Original Article

Chemometric-empowered spectroscopic techniques in pharmaceutical fields: A bibliometric analysis and updated review

Baidillah Zulkifli[1], Fajar Fakri[2], Joan Odigie[3], Loveday Nnabuife[3], Chinwe C. Isitua[4] and Williams Chiari[5,6*]

[1]Laboratory of Physiology, Faculty of Veterinary Medicine, Universitas Syiah Kuala, Banda Aceh, Indonesia; [2]Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh, Indonesia; [3]Department of Medical Laboratory Science, Ring Road State Hospital, Ibadan, Oyo State, Nigeria; [4]Department of Biology and Forensic Science, Faculty of Science, Admiralty University of Nigeria, Ibusa, Delta State, Nigeria; [5]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh, Indonesia; [6]Innovative Sustainability Lab, PT. Biham Riset dan Edukasi, Banda Aceh, Indonesia

*Corresponding author: williamschiari@gmail.com

Abstract

Undeniable increase in pharmaceutical demand has encouraged researchers to develop analytical techniques to analyze drugs and monitor their effects. Chemometrics enabled simple spectroscopic approaches (such as infrared or UV spectroscopy) to analyze complex samples, including drug formulations, excreted fluids and tissues of living organisms. The aim of this study was to evaluate the research trend of this combinatorial technique utilized for pharmaceutical-related sample analysis using a bibliometric analysis. Bibliometric data of published literature from the Scopus database on March 14, 2023 were retrieved using the keyword combinations of "multivariate", "chemometrics", "pattern recognition", "drug", "pharmaceutical" and "spectroscopy". Network visualization analysis was performed using VOSviewer on the co-occurring keywords and authorships, presenting data such as top cited papers (n=10). The literature review was performed based on the research trend revealed by the clusters that emerged in the network visualization. The analysis revealed that the first paper was published in 1973 (n=1) and a total of 3544 records have been published as of March 14, 2023, comprising original research articles (n=3144, 88.71%) and review articles (n=232, 6.55%). The keyword “chemometrics” with Total Link Strength (TLS) of 826 emerged as the most abundant, followed by “metabolomics” (TLS=388), “Raman spectroscopy” (TLS=280), “metabonomics” (TLS=272), “nuclear magnetic resonance” or “NMR” (TLS=271), and “multivariate analysis” (TLS=254). Network visualization revealed that the research falls into two general categories: (1) drug toxicity and efficacy monitoring and (2) quality control of drug manufacturing. The top cited paper (n=3269) was a review article published in 1999 describing the utility of nuclear magnetic resonance combined with multivariate statistics for metabolite profiling of biological samples. The chemometric-empowered spectroscopy techniques were expected to provide objective measurement during clinical studies and monitoring of therapeutic effects.

Keywords: FTIR, infrared, network visualization, NMR, PCA, research landscape

Introduction

The pharmaceutical field has seen an acceleration in drug discovery and development, which consequently leads to the increasing importance of analytical methods [1, 2]. Analytical methods

Received: April 4, 2023 - Accepted: April 27, 2023 - Published online: April 30, 2023

are essential in stages of drug development to secure the quality, efficacy, and safety of pharmaceutical products, starting from drug discovery to quality control [3]. Such methods lead to the use of spectroscopy in combination with chemometrics as critical analytical tools in pharmaceutical development. Spectroscopy measures and interprets the interaction between electromagnetic radiation and matter, including the molecular structure, properties, and reactions [4]. When combined with chemometrics, users could extract useful information from a certain chemical system [5]. Some examples of spectroscopic techniques used in combination with chemometrics are the near-infrared (NIR) spectroscopy [6], Raman spectroscopy [7], and nuclear magnetic resonance (NMR) [8].

In response to the rising demands of drug discovery, researchers have begun to further their attempts to find a breakthrough in chemometrics and spectroscopy in pharmaceutical development. Modernized analytical methods have shown its use, one of them being the advanced Raman spectroscopy which enables the support of novel on-site applications of pharmaceutical processes [9]. Recent years have also shown considerable importance in involving process analytical technology (PAT) to improve pharmaceutical product's quality efficiency, due to its irreplaceable role in ultra-rapid monitoring and control [10]. In particular, multivariate data analysis has been widely reported for its utilization in controlling the quality of manufactured drugs. Its calibration model is able to predict chemical properties from a set of predictor variables, which is then applied to the NIR spectroscopy [11].

The recent issue lies in the lack of literature review and bibliometric analysis on recapping the current progress of chemometrics and spectroscopy usage in pharmaceutical development. This problem leads to the development not pacing well to the rising demand for further pharmaceutical research. Some other potential hotspots are still yet to be assessed further to hasten the aforementioned development. Herein, we performed a bibliometric analysis and literature review based on the ongoing progress of chemometrics and spectroscopy in pharmaceutical research. We gathered the metadata of the fifty-year published literature. Bibliometric studies have been widely used to assess the trend of research in the medical field and to find out any potential research hotspots yet to be further investigated [12-14]. This study delivered more information to researchers dealing with chemometrics and spectroscopy related research in pharmaceutical development.

Methods

Study design

In this study, a bibliometric analysis was conducted to assess the impact of chemometrics and spectroscopy in the pharmaceutical field. The metadata of published papers was retrieved from the Scopus database (dated 14 March 2023), which consists of papers reporting the utility of chemometrics and spectroscopy for pharmaceutical-related products or samples. VOSviewer 1.6.17 (Center for Science and Technology Studies, Leiden University, The Netherlands) was used to analyze the retrieved data via network visualization analysis [15]. Then, the literature review was performed based on the trend suggested by the network visualization analysis.

Search strategy

The literature search on the Scopus database used the following combination of keywords: ("multivariate" OR "chemometrics" OR "pattern recognition" AND "drug" OR "pharmaceutical" AND "spectroscopy") to ensure the relevancy of the retrieved papers. The aforementioned search was used on all titles, abstracts, and keywords of the published paper. Some limitations during the investigation were the document type, publication stage, and exclusion of non-English papers (**Figure 1**). The retrieved information was extracted from the Scopus database, such as the authors' information, title, abstract, and journal's keywords, and title, all of which were exported in a CSV file (.csv).

Keyword analysis and network visualization

Once exported, the retrieved data from the Scopus database was assessed to cover the characteristics, research theme & clusters of the use of chemometrics and spectroscopy on

pharmaceutical research, along with the keywords occurrences. Network visualization analysis was then conducted using VOSviewer to generate mappings for each of the keyword co-occurrences and countries' co-authorship. The mappings also included the density visualization to further assess the focus on the distribution of the keywords and countries involved in the related study.

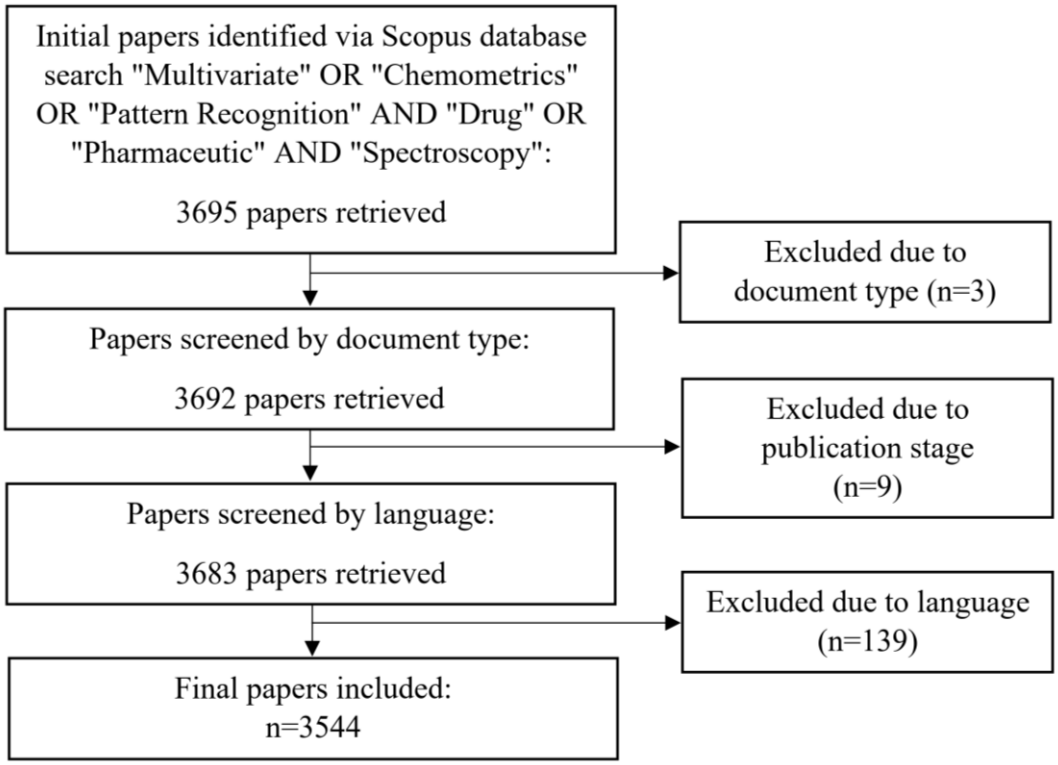

Figure 1. Flowchart of the extraction process of the bibliometric data from reports covering spectroscopy combined with chemometric analysis for pharmaceutical-related samples.

Selection of most cited paper

The most cited papers were selected through the number of citations as the parameter of the scientific report's impact on the related study. The filtering was then enacted to ensure the closeness of the selected papers to chemometrics and spectroscopy studies in pharmaceutical development. Thus, slightly related and unrelated papers will be excluded from the list.

Results

Characteristics of included paper on chemometrics and spectroscopy study in pharmaceutical development research

A total of 3544 papers were retrieved from the Scopus database, consisting of original research articles (n=3144, 88.71%), review articles (n=232, 6.55%) and other documents, including conference papers, book chapters, editorial and letters (n=168, 4.74%). The detailed document types of the metadata were presented in **Table 1**. In terms of the year published, the publication of chemometrics and spectroscopy studies in pharmaceutical research saw a significant increase almost every year, most notable being papers published between 2003 to 2012, with a whopping 438% rise in numbers. The publications then reached a stable number of papers released for the next decade, and never fell below 200 papers per year. The trend of chemometrics and spectroscopy study in pharmaceutical research based on the year published is presented in **Figure 2**.

Table 1. Distribution of the publication type on chemometrics and spectroscopy application in the pharmaceutical field

Document type	Document, n (%)
Article	3144 (88.71)
Review	232 (6.55)
Conference paper	104 (2.93)
Book chapter	38 (1.07)
Editorial	7 (0.2)
Note	6 (0.17)
Short survey	6 (0.17)
Conference review	5 (0.14)
Book	1 (0.03)
Letter	1 (0.03)

In terms of subject field distribution, Chemistry had the highest number of papers published, followed by Biochemistry, Genetics & Molecular Biology and Pharmacology, Toxicology & Pharmaceuticals. These three research themes were the only ones discussed in more than 1000 papers. The top ten related subject fields along with their respective document frequency have been presented in **Figure 3**.

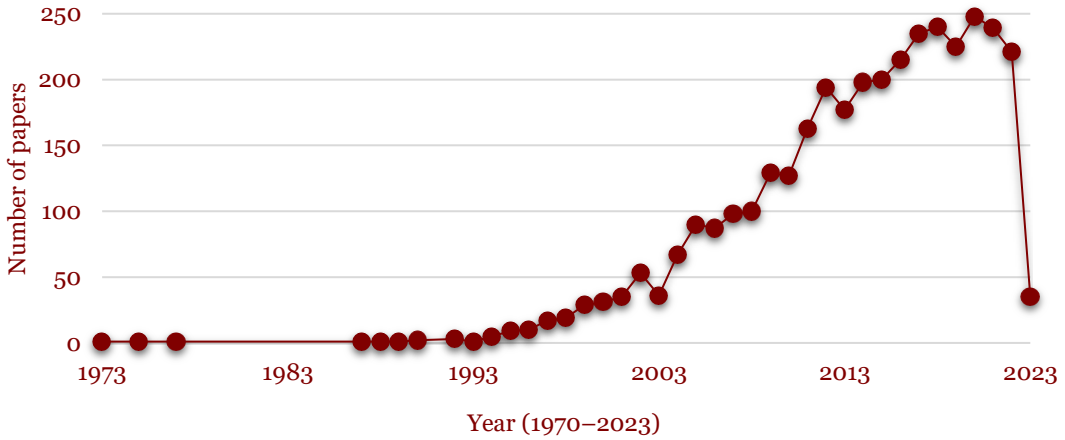

Figure 2. Annual research trend of chemometrics and spectroscopy application for pharmaceutical-related samples (1973–2023).

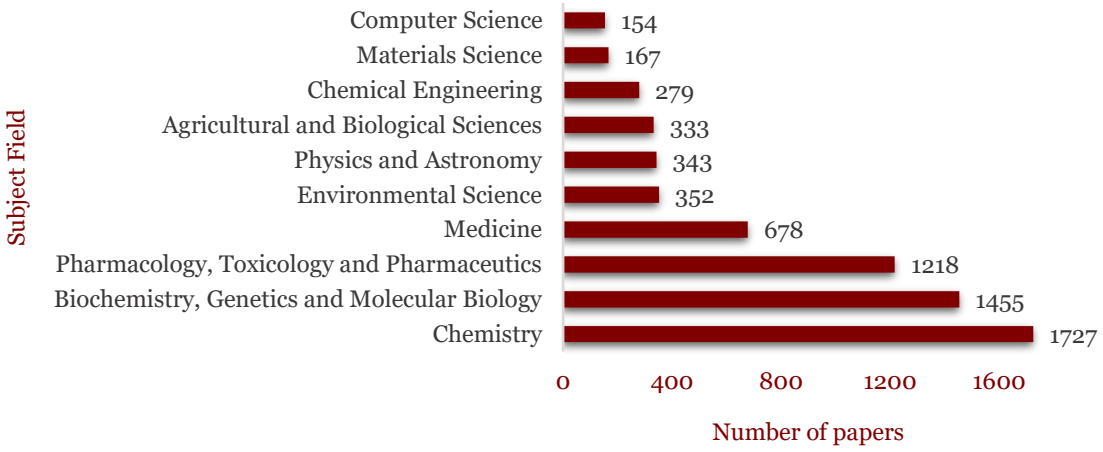

Figure 3. Top ten related subject field distribution of chemometrics and spectroscopy application for pharmaceutical-related samples.

Chemometrics and spectroscopy study in pharmaceutical research based on the keyword occurrence frequency

The ten most common keywords used in chemometrics and spectroscopy studies in pharmaceutical research according to the keyword occurrences are presented in **Figure 4**. The

occurrences showed the number of times a certain keyword is used in the metadata of retrieved papers. 'Chemometrics', 'metabolomics', 'Raman spectroscopy' and 'metabonomics' are the top four keywords occurred in the papers, with 'chemometrics' significantly dominating the others, almost twice as much as the second-ranked most used keyword.

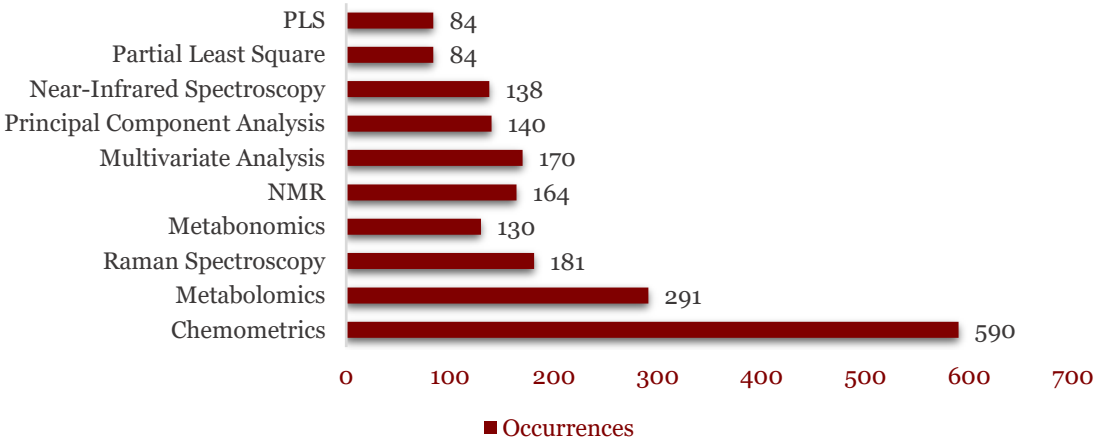

Figure 4. The most frequently occurred keywords of studies reporting chemometrics and spectroscopy application for pharmaceutical-related samples.

Co-occurrences of all keywords

The network visualization and density visualization of the co-occurrences of all keywords have been presented in **Figure 5**. Certain limitations were enacted to better analyze the keywords' co-occurrences, where the minimum number of occurrences of a keyword was set to 15. The most used keyword was determined by identifying the keyword's number of Total Link Strength (TLS), indicating the total strength of the co-occurrence links of a given keyword to other keywords. Of 8246 keywords detected from the metadata, 86 keywords met the threshold. In line with the most frequently occurred keywords, the most used keyword in chemometrics and spectroscopy studies in pharmaceutical research was 'chemometrics' (TLS=826), followed by 'metabolomics' (TLS=388), 'Raman spectroscopy' (TLS=280), 'metabonomics' (TLS=272), 'Nuclear Magnetic Resonance' or 'NMR' (TLS=271), and 'multivariate analysis' (TLS=254). These six keywords were the only ones to pass 250 TLS, representing slightly more than 30% of all other keywords combined.

The colors shown in the network visualization indicated the closeness of certain keywords to a research cluster (**Figure 5**). There were five clusters found in the chemometrics and spectroscopy study in pharmaceutical research, distinguished by their color. These clusters were comprised of keywords associated to 'NMR spectroscopy for metabolite analysis in urinary samples' (red), 'chemometric methods in analysing Raman and NIR spectroscopy data' (green), 'Raman spectroscopy for identifying polymorphism with multivariate statistics' (blue), 'quality control using PCA and FTIR spectroscopy' (yellow), and 'discrimination of drugs using FTIR combined with partial least square' (purple).

The density visualization shows different colors for certain keywords, indicating how other keywords revolve around it (**Figure 5**). Keywords in the red area indicated keywords with the highest density, followed by yellow, green, cyan. and dark blue. 'Chemometrics' and 'metabolomics' were the keywords with the highest density, where all other keywords revolved around other keywords. A small portion of keywords were also seen to revolve around 'Raman spectroscopy' and 'NMR'.

Clusters from the network visualization of keyword co-occurrence

Clusters identified from the previous network visualization showed the research topics found in chemometrics and spectroscopy studies for pharmaceutical-related samples. The clusters and suggested research hotspots ranked based on their size have been presented in **Table 2**. Five clusters were identified from the network visualization, with the main cluster found to suggest

the research hotspot 'metabolomics and NMR spectroscopy of urine samples.' The research hotspot was considered acceptable due to how common NMR spectroscopy is used to perform an analytical technique to study the chemical and physical characteristics of molecules, including metabolites.

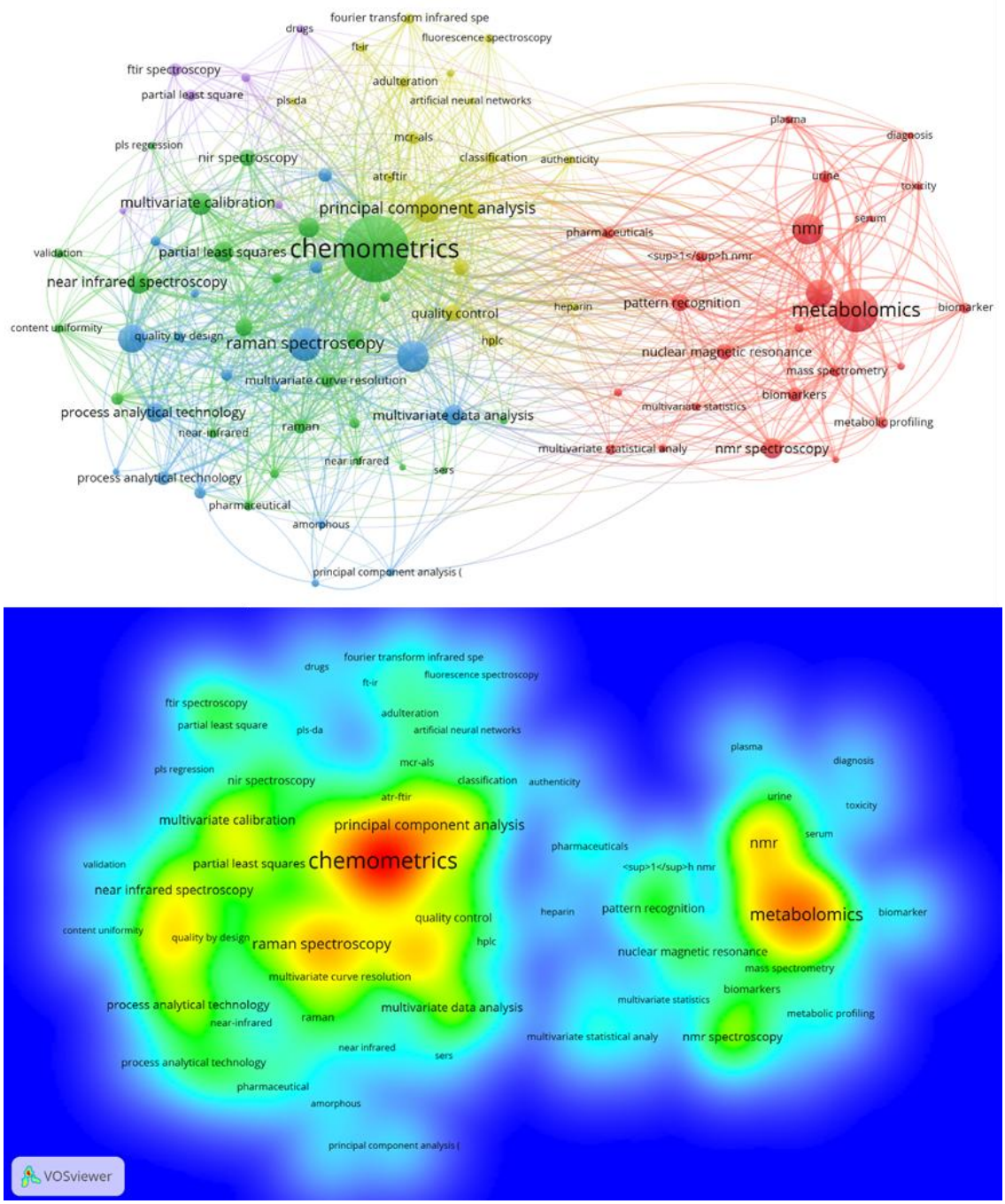

Figure 5. Network visualization of all keywords (weight: occurrences) (top) and density visualization of all keywords (weights: occurrences, score: average publications per year) (bottom). Nodes with the same color in 3A represent a research cluster.

The chemometric methods in analyzing Raman and NIR spectroscopy data' research hotspot indicated the application of statistical and mathematical methods in chemical data, using the partial least square (PLS) to analyze spectroscopy data such as Raman and Near Infrared (NIR). 'Multivariate analysis in polymorphism studies' research hotspot suggested the use of Raman spectroscopy and multivariate analysis to study polymorphism using Process Analytical Technology (PAT). 'Quality control using PCA and FTIR spectroscopy' indicated the use of Principal Component Analysis (PCA) and Fourier Transform-Infrared (FTIR) spectroscopy in

controlling the quality of chemical samples. Lastly, the 'drugs analysis using spectroscopy techniques' research hotspot is likely related to the discriminant analysis and FTIR and attenuated total reflectance (ATR)-FTIR spectroscopy in analyzing drugs.

Table 2. Clusters formed by keywords' co-occurrences of chemometrics and spectroscopy study in pharmaceutical research

Cluster	Top keywords	Research hotspot(s)
I	Metabolomics, metabonomics, NMR, NMR spectroscopy, urine	NMR spectroscopy for metabolite analysis in urinary samples
II	Chemometrics, PLS, Raman, NIR, partial least squares	Chemometric methods in analyzing Raman and NIR spectroscopy data
III	Raman spectroscopy, multivariate analysis, multivariate data analysis, process analytical technology, polymorphism	Raman spectroscopy for identifying polymorphism with multivariate statistics
IV	Quality control, principal component analysis, classification, FTIR, PCA	Quality control using PCA and FTIR spectroscopy
V	Drugs, discriminant analysis, FTIR spectroscopy, partial least square, ATR-FTIR spectroscopy	Discrimination of drugs using FTIR combined with partial least square

Co-authorship country

The network visualization and density visualization mappings of the co-authorship countries have been presented in **Figure 6**, aiming to interpret the countries' co-authorship and collaboration between each identified group of countries. Restrictions were enacted on the mapping to refine the countries' co-authorship analysis, where the minimum number of documents and citations in a country was 5 and 10, respectively. Of 104 identified countries, 47 met the threshold.

The United States (documents=623, citations=24078, TLS=285) and the United Kingdom (documents=355, citations=23444, TLS=278) were the top two countries, as well as the only ones having TLS larger than 200. The ranking was then followed by Germany (documents=195, citations=7210, TLS=171), France (documents=159, citations=5291, TLS=141) and China (documents=583, citations=15633, TLS=140). Interestingly, despite being the second top countries in terms of their link strength, the United Kingdom's number of publications is considered smaller than China, showing how intense the United Kingdom researchers in collaborating with the other countries. Based on the number of link strengths, the United Kingdom and the United States had the highest number of collaborations (LS=49), followed by China and the United States (LS=37) and China and Australia (LS=21), and the United Kingdom and Germany (LS=21).

Most cited papers covering chemometrics and spectroscopy in pharmaceutical topics

The list of the most cited papers of chemometrics and spectroscopy studiesy in pharmaceutical research is presented in **Table 3**. 'Metabonomics: Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data' authored by Jeremy K. Nicholson and colleagues was the most cited article in chemometrics and spectroscopy in pharmaceutical research, cited by 3269 papers since its publication in 1999. The second most cited article—also only the other article to be cited more than 1000 times was a paper authored by Kodo Kawase and colleagues in 2003 titled 'Non-destructive terahertz imaging of illicit drugs using spectral fingerprints', cited by 1242 papers.

It is also worth noting that none of the top ten most cited papers of chemometrics and spectroscopy in pharmaceutical research was published in the last ten years (2013-2023), with the newest article published in 2010. Of the top ten most cited papers in this list, three focused on metabolomics and NMR spectroscopy research (cluster I) and three on multivariate analysis in polymorphism studies (cluster III), covering more than half of the list.

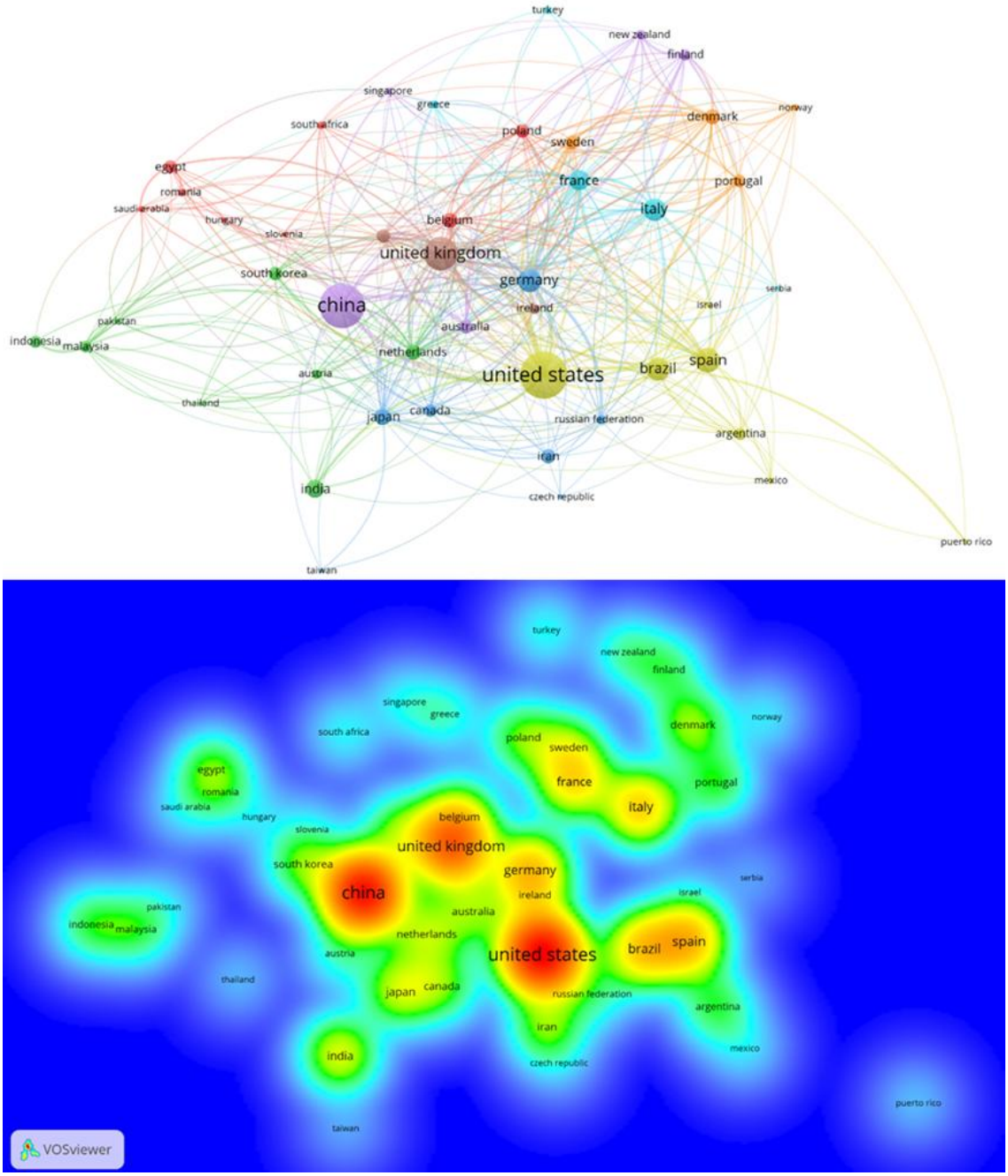

Figure 6. Network visualization of countries' co-authorship (top) and Density visualization of countries' co-authorship (bottom).

Discussion

The 50-year progress

The research about the combination of chemometric and spectroscopy has been established for almost half a century (50 years), with the number of annual publications tend to remain constant after 2017 (<250 publications/year). Throughout those years, a high proportion of original research articles have been published (88.71%), with only 6.55% review articles. This finding indicates that this research field has matured and contrasts with other research themes such as COVID-19 [12] and pollutant degrading enzymes – laccase [26]. The top-cited paper by Nicholson and colleagues has reached 3269 citations, which is relatively higher than others [26]. The inter-countries collaborations are mostly found in the United Kingdom, the United States, China, Australia, and Germany, which these countries are leaders in pharmaceutical industries [27]. Based on the clusters of co-occurring keywords, the application of chemometric and spectroscopy

combinations in the pharmaceutical field could be generalized into two: (1) drug toxicity and efficacy monitoring and (2) quality control of drug manufacturing.

Table 3. Top 10 most cited papers in chemometrics and spectroscopy in pharmaceutical research

Rank	Title	First author	Year of publish	Citation	Ref
1	Metabonomics: Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data	Nicholson JK	1999	3269	[16]
2	Non-destructive terahertz imaging of illicit drugs using spectral fingerprints	Kawase K	2003	1242	[17]
3	A review of near-infrared spectroscopy and chemometrics in pharmaceutical technologies	Roggo Y	2007	932	[18]
4	Symbiotic gut microbes modulate human metabolic phenotypes	Li M	2008	866	[19]
5	Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications	Reich G	2005	701	[20]
6	NMR-based metabolomic analysis of plants	Kim HK	2010	678	[21]
7	3D-QSAR in drug design - A review	Verma J	2010	549	[22]
8	New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids	Sandberg M	1998	497	[23]
9	Liver, muscle, and adipose tissue insulin action is directly related to intrahepatic triglyceride content in obese subjects	Korenblat KM	2008	436	[24]
10	NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition	Bollard ME	2005	413	[25]

Pharmaceutical-related components analysis

Drug toxicity and efficacy monitoring

Chemometric approaches are advantageous when applied to assess the efficacy and side effects of drugs and treatment monitoringbecause their samples are biofluids which could be collected in less invasive procedures [28]. Mostly, the NMR – as the analytical instrument and blood and urine – as biofluid samples were used for this purpose [29]. Animal models (rat or mouse) were widely used in the reports because it was not ethical to induce toxicity in human subjects [30-32]. The combination technique has been suggested to detect multiple organ injuries based on urinary sample analysis [32-34]. Reports reporting chemometric and spectroscopy combinations for pharmaceutical-related samples have been summarized and presented in **Table 4**.

Table 4. Drug toxicity and efficacy monitoring using combinatorial techniques of chemometric and spectroscopy

Author, year (ref)	Chemometric(s) combined with H^1 NMR	Sample	Findings
Sussulini *et al.* 2009 [39]	PLS-DA	Blood serum of patients with bipolar disorder treatment	Distinguished lipids, lipid-metabolism-related molecules, and amino acids profiles related to lithium or other medications
Liu *et al.* 2011 [30]	PLS-DA	Plasma of rats treated with anti-depressant	Reduced chronic unpredictable mild stress-associated metabolites after Xiaoyaosan treatment
Holmes *et al.* 2000 [33]	PCA, SIMCA	Urine of rats treated with hydrazine or $HgCl_2$	Distinguished organ toxicity (liver vs. kidney) with 98% accuracy and 79% sensitivity

Author, year (ref)	Chemometric(s) combined with H^1 NMR	Sample	Findings
Waters *et al.* 2001 [40]	PCA	Urine and plasma of rats treated with α-naphthylisothiocyanate	Identification of key metabolites related to hepatic intoxication
Van Dorsten *et al.* 2006 [41]	PCA, PLS-DA	Urine and plasma of human subjects consuming green tea or black tea	Identification of green tea effects on oxidative energy metabolism-related metabolites
Winnike *et al.* 2010 [34]	PCA, OPLS-DA	Urine of healthy adults receiving acetaminophen	Early detection of liver injury (before ALT increase)
Wei *et al.* 2009 [31]	PCA, PLS-DA	Urine and serum of rats receiving realgar	Identification of metabolites related to realgar-induced oxidative damage
Lenz *et al.* 2004 [32]	PCA	Urine of rats receiving cyclosporin A	Identification of the onset of cyclosporin A-nephropathy
Van Doorn *et al.* 2007 [42]	PCA, DAPC	Urine and blood plasma of diabetic patients receiving thiazolidinediones	Monitoring of thiazolidinediones therapy based on related biomarkers
Daykin *et al.* 2005 [38]	PCA	Urine of healthy subjects receiving	Identification of black tea flavonoid metabolites related to the treatment efficacy monitoring
Huo *et al.* 2009 [43]	PCA, PLS	Serum of diabetic patients receiving metformin	Treatment efficacy monitoring was based on the changes of amino acids and glucose and non-glucose metabolites
Romick-Rosendale *et al.* 2009 [44]	PCA, PLS-DA	Urine and feces of mice receiving enrofloxacin	Detection of gut microbiome dysbiosis

ALT: alanine aminotransferase; PCA: principal component analysis; PLS-DA: partial least square discriminant analysis; PLS: partial least square regression analysis; DAPC: discriminate analysis of principal component; SERS: surface enhanced Raman spectroscopy; SLIPSERS: slippery liquid-infused porous surface-enhanced Raman spectroscopy; OPLS-DA: orthogonal projections to latent structures-discriminant analysis; MCR-ALS: multivariate curve resolution-alternating least square; SIMCA: soft independent modelling of class analogy

Since it can be used to analyze metabolomic profiles in biofluids, the combined techniques could also be used to detect disease development and progression with implications of therapeutic efficacies. The summary of the combinatorial techniques to observe disease development and progression has been presented in **Table 5**. The application ranged from detecting cancer to the most recent infectious disease – coronavirus 2019 (COVID-19) [35, 36]. As a complementary technique, imaging FTIR in combination with PCA and unsupervised hierarchical cluster analysis (UHCA) was used to detect lung cancer in animal tissue, giving a comprehensive outlook on histological images [37]. Sometimes the methods were combined with HPLC to provide more comprehensive metabolomic data [38].

Table 5. Disease development and progression monitoring using combinatorial techniques of chemometric and spectroscopy

Author, year (ref)	Spectroscopy	Chemometric	Sample	Findings
Jakubczyk *et al.* 2022 [45]	FTIR	PLSR, PCA, and machine learning	Follicular fluid	Idiopathic female infertility determination (~100% accuracy)
Bari *et al.* 2022 [46]	SERS	PCA, PLS-DA, PLSR	Centrifugally filtered serum	HBV-infected sample can be differentiated from that of non-infected samples
Carswell *et al.* 2022 [47]	Raman spectroscopy	PCA, DAPC, PLSR	Urine	Quantification of macro- and microhematuria (>90% accuracy)
Cai *et al.* 2022 [35]	SLIPSERS	PCA, PLS-DA, OPLS-DA	Blood	Detection of lung cancer efficiently (~100% accuracy, small samples and rapid process)
Robertson *et al.* 2022 [36]	Raman spectroscopy	PCA, DAPC	Urine	Identification of COVID-19 associated diseases, COVID-19

Author, year (ref)	Spectroscopy	Chemometric	Sample	Findings
				severity, and long COVID-19 (>90% accuracy)
Senger *et al.* 2022 [48]	Raman spectroscopy	PCA, DAPC	Urine	Detection of lime disease (compared with healthy subjects) with 88.7% accuracy, 83.3% sensitivity, and 91.0% specificity
Caixeta *et al.* 2023 [49]	ATR-FTIR	PCA, MCR-ALS	Urine	More accurate estimation of glucose concentration compared with enzyme assays
Koehler *et al.* 2022 [50]	ATR-FTIR	PCA, OPLS-DA	Serum	Correct detection of paracoccidioidomycosis in almost all samples (100% sensitivity and specificity)
Ortega-Hernández *et al.* 2022 [51]	FTIR+ nano-immunosensor	PCA	Urine	Discrimination of acute kidney injury from healthy patients

PCA, principal component analysis; PLS-DA, partial least square discriminant analysis; PLS, partial least square regression analysis; DAPC, discriminate analysis of principal component; SERS, surface-enhanced Raman spectroscopy; SLIPSERS, slippery liquid-infused porous surface-enhanced Raman spectroscopy; OPLS-DA, orthogonal projections to latent structures-discriminant analysis; MCR-ALS, multivariate curve resolution-alternating least square

Quality control of drug manufacturing

Summaries of reports on the use of spectroscopy in combination with chemometrics for drug quality control have been presented in **Table 6**. Studies on the utilization of this combinatorial technique have been reported more recently, suggesting the shifted trend in this research field. Infrared spectroscopy and its modified versions are mostly used for this purpose due to their simplicity and rapid analysis [52-56]. While PCA is the common chemometric discriminant method, PLSR has been used widely to quantify certain compositions of active pharmaceutical ingredients in drug formulations, including monoclonal antibodies (mAbs) and chemotherapeutical formulations [55,57]. This combinatorial technique is particularly popular among traditional medicine manufacturers [52,56,58-61]. Not only to ensure the compositions among the formulations remain the same, a combination of spectroscopy and chemometric could be used to differentiate polymorphism in drugs such as triclabendazole [62].

Table 6. Drug manufacturing quality control using combinatorial techniques of chemometric and spectroscopy

Author, year (ref)	Spectroscopy	Chemometric	Sample	Findings
Geskovski *et al.* 2021 [54]	MIR	PLS	Extracts of cannabis flower	Quantification of tetrahydrocannabinol and cannabidiol
Sakira *et al.* 2021 [53]	NIR	PCA, PLSR	Metronidazole tablet	Low-biased and accurate QC of metronidazole formulations
Gong *et al.* 2021 [52]	UV, FTIR	HCA, PLS	Liquorice tablet	Accurate identification and quantification of liquiritin, glycyrrhizinic acid, and sodium benzoate
Ren *et al.* 2021 [58]	UV, HPLC	PCA, DR, LDA	Curcuma-based herbal medicine	Prediction of bioactive contents in *Curcumae rhizome* from different cultivation places
Wang *et al.* 2021 [59]	UPLC, FTIR	QAMS, AMSQFM	Yankening tablets	Accurate and low-biased quantification of berberine and baicalin
Makki *et al.* 2021 [57]	Raman, FTIR	PLSR	Drugs containing anticancer agents	Accurate and low-biased QC for commercial intravenous chemotherapeutic formulations (Raman>FTIR)
Yao *et al.* 2022 [56]	FTNIR, GC-MS, UHPLC-	PCA, OPLS-DA	Wild and cultivated agarwood	Identification of metabolomic variation

Author, year (ref)	Spectroscopy	Chemometric	Sample	Findings
	Q-Exactive Orbitrap/MS			between wild and cultivated agarwood
Makki *et al.* [55]	Raman	PLSR, PLS-DA	Commercial formulations containing cetuximab, rituximab, trastuzumab, bevacizumab	Validation of mAbs conformity
Yu *et al.* 2022 [61]	HPLC-UV-ELSD	PLSR, CCA	Bushen Huoxue Prescription	Quantification of puerarin, daidzin, salvianolic acid B and ginsenoside Rb1 as active agents
Wang *et al.* [60]	ATR-FTIR, HS-GC-MS	PCA, HCA, BP-NN, KNN, LDA	Dried tuberous roots of Curcumae Radix	Discrimination and characterization of volatile contents from Curcumae Radix (100% discriminant accuracy)
Salazar-Rojas, *et al.* 2021 [62]	NIR	PLS	Triclabendazole	Faster polymorphic control (60%)
Rabiere *et al.* 2022 [63]	HPLC, GC-MS, XPRD, ^{1}H NMR, NIR	PCA, HCA	Omeprazole	Discrimination of omeprazole and omeprazole magnesium based on manufacturing source

AMSQFM: average method of systematic quantified fingerprint method; ATR-FTIR: attenuated total reflection fourier transformed infrared; BP-NN: back propagation neural network; CCA: canonical correlation analysis; DR: decision tree; FTIR: Fourier transform-infrared; FT-NIR:Fourier transform near-infrared; GC–MS: gas chromatography-mass spectrometry; HCA: hierarchical cluster analysis; HPLC, high-performance liquid chromatography; HPLC-UV-ELSD; high-performance liquid chromatography-ultraviolet detector-evaporative light scattering detector; KNN: K-nearest neighbor; LDA: linear discriminant analysis; MIR: mid-infrared; NIR: near-infrared; QAMS: quantitative analysis of multi-components by single marker; QC: quality control; UHPLC-Q-Exactive Orbitrap/MS: ultraperformance liquid chromatography Quadrupole-Exactive Orbitrap tandem mass spectrometry; XPRD: X-ray powder diffractometry

Future direction

Most of the research on the toxicology of drugs was carried out in 2010 and before. Surprisingly, the studies used numerous animal subjects, which might be problematic based on the current ethical standards. For example, a study employed over 400 rats induced with hepatotoxin hydrazine or nephrotoxins $HgCl_2$ [33]. A much fewer number of samples (n=25) were used in the hepatotoxicant model [40], but it remains concerning according to the present ethical standards. It has been suggested that the minimum number of animals in a study is 3, but the maximum should be 6 or 7 [64]. In the future, the development of in vitro models mimicking humane physiological systems or in vivo model with less ethical concerns (such as drosophila) could assist this study [65]. Moreover, our extensive knowledge could make direct intervention human subjects possible.

Future trajectories of chemometric research also include the development of mathematical models assisted by artificial intelligence (AI). Supervised models built by machine learning algorithms have been reported [59, 60]. Advanced modeling could also aid the accuracy and sensitivity of this technique in analyzing complex drug formulations such as those in traditional medicine [59-61]. With data that have been gathered for over 50 years of this research, the study will require less sample in the future, hence overcoming the ethical problems. Further, these methods could be standardized after the system is built on large data collected previously. This allows the system to overcome several biases in diagnosis or disease monitoring, including race/ethnicity, age, sex, unknown clinical variables, limited knowledge, socioeconomic status, and other characteristics [66]. This system indicates that the employment of chemometric and spectroscopy as a combinatorial analytical technique could provide objective assessment on drug efficacy, which contributes significantly to the clinical research in the future [67].

Conclusions

The 50-year research trend of analytical technique spectroscopy combined with chemometrics in pharmaceutical fields has been successfully observed using bibliometric information and network visualization. The research of spectroscopy and chemometrics in the pharmaceutical field could be divided into two categories: (1) drug toxicity and efficacy monitoring and (2) quality control of drug manufacturing. The research using this technique to measure the toxicology and efficacy of drugs through metabolomic profiling has reached a maturity stage and is now shifting to its utility for quality control of drug formulations, especially the traditional medicine. The accuracy of this technique could reach 100% in some studies, where the analysis could be run in a much shorter time as compared to the conventional method. Moreover, the techniques offer a minimum invasive approach. With the development of AI, the chemometric-empowered spectroscopy technique potentially has a significant role in the development of therapeutics.

Ethics approval

Not required.

Acknowledgments

The authors would like to appreciate the collaboration between Universitas Syiah Kuala, Banda Aceh, Indonesia; Admiralty University of Nigeria, Ibusa, Delta State, Nigeria; Ring Road State Hospital, Ibadan, Oyo State, Nigeria; and Innovative Sustainability Lab, PT. Biham Riset dan Edukasi, Banda Aceh, Indonesia who have supported the research experiment and the making of this article, as well as the anonymous reviewer for providing technical and substantial insights in improving the manuscript.

Conflict of interest

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

Underlying data for the bibliometric analysis can be made available by following the reasonable request to the corresponding author.

How to cite

Zulkifli B, Fakri F, Odigie J, *et al.* Chemometric-empowered spectroscopic techniques in pharmaceutical fields: A bibliometric analysis and updated review. Narra X 2023; 1 (1): e80 - http://doi.org/10.52225/narrax.v1i1.80.

References

1. Mennen SM, Alhambra C, Allen CL, *et al.* The evolution of high-throughput experimentation in pharmaceutical development and perspectives on the future. Org Process Res Dev 2019;23(6):1213-1242.
2. N. Politis S, Colombo P, Colombo G, *et al.* Design of experiments (DoE) in pharmaceutical development. Drug Dev Ind Pharm 2017;43(6):889-901.
3. Chow S-C. Pharmaceutical validation and process controls in drug development. Drug Inf J 1997;31(4):1195-1201.
4. Milton EJ, Schaepman ME, Anderson K, *et al.* Progress in field spectroscopy. Remote Sens Environ 2009;113:S92-S109.
5. Wold S. Chemometrics; What do we mean with it, and what do we want from it? Chemom Intell Lab Syst 1995;30(1):109-115.
6. Zhu M-Z, Wen B, Wu H, *et al.* The quality control of tea by near-infrared reflectance (NIR) spectroscopy and chemometrics. J Spectrosc 2019;2019.
7. Ralbovsky NM and Lednev IK. Raman spectroscopy and chemometrics: A potential universal method for diagnosing cancer. Spectrochim Acta A Mol Biomol Spectrosc 2019;219:463-487.

8. Pacholczyk-Sienicka B, Ciepielowski G, Albrecht Ł. The application of NMR spectroscopy and chemometrics in authentication of spices. Molecules 2021;26(2):382.
9. Lomont JP, Smith JP. In situ Raman spectroscopy for real time detection of cysteine. Spectrochim Acta A Mol Biomol Spectrosc 2022;274:121068.
10. Nagy B, Farkas A, Borbás E, *et al.* Raman spectroscopy for process analytical technologies of pharmaceutical secondary manufacturing. AAPS PharmSciTech 2019;20(1):1.
11. Cozzolino D, Cynkar W, Shah N, *et al.* Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. Food Res Int 2011;44(7):1888-1896.
12. Chiari W, Damayanti R, Harapan H, *et al.* Trend of polymer research related to COVID-19 pandemic: Bibliometric analysis. Polymers 2022;14(16):3297.
13. Kumar M, George RJ, PS A. Bibliometric analysis for medical research. Indian J Psychol Med 2022;XX:1-6.
14. Ahmad T, Dhama K, Tiwari R, *et al.* Bibliometric analysis of the top 100 most cited studies in apolipoprotein E (ApoE) research. Narra J 2021;1(1):e2.
15. van Eck NJ and Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010;84(2):523-538.
16. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenibiotica 1999;29(11):1181-1189.
17. Ogawa Y, Kawase K, Yamashita M, *et al.* Non-destructive inspection techniques for illicit drugs using terahertz imaging. CLEO 2004:1;3-12.
18. Roggo Y, Chalus P, Maurer L, *et al.* A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J Pharm Biomed Anal 2007;44(3):683-700.
19. Li M, Wang B, Zhang M, *et al.* Symbiotic gut microbes modulate human metabolic phenotypes. Proc Natl Acad Sci 2008;105(6):2117-2122.
20. Reich G. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. Adv Drug Deliv Rev 2005;57(8):1109-1143.
21. Kim HK, Choi YH, Verpoorte R. NMR-based metabolomic analysis of plants. Nat Protoc 2010;5(3):536-549.
22. Verma J, Khedkar MV, Coutinho CE. 3D-QSAR in drug design-A review. Current Curr Top Med Chem 2010;10(1):95-115.
23. Sandberg M, Eriksson L, Jonsson J, *et al.* New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem 1998;41(14):2481-2491.
24. Korenblat KM, Fabbrini E, Mohammed BS, *et al.* Liver, muscle, and adipose tissue insulin action is directly related to intrahepatic triglyceride content in obese subjects. Gastroenterol 2008;134(5):1369-1375.
25. Bollard ME, Stanley EG, Lindon JC, *et al.* NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR Biomed 2005;18(3):143-162.
26. Puspita K, Chiari W, Abdulmadjid SN, *et al.* Four decades of laccase research for wastewater treatment: Insights from bibliometric analysis. IJERPH 2022;20(1):308.
27. González POI, López ZZMÁ and Cabral RH. Pharmaceuticals market, consumption trends and disease incidence are not driving the pharmaceutical research on water and wastewater. Intl J Environ Res Public Health 2021;18(5):2532.
28. Lindon JC, Holmes E, Bollard ME, *et al.* Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. Biomark 2004;9(1):1-31.
29. Lindon JC, Nicholson JK, Holmes E, *et al.* Contemporary issues in toxicology the role of metabonomics in toxicology and its evaluation by the COMET project. Toxicol Appl Pharmacol 2003;187(3):137-146.
30. Liu X-J, Zhou Y-Z, Li Z-F, *et al.* Anti-depressant effects of Xiaoyaosan on rat model of chronic unpredictable mild stress: A plasma metabonomics study based on NMR spectroscopy. J Pharm Pharmacol 2011;64(4):578-588.
31. Wei L, Liao P, Wu H, *et al.* Metabolic profiling studies on the toxicological effects of realgar in rats by (1)H NMR spectroscopy. Toxicol Appl Pharmacol 2009;234(3):314-325.
32. Lenz EM, Bright J, Knight R, *et al.* Cyclosporin A-induced changes in endogenous metabolites in rat urine: A metabonomic investigation using high field 1H NMR spectroscopy, HPLC-TOF/MS and chemometrics. J Pharm Biomed Anal 2004;35(3):599-608.
33. Holmes E, Nicholls AW, Lindon JC, *et al.* Chemometric models for toxicity classification based on NMR spectra of biofluids. Chem. Res. Toxicol 2000;13(6):471-478.

34. Winnike J, Li Z, Wright F, *et al*. Use of pharmaco-metabonomics for early prediction of acetaminophen-induced hepatotoxicity in humans. Clin. Pharm. Therap 2010;88(1):45-51.
35. Cai C, Liu Y, Li J, *et al*. Serum fingerprinting by slippery liquid-infused porous SERS for non-invasive lung cancer detection. Analyst 2022;147(20):4426-4432.
36. Robertson JL, Senger RS, Talty J, *et al*. Alterations in the molecular composition of COVID-19 patient urine, detected using Raman spectroscopic/computational analysis. PLoS One 2022;17(7):e0270914.
37. Augustyniak K, Chrabaszcz K, Smeda M, *et al*. High-resolution fourier transform infrared (FT-IR) spectroscopic imaging for detection of lung structures and cancer-related abnormalities in a murine model. Appl Spectrosc 2022;76(4):439-450.
38. Daykin CA, Duynhoven JPV, Groenewegen A, *et al*. Nuclear magnetic resonance spectroscopic based studies of the metabolism of black tea polyphenols in humans. J Agric Food Chem 2005;53(5):1428-1434.
39. Sussulini A, Prando A, Maretto DA, *et al*. Metabolic profiling of human blood serum from treated patients with bipolar disorder employing 1H NMR spectroscopy and chemometrics. Anal Chem 2009;81(23):9755-9763.
40. Waters NJ, Holmes E, Williams A, *et al*. NMR and pattern recognition studies on the time-related metabolic effects of α-naphthylisothiocyanate on liver, urine, and plasma in the rat: An integrative metabonomic approach. Chem Res Toxicol 2001;14(10):1401-1412.
41. Van Dorsten FA, Daykin CA, Mulder TP, *et al*. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. J Agric Food Chem 2006;54(18):6929-6938.
42. Van Doorn M, Vogels J, Tas A, *et al*. Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in Type 2 diabetes mellitus patients and healthy volunteers. Br J Clin Pharmacol 2007;63(5):562-574.
43. Huo T, Cai S, Lu X, *et al*. Metabonomic study of biochemical changes in the serum of type 2 diabetes mellitus patients after the treatment of metformin hydrochloride. J Pharm Biomed Anal 2009;49(4):976-982.
44. Romick-Rosendale LE, Goodpaster AM, Hanwright PJ, *et al*. NMR-based metabonomics analysis of mouse urine and fecal extracts following oral treatment with the broad-spectrum antibiotic enrofloxacin (Baytril). Magn Reson Chem 2009;47(S1):S36-S46.
45. Jakubczyk P, Paja W, Pancerz K, *et al*. Determination of idiopathic female infertility from infrared spectra of follicle fluid combined with gonadotrophin levels, multivariate analysis and machine learning methods. Photodiagnosis Photodyn Ther 2022;38:102883.
46. Bari RZA, Nawaz H, Majeed MI, *et al*. Surface-enhanced Raman spectroscopic analysis of centrifugally filtered HBV serum samples. Photodiagnosis Photodyn Ther 2022;38:102808.
47. Carswell W, Robertson JL, Senger RS. Raman spectroscopic detection and quantification of macro- and microhematuria in human urine. Appl Spectrosc 2022;76(3):273-283.
48. Senger RS, Issa AS, Agnor B, *et al*. Disease-associated multimolecular signature in the urine of patients with Lyme disease detected using Raman spectroscopy and chemometrics. Appl Spectrosc 2022;76(3):284-299.
49. Caixeta DC, Lima C, Xu Y, *et al*. Monitoring glucose levels in urine using FTIR spectroscopy combined with univariate and multivariate statistical methods. Spectrochim Acta A Mol Biomol Spectrosc 2023;290:122259.
50. Koehler A, Scroferneker ML, Pereira BAS, *et al*. Using infrared spectroscopy of serum and chemometrics for diagnosis of paracoccidioidomycosis. J Pharm Biomed Anal 2022;221:115021.
51. Ortega-Hernández N, Ortega-Romero M, Medeiros-Domingo M, *et al*. Detection of biomarkers associated with acute kidney injury by a gold nanoparticle based colloidal nano-immunosensor by fourier-transform infrared spectroscopy with principal component analysis. Anal Lett 2022;55(15):2370-2381.
52. Gong D, Chen J, Li X, *et al*. A smart spectral analysis strategy-based UV and FT-IR spectroscopy fingerprint: Application to quality evaluation of compound liquorice tablets. J Pharm Biomed Anal 2021;202:114172.
53. Sakira AK, Mees C, Braekeleer KD, *et al*. Determination of the quality of metronidazole formulations by near-infrared spectrophotometric analysis. Talanta 2021;3:100027.
54. Geskovski N, Stefkov G, Gigopulu O, *et al*. Mid-infrared spectroscopy as process analytical technology tool for estimation of THC and CBD content in Cannabis flowers and extracts. Spectrochim Acta A Mol Biomol Spectrosc 2021;251:119422.
55. Makki AA, Massot V, Byrne HJ, *et al*. Understanding the discrimination and quantification of monoclonal antibodies preparations using Raman spectroscopy. J Pharm Biomed Anal 2021;194:113734.
56. Yao C, Qi L, Zhong F, *et al*. An integrated chemical characterization based on FT-NIR, GC–MS and LC-MS for the comparative metabolite profiling of wild and cultivated agarwood. J Chromatogr B Biomed Appl 2022;1188:123056.

57. Makki AA, Massot V, Byrne HJ, *et al*. Vibrational spectroscopy for discrimination and quantification of clinical chemotherapeutic preparations. Vib Spectrosc 2021;113:103200.
58. Ren X, He T, Wang J, *et al*. UV spectroscopy and HPLC combined with chemometrics for rapid discrimination and quantification of Curcumae Rhizoma from three botanical origins. J Pharm Biomed Anal 2021;202:114145.
59. Wang X, Liu X, Wang J, *et al*. Study on multiple fingerprint profiles control and quantitative analysis of multi-components by single marker method combined with chemometrics based on Yankening tablets. Spectrochim Acta A Mol Biomol Spectrosc 2021;253:119554.
60. Wang L, Li X, Wang Y, *et al*. Rapid discrimination and screening of volatile markers for varietal recognition of Curcumae Radix using ATR-FTIR and HS-GC-MS combined with chemometrics. J Ethnopharmacol 2021;280:114422.
61. Yu Y, Zhu Z, Xie M, *et al*. Investigation on the Q-markers of Bushen Huoxue prescriptions for DR treatment based on chemometric methods and spectrum-effect relationship. J Ethnopharmacol 2022;285:114800.
62. Salazar-Rojas D, Kaufman TS, Maggio RM. A comprehensive approach toward concomitant triclabendazole polymorphism in pharmaceutical products. J Drug Deliv Sci Technol 2021;62:102386.
63. Rebiere H, Grange Y, Deconinck E, *et al*. European fingerprint study on omeprazole drug substances using a multi analytical approach and chemometrics as a tool for the discrimination of manufacturing sources. J Pharm Biomed Anal 2022;208:114444.
64. Fitts DA. Ethics and animal numbers: Informal analyses, uncertain sample sizes, inefficient replications, and type I errors. J Am Assoc Lab Anim Sci 2011;50(4):445-453.
65. Yamaguchi M and Yamamoto S. Role of drosophila in human disease research 2.0. Int J Mol Sci 2022;23(8):4203.
66. Garb HN. Race bias and gender bias in the diagnosis of psychological disorders. Clin Psychol Rev 2021;90:102087.
67. Park Y, Jackson GP, Foreman MA, *et al*. Evaluating artificial intelligence in medicine: phases of clinical research. JAMIA Open 2020;3(3):326-331.