-level analysis shows several key findings. Firstly, the most prominent authors are identified alongside essential bibliometric indicators, as outlined in **Tab. 6** and **Tab. 7**. Notably, Kikuchi J stands out as the author with the highest article count and concurrently boasts the highest H-index and G-index, indicative of substantial scholarly impact. Secondly, the authors' productivity trends over time will be investigated, shedding light on the temporal distribution and output characteristics of their work. As illustrated in **Fig. 10**, a group of authors, including Kikuchi J, Date Y, Troisi J, and Wang X, has demonstrated consistent and prolific productivity, maintaining an annual publication record over the past few years and garnering sustained high citation counts (with a temporary interruption in 2019). Conversely, authors like Wang J, Wang Y, Adamski J, and Graham SF entered the field of applying AI in metabolomics more recently, starting in 2019. Despite their relatively shorter tenure, they have exhibited impressive research output and garnered substantial total citations over the last 4 years. Additionally, we note that authors Chen ZJ and Li Y made their entry into the realm of applying AI in metabolomics in 2020. Despite their relatively brief involvement, Chen ZJ notably secured the second position among the most influential authors in this field, owing to a substantial number of publications and a remarkable annual citation count. These findings collectively underscore the dynamic and evolving landscape of author contributions in the intersection of AI and metabolomics. By analyzing the collaboration network, a total of 11 clusters were generated and 5 of them were interconnected with each other (**Fig. 11**).

Table 6. The top 10 contributing authors.

Authors	Articles	Articles fractionalized [a]
Kikuchi J	10	2.57738095238095
Chen ZJ	9	1.20324675324675
Date Y	9	2.32738095238095
Li Y	9	0.780696058327637
Troisi J	9	0.792735042735043
Wang J	9	0.571570972886762
Wang X	9	0.863475310379335
Wang Y	9	0.771785567373803
Adamski J	8	0.495960170098101
Graham SF	8	0.792893217893218

[a] In Biblioshiny, "articles fractionalized" indicate an individual author's contribution to a published set of papers (uniform contribution of all co-authors at each document is hypothesized).

Table 7. The top 10 H-index authors.

Authors	H-index	G-index	M-index	TC	NP	PY_start
Kikuchi J	7	10	1	178	10	2017
Scala G	7	8	1	142	8	2017
Troisi J	7	9	1	164	9	2017
Adamski J	6	8	1.2	114	8	2019
Date Y	6	9	0.857142857142857	157	9	2017
Graham SF	6	8	1.2	129	8	2019
Guida M	6	6	0.857142857142857	132	6	2017
Yilmaz A	6	8	1.2	129	8	2019
Bahado-Singh RO	5	5	1	97	5	2019
Chen ZJ	5	7	1.25	57	9	2020

H-index: it is an author-level metric that measures both the productivity and citation impact of the publications; G-index: it is a variant of the H-index that, in its calculation, gives credit for the most highly cited papers in a data set; M-index: it is another variant of the H-index that displays H-index per year since first publication; TC: total citations; NP: number of publications; PY_start: start of publishing (year).

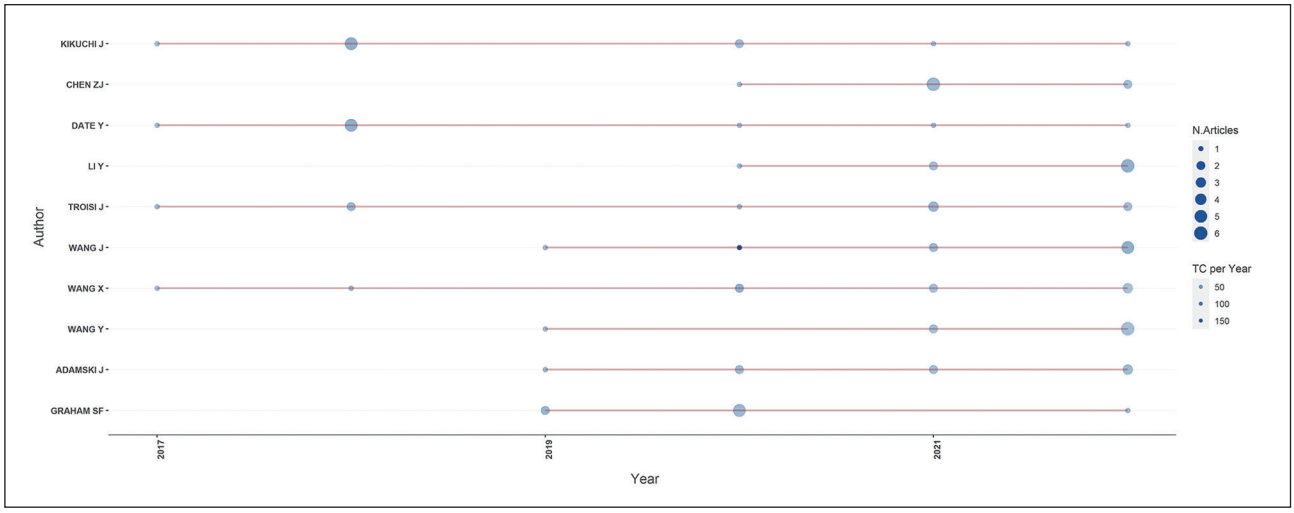

Figure 10. The top 10 contributing authors' production over time.
TC: total citations.

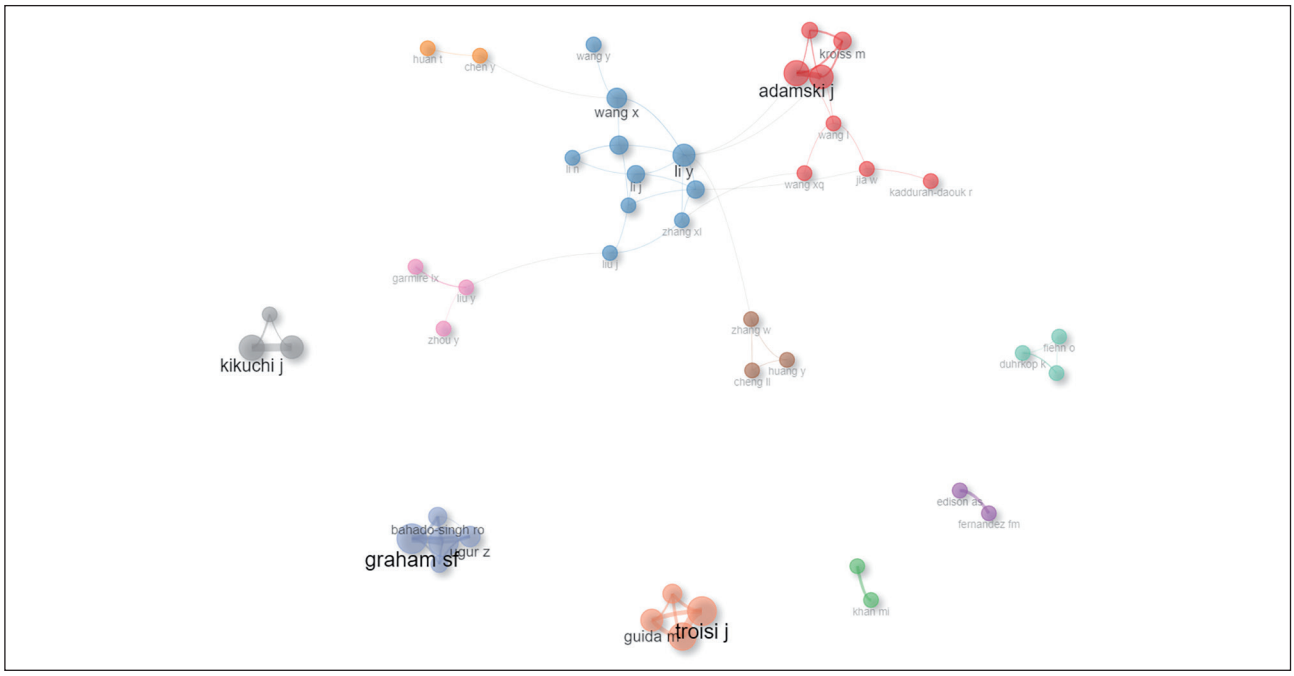

Figure 11. Author collaboration network.

Keywords analysis

In the course of investigation, the capabilities of Biblioshiny to delineate the foremost 10 keywords associated with the application of AI in metabolomics spanning the years 2017 to 2022 were utilized. The most frequent words are metabolomics, identification, biomarkers, mass-spectrometry, risk, diagnosis and so on (**Tab. 8**, **Fig. 12**). The prevalence of specific keywords often serves as a barometer for emerging trends or prevalent themes within a particular field of study.

The co-occurrence network visualisation was made to show the correlation between the keywords.

Table 8. The top 10 keywords.

Words	Occurrences
Metabolomics	143
Identification	83
Biomarkers	69
Mass-spectrometry	65
Risk	59
Diagnosis	48
Expression	44
Cancer	43
Metabolism	41
Prediction	40

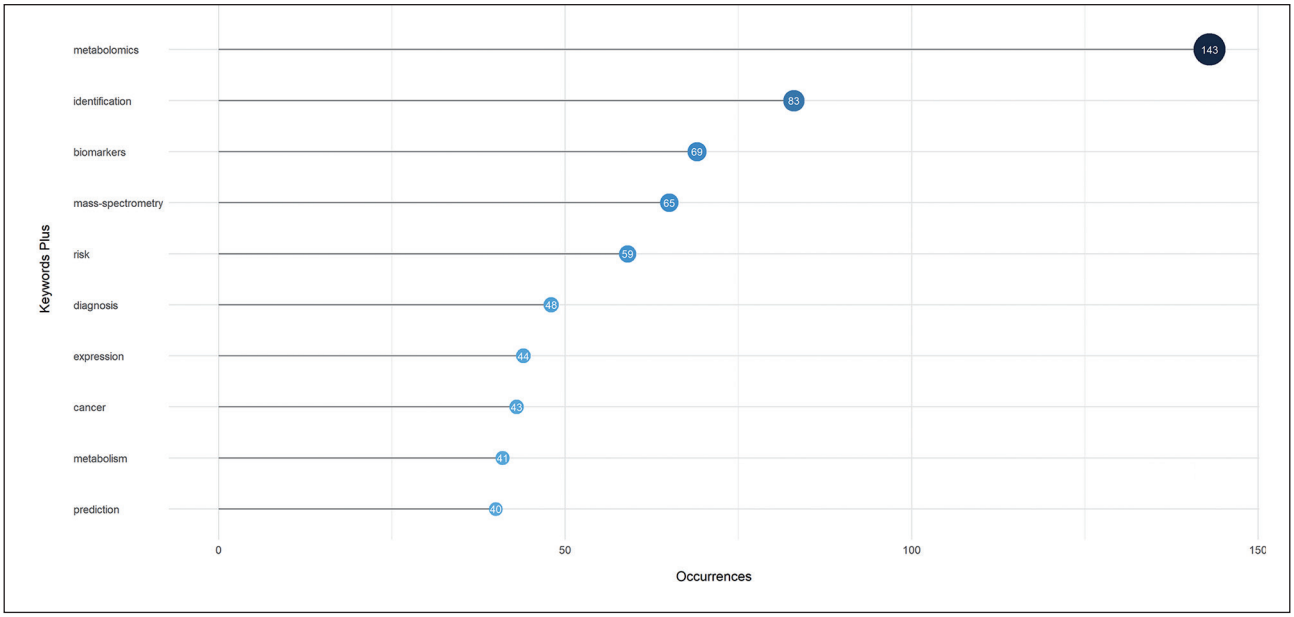

Figure 12. The top 10 keywords.

In the co-word analysis (**Fig. 13**), various aspects are depicted through visual indications: the size of the items corresponds to the frequency of the term, different colors distinguish different clusters, and the length of connections between items signifies the strength of their relationships. This analysis reveals the presence of 2 distinct clusters, each indicative of specific areas of research focus. Cluster 1 primarily centers around scholarly investigations related to biomarkers, indicating a vibrant research domain with a substantial volume of publications from 2017 to 2022. The prominence of this cluster underscores the significance of biomarker research within the academic community. In contrast, cluster 2 appears more comprehensive, as it contains centrally positioned keywords, such as "metabolomics" and "identification." This cluster suggests a multifaceted research landscape, signifying the interconnectedness of topics and a broader scope of inquiry. Notably, this co-word analysis, derived from keywords, provides valuable insights into the prevailing research interests and emphases within academia, elucidating the diverse areas of scholarly exploration and collaboration.

An examination of research trends, as illustrated in **Fig. 14**, reveals a notable evolution in the focus of studies over the years, particularly since 2018. In

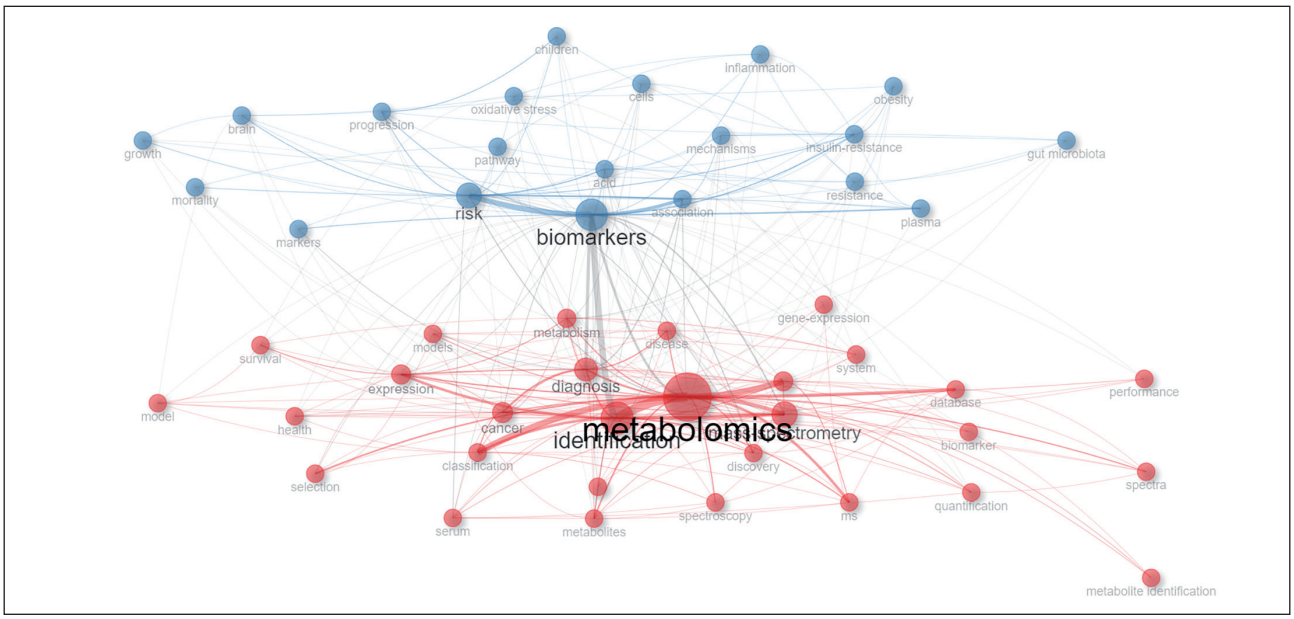

Figure 13. Keywords co-occurrences network.

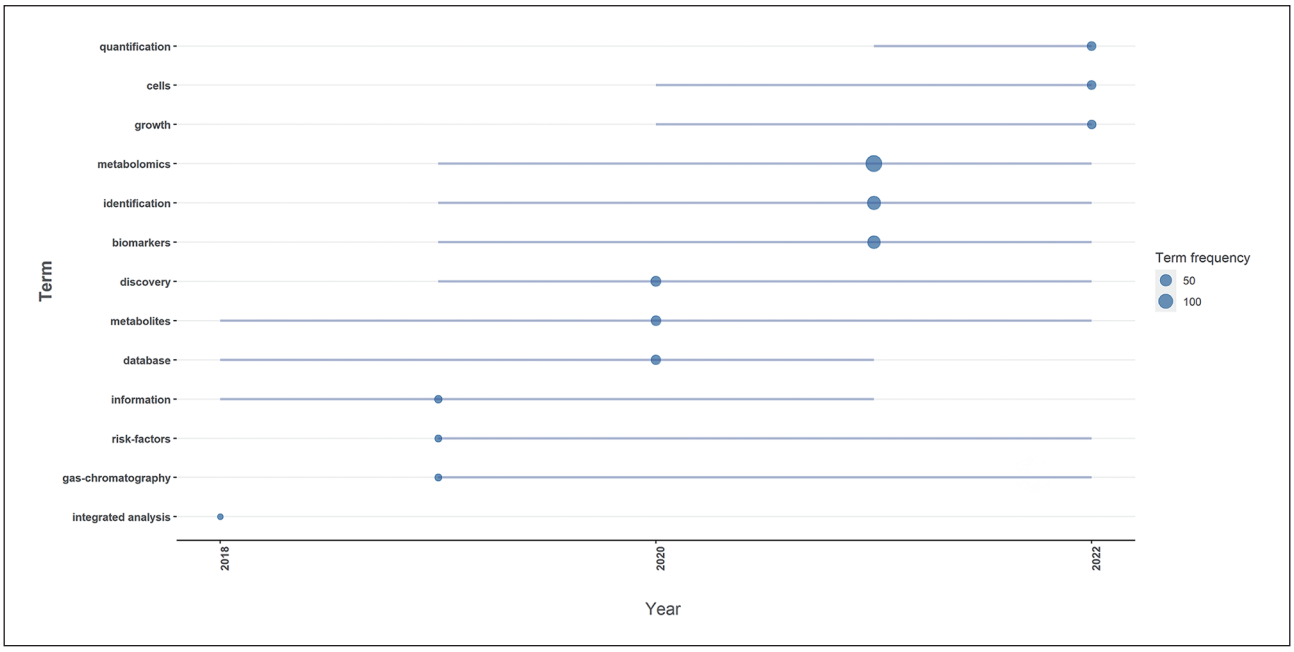

Figure 14. Trend topics over time.

a comparative context, recent years have witnessed a shift towards hot topics such as quantification, cells, growth, metabolomics, identification, and biomarkers, indicating the dynamic nature of research interests in this field.

Extensive analysis of 88 articles

In this study, an extensive analysis of 88 articles sourced from the Web of Science™ database, focusing on the application of AI techniques in the field of metabolomics, was conducted.

After reviewing the initial pool of 824 articles, the selection process for the 88 articles was based on the inclusion of experimental studies. A preference was given to experimental works for their rigorous methodology and direct investigation of the research questions at hand.

The findings of this comprehensive review are summarized in **Tab. 9**.

Notably, the primary objective of AI utilization in these studies was the identification of biomarkers, representing the most prevalent outcome across the analyzed literature.

Furthermore, it is noteworthy that AI has been deployed across a diverse spectrum of diseases for the analysis of metabolomic profiles. Among the various cancers studied, lung and breast cancer emerged as the most frequently investigated types (**Fig. 15**). This highlights the particular relevance of AI-driven metabolomics research in the context of malignancies.

Beyond cancer, other prevalent diseases subject to metabolomic profiling included neurodegenerative diseases such as Alzheimer's and Parkinson's disease, and the acute respiratory disease COVID-19 (**Fig. 15**). These conditions garnered significant attention within the research community, reflecting the pressing need for effective disease characterization and biomarker discovery in these domains.

In terms of the ML techniques employed, the analysis revealed that RF, SVM and LR were the most frequently utilized methods (**Fig. 16**). These algorithms were favored for their ability to effectively handle and interpret the complex data inherent to

Table 9. Extensive analysis of 88 articles (continues on the next page).

Study	Pathology	Model	Main outcome
Oktay et al., 2020 [15]	Breast cancer	RF, LR	Evaluation of blood biomarkers
Dong et al., 2020 [16]	Obesity/food addiction	RF	Brain-gut-microbiome profile
Chen et al., 2021 [17]	Tumor brain metastases	ANN	Identification of biomarkers
Miller et al., 2022 [18]	Low-level or sub-concussive blast overpressure	RF, t-SNE	Identification of biomarkers

Table 9. Extensive analysis of 88 articles (continues from the previous page and on the next page).

Study	Pathology	Model	Main outcome
Massetti et al., 2022 [19]	Alzheimer's disease spectrum	RF	Prediction of the clinical course
Muller et al., 2021 [20]	Inflammatory bowel disease, gastric cancer, irritable bowel syndrome, colorectal adenomas or cancer	RF, REM	Prediction of metabolite levels on microbiome composition
Stamate et al., 2019 [21]	Alzheimer's disease	DL, XGBoost, RF	Prediction of the disease
Pannkuk et al., 2018 [22]	Hematopoietic acute radiation syndrome	RF	Identification of biomarkers
Troisi et al., 2017 [23]	Fetal aneuploidies	PLS-DA, LDA, NB, DT, RF, k-NN, ANN, SVM	Evaluation of the diagnostic performance
Troisi et al., 2018 [24]	Congenital anomalies of the central nervous system	PLS-DA, LDA, NB, DT, RF, k-NN, ANN, SVM, LR	Building a metabolomic fingerprint
Troisi et al., 2022 [25]	Colorectal cancer	NB, GLM, LR, DL, DT, RF, GBT, SVM, PLS-DA	Evaluation of the metabolic alterations associated with the disease
Liang et al., 2022 [26]	Lung cancer	PLS-DA, OPLS-DA	Identification of biomarkers
Chung and Kang, 2019 [27]	Breast cancer, ovarian cancer	Neural network-based method ATHENA	Disease classification
Wang et al., 2018 [28]	Sepsis	KELM, FOA, RF, PSO-based KELM, GA-based KELM, FOA-based KELM, ANN, SVM	Prediction of the disease, identification of the most important biomarkers
Neumann et al., 2021 [29]	Urachal cancer, urachal adenocarcinomas	k-NN, SVM, RF	Detection of biomarkers
Gupta et al., 2022 [30]	Endometrial, breast, cervical, ovarian cancer	K-NN, LR	Detection of disease
Wang et al., 2022 [31]	Type 2 diabetes	RF, XGBoost, OPLS-DA, SVM, DNN	Prediction of clinical outcomes
Zacharias et al., 2019 [32]	Chronic kidney disease	PH, LASSO Cox PH	Prediction of the requirement of dialysis or renal transplantation
Blasco et al., 2018 [33]	Amyotrophic lateral sclerosis	OPLS-DA, RF, SVM	Detection of biomarkers
Bocca et al., 2018 [34]	Dominant optic atrophy	OPLS-DA, Biosigner, RF, SVM	Detection of biomarkers
Olin et al., 2020 [35]	Fibromuscular dysplasia	k-NN, RF	Identification of biomarkers
Bao et al., 2022 [36]	COVID-19	RF, CoxBoost	Prediction, identification of biomarkers
Troisi et al., 2021 [37]	Bladder cancer	PLS-DA, NB, DT, RF, k- NN, ANN, SVM, LR, DL	Creation of a metabolomics-based profile of the disease
Rawshani et al., 2020 [38]	Type 2 diabetes	RF, Gradient boosting	Identification of predictors of cardiometabolic risk profle
Anwar et al., 2018 [39]	Autism spectrum disorder	RF, LR, Ensemble classifier, SVM	Biomarker selection
Tang et al., 2019 [40]	Lymphangioleiomyomatosis	EBAM, SAM	Identification of biomarkers
Chang et al., 2021 [41]	Amyotrophic lateral sclerosis	OPLS-DA, SVM	Identification of biomarkers
Sapkota et al., 2018 [42]	Alzheimer's disease	OPLS-DA, SVM, RF	Identification of biomarkers
Vicente-Dueñas et al., 2020 [43]	Leukemia	RF	Identification of biomarkers
Siegel et al., 2017 [44]	Type 1 diabetes	Bootstrap, LDA	Detection of biomarkers

Table 9. Extensive analysis of 88 articles (continues from the previous page and on the next page).

Study	Pathology	Model	Main outcome
Huang et al., 2021 [45]	Inflammatory bowel disease	SVM, AdaBoost, RF, DNN	Multi-classification of inflammatory bowel disease and its subtypes
Irajizad et al., 2022 [46]	Triple-negative breast cancer	DL, RF, Ensemble learning, Gradient boosting	Identification of biomarkers
Bahado-Singh et al., 2022 [47]	Alzheimer's disease	k-NN, SVM, GLM, PAM, RF, LDA, DL	Identification of biomarkers
Proitsi et al., 2017 [48]	Alzheimer's disease	RF	Identification of biomarkers
Borkowski et al., 2021 [49]	Alzheimer's disease	PLS-DA, LR	Identification of biomarkers
Kehoe et al., 2022 [50]	Lyme disease	SSVM, k-NN	Selection of metabolic biomarkers, building a metabolite-based diagnostic
Varma et al., 2018 [51]	Alzheimer's disease	SVM, RF	Identification of biomarkers
Khan et al., 2022 [52]	Insulin resistance	RF, OPLS	Identification of biomarkers
Miller-Atkins et al., 2020 [53]	Hepatocellular carcinoma	LR, RF	Predictive model
Gbaoui et al., 2022 [54]	Major depressive disorder	RF, LR	Identification of biomarkers
Kaur et al., 2022 [55]	Breast cancer	DNN, GBM, DRF	Prediction of the occurrence, reoccurrence, and survival
Zhou et al., 2022 [56]	Postherpetic neuralgia	PLS-DA, OPLS-DA, RF, SMV, LR	Identification of biomarkers
Tiedt et al., 2020 [57]	Ischemic stroke, stroke mimics	RF, LDA, LR, k-NN, NB, SVM	Discrimination of patients with ischemic stroke from stroke mimics
Zhou et al., 2021 [58]	Rheumatoid arthritis	PLS-DA, OPLS-DA, RF, BLR, CP-ANN	Identification of biomarkers
Li et al., 2021 [59]	Malignant mesothelioma	RF, OPLS-DA	Identification of biomarkers
Lai et al., 2022 [60]	Lung cancer	PLS-DA, RF, XGBoost, LightGBM, k-NN, SVM, LR, ExtraTree	Identification of biomarkers
Hu et al., 2022 [61]	Tuberculosis	OPLS-DA, Metabolite enriched pathways, RF, SVM, MLP	Identification of biomarkers
Yang et al., 2022 [62]	Retinopathy of prematurity	RF, OPLS-DA	Identification of biomarkers
Hao et al., 2018 [63]	Alzheimer's disease	SVM	Identification of biomarkers
Koureas et al., 2021 [64]	Lung cancer	RF	Identification of biomarkers
Gal et al., 2020 [65]	Breast cancer	k-means, SIMLR, k-sparse and Spectral clustering	Classification of the disease
Xiao et al., 2022 [66]	Triple-negative breast cancer	LASSO, SVM	Identification of biomarkers
Delafiori et al., 2021 [67]	COVID-19	Tree boosting (ADA), RF, XRF, PLS, SVM	Diagnosis
Villagrana-Bañuelos et al., 2022 [68]	COVID-19	GA, RF	Prediction outcomes
Liu et al., 2022 [69]	Acute myocardial infarction	PRA, RF, RFE, GA	Identification of biomarkers
Alakwaa et al., 2018 [70]	Breast cancer	DL, RF, SVM RPART, LDA, PAM, GBM	Identification of biomarkers
Zhang et al., 2022 [71]	Stroke	XGBoost, OPLS-DA	Detection of the disease
Johno et al., 2018 [72]	Atherosclerosis	PLS, LR	Identification of biomarkers
Wang et al., 2019 [73]	Osteoporosis	PCA, PLS-DA	Identification of biomarkers

Table 9. Extensive analysis of 88 articles (continues from the previous page).

Study	Pathology	Model	Main outcome
Miller et al., 2021 [74]	Lung cancer	PCA, PLS-DA, k-NN, Bayesian principal component analysis	Identification of biomarkers
Trezzi et al., 2017 [75]	Parkinson's disease	LR	Identification of biomarkers
Xie et al., 2021 [76]	Lung cancer	K-NN, NB, AdaBoost, SVM, RF, ANN	Identification of biomarkers
Peddinti et al., 2017 [77]	Type 2 diabetes	LR, RLS	Identification of biomarkers
Miller et al., 2021 [78]	Lung cancer	PLS-DA, SVM, ANN, RF	Identification of biomarkers
Akyol et al., 2020 [79]	Dementia	DL, RF, SVM, LDA, PAM, GLM	Identification of biomarkers
Sinha et al., 2017 [80]	Asthma	RF, k-means	Identification of biomarkers
Beccaria et al., 2018 [81]	Tuberculosis	SVM, PLS-DA, RF	Identification of biomarkers
Iwano et al., 2021 [82]	Pancreatic cancer	PLSR, SVM	Identification of biomarkers
Abdullah at al. 2022 [83]	Alzheimer's disease	LR	Identification of biomarkers
Lin et al., 2022 [84]	Carotid artery stenosis	OPLS-DA, DT, RF	Identification of biomarkers
Eng et al., 2021 [85]	Cystic fibrosis	RF	Identification of biomarkers
Webb-Robertson et al., 2022 [86]	Type 1 diabetes	NB	Identification of biomarkers
Glaab et al., 2019 [87]	Parkinson's disease	SSVM, RF	Diagnosis
An et al., 2022 [88]	Breast cancer	LASSO, RF, SVM	Identification of biomarkers
Gilard et al., 2021 [89]	Glioblastoma	RF	Mechanisms of disease
Celaya-Padilla et al., 2021 [90]	COVID-19	SSVM, LR	Identification of biomarkers
Buszewska-Forajta et al., 2021 [91]	Prostate cancer	ANN, RF	Identification of biomarkers
Wang et al., 2022 [92]	Lung cancer	SVM, RF, AdaBoost	Identification of biomarkers
Miller et al., 2022 [93]	Lung cancer	RF, ANN, MLP, NSC, NB, BGLM, k-NN, SVM, SPLS, LR, RF	Identification of biomarkers
Gao et al., 2022 [94]	Alcoholic hepatitis	RF, LR, SVM	Prediction of mortality
Kosyakovsky et al., 2022 [95]	Sepsis	RF, SVM, k-NN, NSC, LASSO, PLS-DA, LR	Prediction of mortality
Yilmaz et al., 2020 [96]	Parkinson's disease	LR, SVM	Metabolic profiling
Liu et al., 2022 [97]	COVID-19	RF, LR	Identification of biomarkers
Rahnavard et al., 2022 [98]	COVID-19	DNN, k-NN, RF, LR	Disease severity prediction
Njoku et al., 2021 [99]	Endometrial cancer	RF	Identification of biomarkers
Cheng et al., 2019 [100]	Endometrial cancer	SVM, PLS-DA, RF, LR	Identification of biomarkers
Qureshi et al., 2022 [101]	Autism spectrum disorder	FDA, SVM	Identification of biomarkers
Dimitri et al., 2022 [102]	Parkinson's disease	SVM, ElasticNet, PLS	Identification of biomarkers

ANN: Artificial Neural Networks; BGLM: Boosted General Linear Model; BLR: Binary Logistic Regression; CP-ANN: Counter Propagation Artificial Neural Network; DL: Deep Learning; DNN: Deep Neural Network; DRF: Distributed Random Forest; DT: Decision Tree; EBAM: Empirical Bayes Analysis of Microarrays; ExtraTree: Extremely Randomized Trees; FDA: Fisher Discriminant Analysis; FLM: Fast Large Margin; FOA: Fruit Fly Optimization Algorithm; GA: Genetic Algorithms; GBM: Gradient Boosting Machine; GBT: Gradient Boosted Trees; GLM: Generalized Linear Model; k-NN: k-Nearest Neighbor; KELM: Kernel Extreme Learning Machine; LASSO: Least Absolute Shrinkage and Selection Operator; LDA: Linear Discriminant Analysis; LightGBM: Light Gradient Boosting Machine; LR: Logistic Regression; MLP: Multilayer Perceptron Neural Network; NB: Naïve Bayes; NSC: Nearest Shrunken Centroids; OPLS-DA: Orthogonal Partial Least Squares Discriminant Analysis; PAM: Prediction Analysis for Microarrays; PCA: Principal Component Analysis; PH: Cox Proportional Hazards; PLS: Partial Least Squares; PLS-DA: Partial Least Square Discriminant Analysis; PLSR: Partial Least Square Regression; PRA: Poisson Regression Analysis; PSO: Particle Swarm Optimization; REM: Random-Effects Models; RF: Random Forest; RFE: Recursive Feature Elimination; RLS: Regularised Least-Squares; RPART: Recursive Partitioning And Regression Trees; SAM: Significance Analysis of Microarrays; SIMLR: Single-cell Interpretation via Multi-kernel LeaRning; SPLS: Sparse Partial Least Squares; SSVM: Sparse Support Vector Machines; SVM: Support Vector Machine; t-SNE: t-distributed Stochastic Nearest Neighbor Embedding; XGBoost: Extreme Gradient Boosting; XRF: Extreme Random Forest.

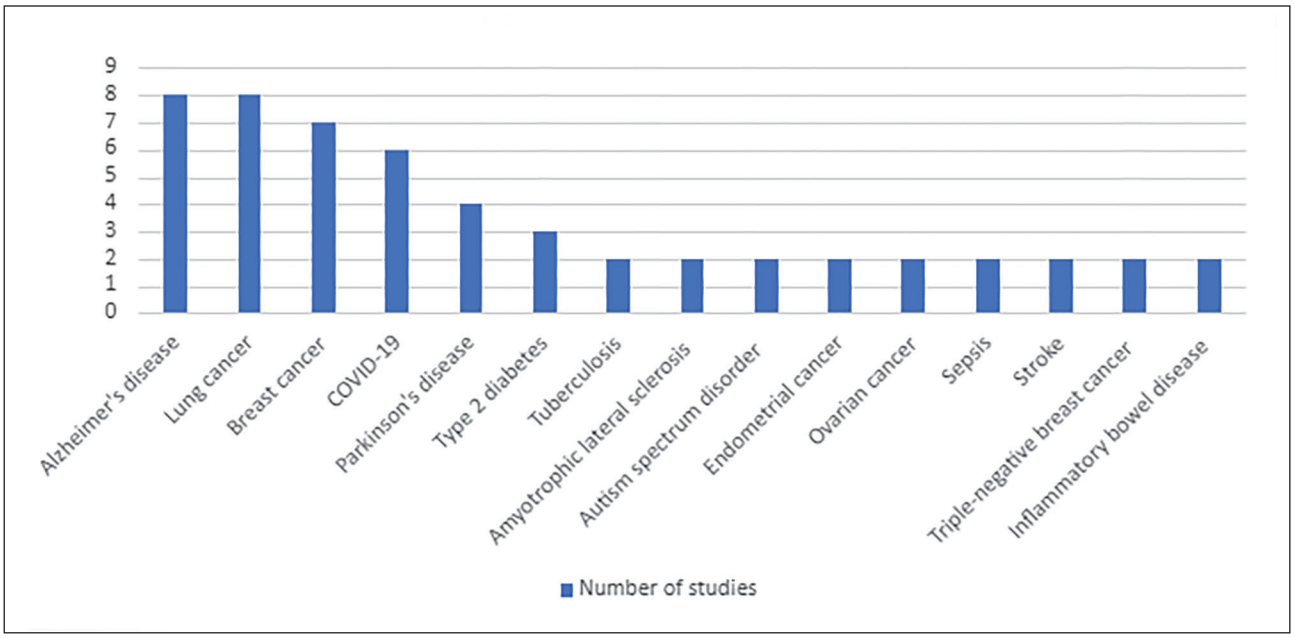

Figure 15. Pathological distribution of studies.

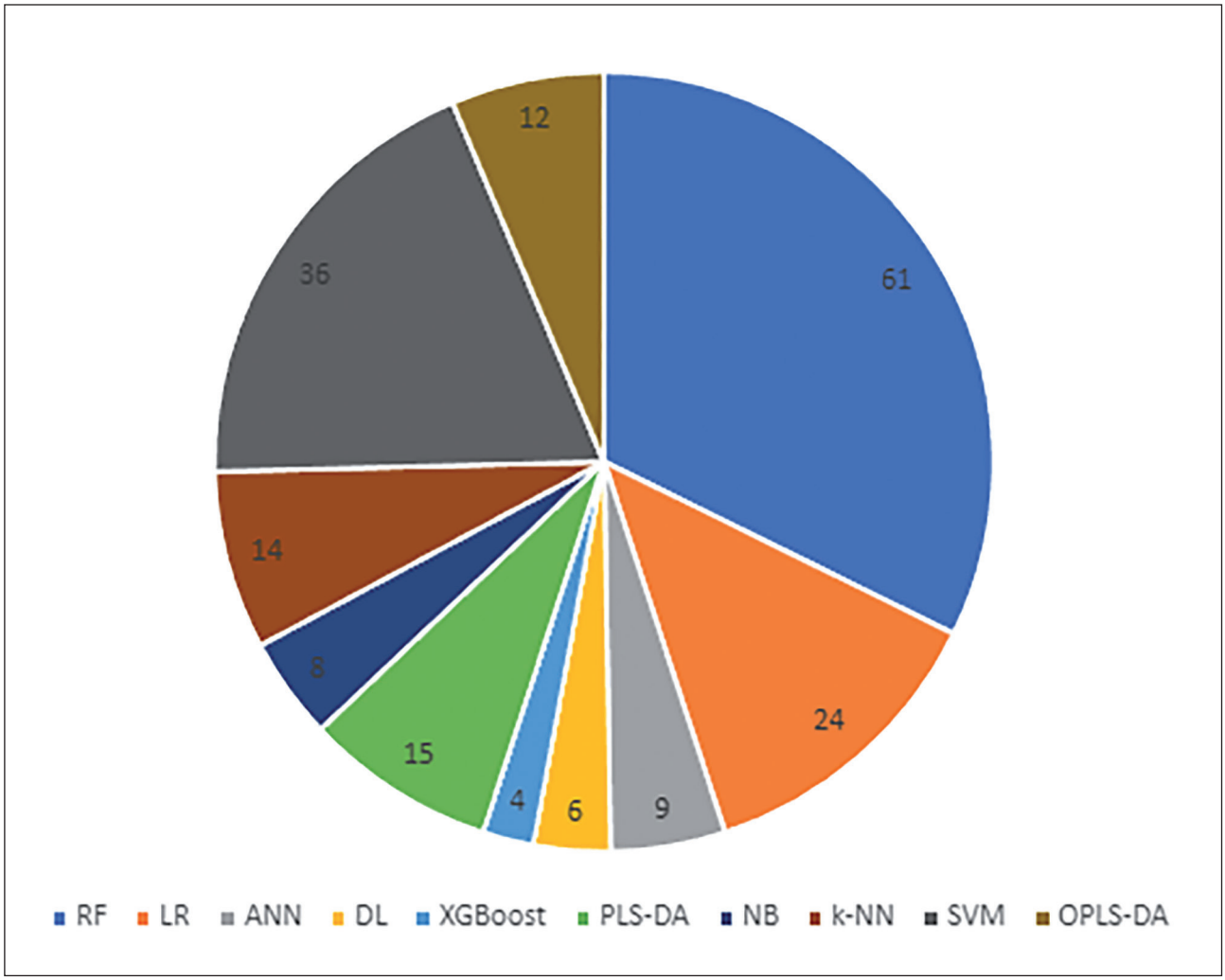

Figure 16. Distribution of Machine Learning (ML) methods.
ANN: Artificial Neural Networks; DL: Deep Learning; k-NN: k-Nearest Neighbor; LR: Logistic Regression; NB: Naïve Bayes; OPLS-DA: Orthogonal Partial Least Squares Discriminant Analysis; PLS-DA: Partial Least Square Discriminant Analysis; RF: Random Forest; SVM: Support Vector Machine; XGBoost: Extreme Gradient Boosting.

metabolomics studies. Such ML techniques have demonstrated their efficacy in achieving accurate biomarker identification and disease profiling, making them prominent choices in the realm of AI-driven metabolomics research.

Conclusions

This bibliometric investigation has explored the global landscape of scientific production concerning the application of AI in metabolomics from 2017 to 2022. The analysis has uncovered a visible upswing in the number of information sources, authors, and scholarly documents, underscoring the interest and engagement in this dynamic domain. Furthermore, the collaborative endeavors among authors have witnessed a notable increase, manifesting in a substantial international collaboration rate of 34.95%.

Prominent journals, including *Metabolites*, *Analytical Chemistry*, *Scientific Reports*, and *Metabolomics*, have emerged as pivotal platforms for disseminating articles related to AI in metabolomics. Noteworthy authors such as Kikuchi J, Chen ZJ, and Date Y have made significant contributions to advancing this field's discourse and knowledge base.

In summary, our comprehensive analysis in the field of metabolomics research reveals several key insights. AI has emerged as a powerful tool for biomarker identification and disease profiling, with the majority of studies focusing on these objectives. ML techniques, including RF, LR, and SVM, have been extensively employed, underscoring their effectiveness in handling the complex data inherent to metabolomics.

Furthermore, our findings highlight specific areas of interest within metabolomics research. Lung and breast cancer stand out as prominent subjects of investigation, emphasizing the relevance of AI-driven metabolomics in oncology. Additionally, Alzheimer's and Parkinson's disease and COVID-19 have garnered substantial attention, reflecting the urgency in characterizing these conditions and identifying potential biomarkers.

The distribution of ML methods in metabolomics showcases the diversity of approaches used to extract meaningful insights from metabolomic data. This diversity underscores the adaptability of AI techniques to the unique challenges posed by different pathologies and research contexts.

Overall, our study underscores the growing significance of AI in metabolomics research and its potential to revolutionize disease diagnosis and biomarker discovery. As the field continues to evolve, researchers should remain attentive to emerging trends and explore innovative applications of AI to further advance our understanding of metabolic pathways and disease mechanisms.

These findings illuminate the international dimensions and the evolving research themes within the realm of AI in metabolomics. Prospective research in this domain could delve into emerging areas such as biomarker discovery, personalized medicine, elucidating disease mechanisms, innovative pharmaceuticals, and novel therapeutic interventions. Such exploration promises to further enrich our understanding and application of AI in the context of metabolomics.

Declaration of interest

The Authors declare no conflict of interest. Funding: this research received no external funding.

References

1. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017;69S:S36-40.
2. Kirk D, Kok E, Tufano M, Tekinerdogan B, Feskens EJM, Camps G. Machine Learning in Nutrition Research. Adv Nutr. 2022;13(6):2573-89.
3. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. Transl Vis Sci Technol. 2020;9(2):14.
4. Park G, Lee YG, Yoon YS, Ahn JY, Lee JW, Jang YP. Machine Learning-Based Species Classification Methods Using DART-TOF-MS Data for Five Coniferous Wood Species. Forests. 2022;13(10):1688.
5. Muthubharathi BC, Gowripriya T, Balamurugan K. Metabolomics: small molecules that matter more. Mol Omics. 2021;17(2):210-29.
6. Cuperlovic-Culf M. Machine Learning Methods for Analysis of Metabolic Data and Metabolic Pathway Modeling. Metabolites. 2018;8(1):4.
7. Haug C, Drazen J. Artificial Intelligence and Machine Learning in Clinical Medicine. N Engl J Med. 2023;388(13):1201-8.
8. Wang XB, Cui NH, Liu X. A novel 6-metabolite signature for prediction of clinical outcomes in type 2 diabetic patients undergoing percutaneous coronary intervention. Cardiovasc Diabetol. 2022;21(1):126.
9. Petrick LM, Shomron N. AI/ML-driven advances in untargeted metabolomics and exposomics for biomedical applications. Cell Rep Phys Sci. 2022;3(7):100978.
10. Liu B, Zhou CJ, Ma HW, Gong B. Mapping the youth soccer: A bibliometrix analysis using R-tool. Digit Health. 2023;9:20552076231183550.
11. Campra M, Riva P, Oricchio G, Brescia V. Bibliometrix analysis of medical tourism. Health Serv Manage Res. 2022;35(3):172-88.

12. Koukopoulos A, Neimeyer RA. Prolonged grief disorder: A bibliometric analysis. Death Stud. 2023:1-14.
13. Guo XJ, Wu P, Jia X, Dong YM, Zhao CM, Chen NN, Zhang ZY, Miao YT, Yun KM, Gao CR, Ren Y. Mapping the structure of depression biomarker research: A bibliometric analysis. Front Psychiatry. 2022;13:943996.
14. Aria M, Cuccurullo C. Bibliometrix: An R-tool for comprehensive science mapping analysis, J Informetrics. 2017;11(4):959-75.
15. Oktay K, Santaliz-Casiano A, Patel M, Marino N, Storniolo AMV, Torun H, Acar B, Madak Erdogan Z. A Computational Statistics Approach to Evaluate Blood Biomarkers for Breast Cancer Risk Stratification. Horm Cancer. 2020;11(1):17-33.
16. Dong TS, Mayer EA, Osadchiy V, Chang C, Katzka W, Lagishetty V, Gonzalez K, Kalani A, Stains J, Jacobs JP, Longo VD, Gupta A. A Distinct Brain-Gut-Microbiome Profile Exists for Females with Obesity and Food Addiction. Obesity (Silver Spring). 2020;28(8):1477-86.
17. Chen W, Li C, Shi Y, Zhang Y, Jin D, Zhang M, Bo M, Li G. A Comprehensive Analysis of Metabolomics and Transcriptomics Reveals Novel Biomarkers and Mechanistic Insights on Lorlatinib Crosses the Blood-Brain Barrier. Front Pharmacol. 2021;12:722627.
18. Miller MR, DiBattista A, Patel MA, Daley M, Tenn C, Nakashima A, Rhind SG, Vartanian O, Shiu MY, Caddy N, Garrett M, Saunders D, Smith I, Jetly R, Fraser DD. A Distinct Metabolite Signature in Military Personnel Exposed to Repetitive Low-Level Blasts. Front Neurol. 2022;13:831792.
19. Massetti N, Russo M, Franciotti R, Nardini D, Mandolini GM, Granzotto A, Bomba M, Delli Pizzi S, Mosca A, Scherer R, Onofrj M, Sensi SL; Alzheimer's Disease Neuroimaging Initiative (ADNI); Alzheimer's Disease Metabolomics Consortium (ADMC). A Machine Learning-Based Holistic Approach to Predict the Clinical Course of Patients within the Alzheimer's Disease Spectrum. J Alzheimers Dis. 2022;85(4):1639-55.
20. Muller E, Algavi YM, Borenstein E. A meta-analysis study of the robustness and universality of gut microbiome-metabolome associations. Microbiome. 2021;9(1):203.
21. Stamate D, Kim M, Proitsi P, Westwood S, Baird A, Nevado-Holgado A, Hye A, Bos I, Vos SJB, Vandenberghe R, Teunissen CE, Kate MT, Scheltens P, Gabel S, Meersmans K, Blin O, Richardson J, De Roeck E, Engelborghs S, Sleegers K, Bordet R, Ramit L, Kettunen P, Tsolaki M, Verhey F, Alcolea D, Lléo A, Peyratout G, Tainta M, Johannsen P, Freund-Levi Y, Frölich L, Dobricic V, Frisoni GB, Molinuevo JL, Wallin A, Popp J, Martinez-Lage P, Bertram L, Blennow K, Zetterberg H, Streffer J, Visser PJ, Lovestone S, Legido-Quigley C. A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: Results from the European Medical Information Framework for Alzheimer disease biomarker discovery cohort. Alzheimers Dement (N Y). 2019;5: 933-8.
22. Pannkuk EL, Laiakis EC, Fornace AJ Jr, Fatanmi OO, Singh VK. A Metabolomic Serum Signature from Nonhuman Primates Treated with a Radiation Countermeasure, Gamma-tocotrienol, and Exposed to Ionizing Radiation. Health Phys. 2018;115(1):3-11.
23. Troisi J, Sarno L, Martinelli P, Di Carlo C, Landolfi A, Scala G, Rinaldi M, D'Alessandro P, Ciccone C, Guida M. A metabolomics-based approach for non-invasive diagnosis of chromosomal anomalies. Metabolomics. 2017;13:11.
24. Troisi J, Landolfi A, Sarno L, Richards S, Symes S, Adair D, Ciccone C, Scala G, Martinelli P, Guida M. A metabolomics-based approach for non-invasive screening of fetal central nervous system anomalies. Metabolomics. 2018;14(6):77.
25. Troisi J, Tafuro M, Lombardi M, Scala G, Richards SM, Symes SJK, Ascierto PA, Delrio P, Tatangelo F, Buonerba C, Pierri B, Cerino P. A Metabolomics-Based Screening Proposal for Colorectal Cancer. Metabolites. 2022;12(2):110.
26. Liang TL, Li RZ, Mai CT, Guan XX, Li JX, Wang XR, Ma LR, Zhang FY, Wang J, He F, Pan HD, Zhou H, Yan PY, Fan XX, Wu QB, Neher E, Liu L, Xie Y, Leung EL, Yao XJ. A method establishment and comparison of in vivo lung cancer model development platforms for evaluation of tumour metabolism and pharmaceutical efficacy. Phytomedicine. 2022;96:153831.
27. Chung RH, Kang CY. A multi-omics data simulator for complex disease studies and its application to evaluate multi-omics data analysis methods for disease classification. Gigascience. 2019;8(5):giz045.
28. Wang X, Wang Z, Weng J, Wen C, Chen H, Wang X. A New Effective Machine Learning Framework for Sepsis Diagnosis. Ieee Access. 2018;6:48300-10.
29. Neumann JM, Niehaus K, Neumann N, Knobloch HC, Bremmer F, Krafft U, Kellner U, Nyirády P, Szarvas T, Bednarz H, Reis H. A new technological approach in diagnostic pathology: mass spectrometry imaging-based metabolomics for biomarker detection in urachal cancer. Lab Invest. 2021;101(9):1281-8.
30. Gupta A, Sagar G, Siddiqui Z, Rao K, Nayak S, Saquib N, Anand R. A non-invasive method for concurrent detection of early-stage women-specific cancers. Sci Rep. 2022;12(1):2301.
31. Wang X, Cui N, Liu X. A novel 6-metabolite signature for prediction of clinical outcomes in type 2 diabetic patients undergoing percutaneous coronary intervention. Cardiovasc Diabetol. 2022;21(1):126.
32. Zacharias H, Altenbuchinger M, Schultheiss U, Samol C, Kotsis F, Poguntke I, Sekula P, Krumsiek J, Köttgen A, Spang R, Oefner P, Gronwald W. A Novel Metabolic Signature To Predict the Requirement of Dialysis or Renal Transplantation in Patients with Chronic Kidney Disease. J Proteome Res. 2019;18(4):1796-805.
33. Blasco H, Patin F, Descat A, Garçon G, Corcia P, Gelé P, Lenglet T, Bede P, Meininger V, Devos D, Gossens JF, Pradat PF. A pharmaco-metabolomics approach in a clinical trial of ALS: Identification of predictive markers of progression. PLoS One. 2018;13(6):e0198116.
34. Bocca C, Kouassi Nzoughet J, Leruez S, Amati-Bonneau P, Ferré M, Kane MS, Veyrat-Durebex C, Chao de la Barca JM, Chevrollier A, Homedan C, Verny C, Miléa D, Procaccio V, Simard G, Bonneau D, Lenaers G, Reynier P. A Plasma Metabolomic Signature Involving Purine Metabolism in Human Optic Atrophy 1 (OPA1)-Related Disorders. Invest Ophthalmol Vis Sci. 2018;59(1):185-95.
35. Olin JW, Di Narzo AF, d'Escamard V, Kadian-Dodov D, Cheng H, Georges A, King A, Thomas A, Barwari T, Michelis KC, Bouchareb R, Bander E, Anyanwu A, Stelzer P, Filsoufi F, Florman S, Civelek

M, Debette S, Jeunemaitre X, Björkegren JLM, Mayr M, Bouatia-Naji N, Hao K, Kovacic JC. A plasma proteogenomic signature for fibromuscular dysplasia. Cardiovasc Res. 2020;116(1):63-77.

36. Bao J, Liu S, Liang X, Wang C, Cao L, Li Z, Wei F, Fu A, Shi Y, Shen B, Zhu X, Zhao Y, Liu H, Miao L, Wang Y, Liang S, Wu L, Huang J, Guo T, Liu F. A prediction model for COVID-19 liver dysfunction in patients with normal hepatic biochemical parameters. Life Sci Alliance. 2022;6(1):e202201576.
37. Troisi J, Colucci A, Cavallo P, Richards S, Symes S, Landolfi A, Scala G, Maiorino F, Califano A, Fabiano M, Silvestre G, Mastella F, Caputo A, D'Antonio A, Altieri V. A Serum Metabolomic Signature for the Detection and Grading of Bladder Cancer. Applied Sci. 2021;11(6):2835.
38. Rawshani A, Eliasson B, Rawshani A, Henninger J, Mardinoglu A, Carlsson Å, Sohlin M, Ljungberg M, Hammarstedt A, Rosengren A, Smith U. Adipose tissue morphology, imaging and metabolomics predicting cardiometabolic risk and family history of type 2 diabetes in non-obese men. Sci Rep. 2020;10(1):9973.
39. Anwar A, Abruzzo PM, Pasha S, Rajpoot K, Bolotta A, Ghezzo A, Marini M, Posar A, Visconti P, Thornalley PJ, Rabbani N. Advanced glycation endproducts, dityrosine and arginine transporter dysfunction in autism – a source of biomarkers for clinical diagnosis. Mol Autism. 2018;9:3.
40. Tang Y, El-Chemaly S, Taveira-Dasilva A, Goldberg HJ, Bagwe S, Rosas IO, Moss J, Priolo C, Henske EP. Alterations in Polyamine Metabolism in Patients With Lymphangioleiomyomatosis and Tuberous Sclerosis Complex 2-Deficient Cells. Chest. 2019;156(6):1137-48.
41. Chang KH, Lin CN, Chen CM, Lyu RK, Chu CC, Liao MF, Huang CC, Chang HS, Ro LS, Kuo HC. Altered Metabolic Profiles of the Plasma of Patients with Amyotrophic Lateral Sclerosis. Biomedicines. 2021;9(12):1944.
42. Sapkota S, Huan T, Tran T, Zheng J, Camicioli R, Li L, Dixon RA. Alzheimer's Biomarkers From Multiple Modalities Selectively Discriminate Clinical Status: Relative Importance of Salivary Metabolomics Panels, Genetic, Lifestyle, Cognitive, Functional Health and Demographic Risk Markers. Front Aging Neurosci. 2018;10:296.
43. Vicente-Dueñas C, Janssen S, Oldenburg M, Auer F, González-Herrero I, Casado-García A, Isidro-Hernández M, Raboso-Gallego J, Westhoff P, Pandyra AA, Hein D, Gössling KL, Alonso-López D, De Las Rivas J, Bhatia S, García-Criado FJ, García-Cenador MB, Weber APM, Köhrer K, Hauer J, Fischer U, Sánchez-García I, Borkhardt A. An intact gut microbiome protects genetically predisposed mice against leukemia. Blood. 2020;136(18):2003-17.
44. Siegel AP, Daneshkhah A, Hardin DS, Shrestha S, Varahramyan K, Agarwal M. Analyzing breath samples of hypoglycemic events in type 1 diabetes patients: towards developing an alternative to diabetes alert dogs. J Breath Res. 2017;11(2):026007.
45. Huang Q, Zhang X, Hu Z. Application of Artificial Intelligence Modeling Technology Based on Multi-Omics in Noninvasive Diagnosis of Inflammatory Bowel Disease. J Inflamm Res. 2021;14:1933-43.
46. Irajizad E, Wu R, Vykoukal J, Murage E, Spencer R, Dennison JB, Moulder S, Ravenberg E, Lim B, Litton J, Tripathym D, Valero V, Damodaran S, Rauch GM, Adrada B, Candelaria R, White JB, Brewster A, Arun B, Long JP, Do KA, Hanash S, Fahrmann JF. Application of Artificial Intelligence to Plasma Metabolomics Profiles to Predict Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. Front Artif Intell. 2022;5:876100.
47. Bahado-Singh RO, Radhakrishna U, Gordevičius J, Aydas B, Yilmaz A, Jafar F, Imam K, Maddens M, Challapalli K, Metpally RP, Berrettini WH, Crist RC, Graham SF, Vishweswaraiah S. Artificial Intelligence and Circulating Cell-Free DNA Methylation Profiling: Mechanism and Detection of Alzheimer's Disease. Cells. 2022;11(11):1744.
48. Proitsi P, Kim M, Whiley L, Simmons A, Sattlecker M, Velayudhan L, Lupton MK, Soininen H, Kloszewska I, Mecocci P, Tsolaki M, Vellas B, Lovestone S, Powell JF, Dobson RJ, Legido-Quigley C. Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis. Alzheimers Dement. 2017;13(2):140-51.
49. Borkowski K, Pedersen TL, Seyfried NT, Lah JJ, Levey AI, Hales CM, Dammer EB, Blach C, Louie G, Kaddurah-Daouk R, Newman JW; Alzheimer's Disease Metabolomics Consortium. Association of plasma and CSF cytochrome P450, soluble epoxide hydrolase, and ethanolamide metabolism with Alzheimer's disease. Alzheimers Res Ther. 2021;13(1):149.
50. Kehoe ER, Fitzgerald BL, Graham B, Islam MN, Sharma K, Wormser GP, Belisle JT, Kirby MJ. Biomarker selection and a prospective metabolite-based machine learning diagnostic for lyme disease. Sci Rep. 2022;12(1):1478.
51. Varma VR, Oommen AM, Varma S, Casanova R, An Y, Andrews RM, O'Brien R, Pletnikova O, Troncoso JC, Toledo J, Baillie R, Arnold M, Kastenmueller G, Nho K, Doraiswamy PM, Saykin AJ, Kaddurah-Daouk R, Legido-Quigley C, Thambisetty M. Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. PLoS Med. 2018;15(1):e1002482.
52. Khan MS, Cuda S, Karere GM, Cox LA, Bishop AC. Breath biomarkers of insulin resistance in pre-diabetic Hispanic adolescents with obesity. Sci Rep. 2022;12(1):339.
53. Miller-Atkins G, Acevedo-Moreno LA, Grove D, Dweik RA, Tonelli AR, Brown JM, Allende DS, Aucejo F, Rotroff DM. Breath Metabolomics Provides an Accurate and Noninvasive Approach for Screening Cirrhosis, Primary, and Secondary Liver Tumors. Hepatol Commun. 2020;4(7):1041-55.
54. Gbaoui L, Fachet M, Lüno M, Meyer-Lotz G, Frodl T, Hoeschen C. Breathomics profiling of metabolic pathways affected by major depression: Possibilities and limitations. Front Psychiatry. 2022;13:1061326.
55. Kaur P, Singh A, Chana I. BSense: A parallel Bayesian hyperparameter optimized Stacked ensemble model for breast cancer survival prediction. J Computational Sci. 2022;60:101570.
56. Zhou R, Li J, Zhang Y, Xiao H, Zuo Y, Ye L. Characterization of plasma metabolites and proteins in patients with herpetic neuralgia and development of machine learning predictive models based on metabolomic profiling. Front Mol Neurosci. 2022;15:1009677.

57. Tiedt S, Brandmaier S, Kollmeier H, Duering M, Artati A, Adamski J, Klein M, Liebig T, Holdt LM, Teupser D, Wang-Sattler R, Schwedhelm E, Gieger C, Dichgans M. Circulating Metabolites Differentiate Acute Ischemic Stroke from Stroke Mimics. Ann Neurol. 2020;88(4):736-46.
58. Zhou G, Lu J, Xu T, Lu Y, Chen W, Wang J, Ke M. Clinical lipidomics analysis reveals biomarkers of lipid peroxidation in serum from patients with rheumatoid arthritis. Microchem J. 2021;169:106607.
59. Li N, Yang C, Zhou S, Song S, Jin Y, Wang D, Liu J, Gao Y, Yang H, Mao W, Chen Z. Combination of Plasma-Based Metabolomics and Machine Learning Algorithm Provides a Novel Diagnostic Strategy for Malignant Mesothelioma. Diagnostics (Basel). 2021;11(7):1281.
60. Lai X, Guo K, Huang W, Su Y, Chen S, Li Q, Liang K, Gao W, Wang X, Chen Y, Wang H, Lin W, Wei X, Ni W, Lin Y, Jiang D, Cheng YH, Che CM, Ng KM. Combining MALDI-MS with machine learning for metabolomic characterization of lung cancer patient sera. Anal Methods. 2022;14(5):499-507.
61. Hu X, Wang J, Ju Y, Zhang X, Qimanguli W, Li C, Yue L, Tuohetaerbaike B, Li Y, Wen H, Zhang W, Chen C, Yang Y, Wang J, Chen F. Combining metabolome and clinical indicators with machine learning provides some promising diagnostic markers to precisely detect smear-positive/negative pulmonary tuberculosis. BMC Infect Dis. 2022;22(1):707.
62. Yang Y, Yang Q, Luo S, Zhang Y, Lian C, He H, Zeng J, Zhang G. Comparative Analysis Reveals Novel Changes in Plasma Metabolites and Metabolomic Networks of Infants With Retinopathy of Prematurity. Invest Ophthalmol Vis Sci. 2022;63(1):28.
63. Hao L, Wang J, Page D, Asthana S, Zetterberg H, Carlsson C, Okonkwo OC, Li L. Comparative Evaluation of MS-based Metabolomics Software and Its Application to Preclinical Alzheimer's Disease. Sci Rep. 2018;8(1):9291.
64. Koureas M, Kalompatsios D, Amoutzias GD, Hadjichristodoulou C, Gourgoulianis K, Tsakalof A. Comparison of Targeted and Untargeted Approaches in Breath Analysis for the Discrimination of Lung Cancer from Benign Pulmonary Diseases and Healthy Persons. Molecules. 2021;26(9):2609.
65. Gal J, Bailleux C, Chardin D, Pourcher T, Gilhodes J, Jing L, Guigonis JM, Ferrero JM, Milano G, Mograbi B, Brest P, Chateau Y, Humbert O, Chamorey E. Comparison of unsupervised machine-learning methods to identify metabolomic signatures in patients with localized breast cancer. Comput Struct Biotechnol J. 2020;3;18:1509-24.
66. Xiao Y, Ma D, Yang YS, Yang F, Ding JH, Gong Y, Jiang L, Ge LP, Wu SY, Yu Q, Zhang Q, Bertucci F, Sun Q, Hu X, Li DQ, Shao ZM, Jiang YZ. Comprehensive metabolomics expands precision medicine for triple-negative breast cancer. Cell Res. 2022;32(5): 477-90.
67. Delafiori J, Navarro LC, Siciliano RF, de Melo GC, Busanello ENB, Nicolau JC, Sales GM, de Oliveira AN, Val FFA, de Oliveira DN, Eguti A, Dos Santos LA, Dalçóquio TF, Bertolin AJ, Abreu-Netto RL, Salsoso R, Baía-da-Silva D, Marcondes-Braga FG, Sampaio VS, Judice CC, Costa FTM, Durán N, Perroud MW, Sabino EC, Lacerda MVG, Reis LO, Fávaro WJ, Monteiro WM, Rocha AR, Catharino RR. Covid-19 Automated Diagnosis and Risk Assessment through Metabolomics and Machine Learning. Anal Chem. 2021;2;93(4):2471-9.
68. Villagrana-Bañuelos KE, Maeda-Gutiérrez V, Alcalá-Rmz V, Oropeza-Valdez JJ, Herrera-Van Oostdam AS, Castañeda-Delgado JE, López JA, Borrego Moreno JC, Galván-Tejada CE, Galván-Tejeda JI, Gamboa-Rosales H, Luna-García H, Celaya-Padilla JM, López-Hernández Y. COVID-19 Outcome Prediction by Integrating Clinical and Metabolic Data using Machine Learning Algorithms. Rev Invest Clin. 2022;74(6):314-27.
69. Liu W, Zhang L, Shi X, Shen G, Feng J. Cross-comparative metabolomics reveal sex-age specific metabolic fingerprints and metabolic interactions in acute myocardial infarction. Free Radic Biol Med. 2022;183:25-34.
70. Alakwaa FM, Chaudhary K, Garmire LX. Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. J Proteome Res. 2018;5;17(1):337-47.
71. Zhang Y, Zhu D, Li T, Wang X, Zhao L, Yang X, Dang M, Li Y, Wu Y, Lu Z, Lu J, Jian Y, Wang H, Zhang L, Lu X, Shen Z, Fan H, Cai W, Zhang G. Detection of acute ischemic stroke and backtracking stroke onset time via machine learning analysis of metabolomics. Biomed Pharmacother. 2022;155:113641.
72. Johno H, Yoshimura K, Mori Y, Kimura T, Niimi M, Yamada M, Tanigawa T, Fan J, Takeda S. Detection of potential new biomarkers of atherosclerosis by probe electrospray ionization mass spectrometry. Metabolomics. 2018;27;14(4):38.
73. Wang J, Yan D, Zhao A, Hou X, Zheng X, Chen P, Bao Y, Jia W, Hu C, Zhang ZL, Jia W. Discovery of potential biomarkers for osteoporosis using LC-MS/MS metabolomic methods. Osteoporos Int. 2019;30(7):1491-9.
74. Miller HA, Emam R, Lynch CM, Bockhorst S, Frieboes HB. Discrepancies in metabolomic biomarker identification from patient-derived lung cancer revealed by combined variation in data pre-treatment and imputation methods. Metabolomics. 2021;27;17(4):37.
75. Trezzi JP, Galozzi S, Jaeger C, Barkovits K, Brockmann K, Maetzler W, Berg D, Marcus K, Betsou F, Hiller K, Mollenhauer B. Distinct metabolomic signature in cerebrospinal fluid in early Parkinson's disease. Mov Disord. 2017;32(10):1401-8.
76. Xie Y, Meng WY, Li RZ, Wang YW, Qian X, Chan C, Yu ZF, Fan XX, Pan HD, Xie C, Wu QB, Yan PY, Liu L, Tang YJ, Yao XJ, Wang MF, Leung EL. Early lung cancer diagnostic biomarker discovery by machine learning methods. Transl Oncol. 2021;14(1):100907.
77. Peddinti G, Cobb J, Yengo L, Froguel P, Kravić J, Balkau B, Tuomi T, Aittokallio T, Groop L. Early metabolic markers identify potential targets for the prevention of type 2 diabetes. Diabetologia. 2017;60(9):1740-50.
78. Miller HA, Yin X, Smith SA, Hu X, Zhang X, Yan J, Miller DM, van Berkel VH, Frieboes HB. Evaluation of disease staging and chemotherapeutic response in non-small cell lung cancer from patient tumor-derived metabolomic data. Lung Cancer. 2021;156: 20-30.
79. Akyol S, Yilmaz A, Oh KJ, Ugur Z, Aydas B, McGuinness B, Passmore P, Kehoe PG, Maddens M, Green BD, Graham SF. Evidence that the Kennedy and polyamine pathways are dysregulated

in human brain in cases of dementia with Lewy bodies. Brain Res. 2020;15;1743:146897.

80. Sinha A, Desiraju K, Aggarwal K, Kutum R, Roy S, Lodha R, Kabra SK, Ghosh B, Sethi T, Agrawal A. Exhaled breath condensate metabolome clusters for endotype discovery in asthma. J Transl Med. 2017;22;15(1):262.
81. Beccaria M, Bobak C, Maitshotlo B, Mellors TR, Purcaro G, Franchina FA, Rees CA, Nasir M, Stevens WS, Scott LE, Black A, Hill JE. Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. J Breath Res. 2018;5;13(1):016005.
82. Iwano T, Yoshimura K, Watanabe G, Saito R, Kiritani S, Kawaida H, Moriguchi T, Murata T, Ogata K, Ichikawa D, Arita J, Hasegawa K, Takeda S. High-performance Collective Biomarker from Liquid Biopsy for Diagnosis of Pancreatic Cancer Based on Mass Spectrometry and Machine Learning. J Cancer. 2021;4;12(24): 7477-87.
83. Abdullah M, Wah Y, Majeed A, Zakaria Y, Shaadan N. Identification of Blood-Based Multi-Omics Biomarkers for Alzheimer's Disease Using Firth's Logistic Regression. Pertanika J Sci Technol. 2022;30:1197-218.
84. Lin CN, Hsu KC, Huang KL, Huang WC, Hung YL, Lee TH. Identification of Metabolomics Biomarkers in Extracranial Carotid Artery Stenosis. Cells. 2022;27;11(19):3022.
85. Eng A, Hayden HS, Pope CE, Brittnacher MJ, Vo AT, Weiss EJ, Hager KR, Leung DH, Heltshe SL, Raftery D, Miller SI, Hoffman LR, Borenstein E. Infants with cystic fibrosis have altered fecal functional capacities with potential clinical and metabolic consequences. BMC Microbiol. 2021;15;21(1):247.
86. Webb-Robertson BM, Nakayasu ES, Frohnert BI, Bramer LM, Akers SM, Norris JM, Vehik K, Ziegler AG, Metz TO, Rich SS, Rewers MJ. Integration of Infant Metabolite, Genetic, and Islet Autoimmunity Signatures to Predict Type 1 Diabetes by Age 6 Years. J Clin Endocrinol Metab. 2022;14;107(8):2329-38.
87. Glaab E, Trezzi JP, Greuel A, Jäger C, Hodak Z, Drzezga A, Timmermann L, Tittgemeyer M, Diederich NJ, Eggers C. Integrative analysis of blood metabolomics and PET brain neuroimaging data for Parkinson's disease. Neurobiol Dis. 2019;124: 555-62.
88. An R, Yu H, Wang Y, Lu J, Gao Y, Xie X, Zhang J. Integrative analysis of plasma metabolomics and proteomics reveals the metabolic landscape of breast cancer. Cancer Metab. 2022;17;10(1):13.
89. Gilard V, Ferey J, Marguet F, Fontanilles M, Ducatez F, Pilon C, Lesueur C, Pereira T, Basset C, Schmitz-Afonso I, Di Fioré F, Laquerrière A, Afonso C, Derrey S, Marret S, Bekri S, Tebani A. Integrative Metabolomics Reveals Deep Tissue and Systemic Metabolic Remodeling in Glioblastoma. Cancers (Basel). 2021;14;13(20):5157.
90. Celaya-Padilla JM, Villagrana-Bañuelos KE, Oropeza-Valdez JJ, Monárrez-Espino J, Castañeda-Delgado JE, Oostdam ASH, Fernández-Ruiz JC, Ochoa-González F, Borrego JC, Enciso-Moreno JA, López JA, López-Hernández Y, Galván-Tejada CE. Kynurenine and Hemoglobin as Sex-Specific Variables in COVID-19 Patients: A Machine Learning and Genetic Algorithms Approach. Diagnostics (Basel). 2021;25;11(12):2197.
91. Buszewska-Forajta M, Pomastowski P, Monedeiro F, Walczak-Skierska J, Markuszewski M, Matuszewski M, Markuszewski MJ, Buszewski B. Lipidomics as a Diagnostic Tool for Prostate Cancer. Cancers (Basel). 2021;21;13(9):2000.
92. Wang G, Qiu M, Xing X, Zhou J, Yao H, Li M, Yin R, Hou Y, Li Y, Pan S, Huang Y, Yang F, Bai F, Nie H, Di S, Guo L, Meng Z, Wang J, Yin Y. Lung cancer scRNA-seq and lipidomics reveal aberrant lipid metabolism for early-stage diagnosis. Sci Transl Med. 2022;2;14(630):eabk2756.
93. Miller HA, van Berkel VH, Frieboes HB. Lung cancer survival prediction and biomarker identification with an ensemble machine learning analysis of tumor core biopsy metabolomic data. Metabolomics. 2022;20;18(8):57.
94. Gao B, Wu TC, Lang S, Jiang L, Duan Y, Fouts DE, Zhang X, Tu XM, Schnabl B. Machine Learning Applied to Omics Datasets Predicts Mortality in Patients with Alcoholic Hepatitis. Metabolites. 2022;5;12(1):41.
95. Kosyakovsky LB, Somerset E, Rogers AJ, Sklar M, Mayers JR, Toma A, Szekely Y, Soussi S, Wang B, Fan CS, Baron RM, Lawler PR. Machine learning approaches to the human metabolome in sepsis identify metabolic links with survival. Intensive Care Med Exp. 2022;17;10(1):24.
96. Yilmaz A, Ugur Z, Ustun I, Akyol S, Bahado-Singh RO, Maddens M, Aasly JO, Graham SF. Metabolic Profiling of CSF from People Suffering from Sporadic and LRRK2 Parkinson's Disease: A Pilot Study. Cells. 2020;31;9(11):2394.
97. Liu J, Li ZB, Lu QQ, Yu Y, Zhang SQ, Ke PF, Zhang F, Li JC. Metabolite profile of COVID-19 revealed by UPLC-MS/MS-based widely targeted metabolomics. Front Immunol. 2022;18;13: 894170.
98. Rahnavard A, Mann B, Giri A, Chatterjee R, Crandall K. Metabolite, protein, and tissue dysfunction associated with COVID-19 disease severity. Sci Rep. 2022;12(1):12204.
99. Njoku K, Campbell AE, Geary B, MacKintosh ML, Derbyshire AE, Kitson SJ, Sivalingam VN, Pierce A, Whetton AD, Crosbie EJ. Metabolomic Biomarkers for the Detection of Obesity-Driven Endometrial Cancer. Cancers (Basel). 2021;10;13(4):718.
100. Cheng SC, Chen K, Chiu CY, Lu KY, Lu HY, Chiang MH, Tsai CK, Lo CJ, Cheng ML, Chang TC, Lin G. Metabolomic biomarkers in cervicovaginal fluid for detecting endometrial cancer through nuclear magnetic resonance spectroscopy. Metabolomics. 2019;29;15(11):146.
101. Qureshi F, Adams JB, Audhya T, Hahn J. Multivariate Analysis of Metabolomic and Nutritional Profiles among Children with Autism Spectrum Disorder. J Pers Med. 2022;1;12(6):923.
102. Dimitri GM, Meoni G, Tenori L, Luchinat C, Lió P; on behalf of the PROPAG-AGEING Consortium. NMR Spectroscopy Combined with Machine Learning Approaches for Age Prediction in Healthy and Parkinson's Disease Cohorts through Metabolomic Fingerprints. Appl Sci. 2022;12;8954.