IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 6, December 2025, pp. 4403~4414
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i6.pp4403-4414

A bibliometric analysis of feature selection techniques: trends, innovations, and future directions

Oumaima Semmar, Wissal El Habti, Donalson Wilson, Abdellah Azmani
Laboratory of Intelligent Automation and BioMedGenomics, Faculty of Sciences and Technologies of Tangier, Abdelmalek Essaâdi University, Tetouan, Morocco

Article Info

Article history:

Received Nov 30, 2024
Revised Sep 19, 2025
Accepted Oct 16, 2025

Keywords:

Artificial intelligence
Bibliometric analysis
Feature selection
High dimensionality
Machine learning

ABSTRACT

Feature selection techniques have become increasingly important in addressing the challenges of high dimensionality in machine learning and other artificial intelligence domains. In this study, we present a comprehensive bibliometric analysis of research on feature selection techniques over the past decade, focusing on mapping the intellectual structure, identifying emerging trends, and highlighting productive collaborations in the field. Using merged data from Scopus and Web of Science databases, we collected and analyzed 2,079 relevant documents published between 2014 and 2024, applying citation analysis, co-authorship networks, and keyword co-occurrence mapping. Our findings reveal that feature selection methodologies including supervised, unsupervised, and hybrid approaches across filter, wrapper, and embedded techniques, have been widely applied across various domains. The authors who have most contributed to the development of these methods are primarily affiliated with institutions in China, India, and the USA. The insights provided by this analysis offer researchers and practitioners a valuable foundation for guiding future research directions in feature selection.

Corresponding Author:

Oumaima Semmar
Laboratory of Intelligent Automation and BioMedGenomics
Faculty of Sciences and Technologies of Tangier, Abdelmalek Essaâdi University
Km. 10 Old Airport Road, PO Box: 416, Tangier, Morocco
Email: oumaima.semmar1@etu.uae.ac.ma

1. INTRODUCTION

In the era of big data, the explosion of multidimensional data has heightened the need for efficient machine learning algorithms while exacerbating the "curse of dimensionality" [1]. This phenomenon, where adding features leads to data dispersion, complicates model generalization and increases the risk of overfitting [2]. To address these challenges, feature selection has become a crucial strategy in artificial intelligence and machine learning, aiming to identify the most relevant variables to improve model efficiency [3], thereby establishing itself as a prominent research field [4]. The field encompasses various methodologies, including supervised, unsupervised, and hybrid approaches, with selection strategies categorized into wrapper, filter, and embedded methods, each offering specific advantages and limitations [3], [5]. Recently, significant progress has been made in integrating deep learning techniques for feature selection, promising better management of complex and large-scale data [6], [7]. Moreover, the emergence of adaptive methods capable of dynamically adjusting feature selection based on evolving data opens up new perspectives for more robust and flexible models [8].

The versatility of feature selection is demonstrated across diverse domains. In healthcare, particularly oncology, these techniques have revolutionized cancer detection and diagnostic precision [9], [10].

The financial sector harnesses them for risk assessment and insurance optimization [11], while manufacturing industries benefit from their application in predictive maintenance [12]. The technology sector has integrated these methods into intrusion detection systems and natural language processing applications [13], [14], and bioinformatics leverages them for gene expression analysis and protein structure prediction [15]. Additionally, transportation systems have achieved significant improvements in traffic optimization and travel time prediction through these techniques [16].

The exponential growth of scientific literature has necessitated the employment of systematic methods for analyzing research developments. Bibliometric analysis, encompassing citation analysis, co-authorship networks, and keyword co-occurrence patterns, provide a robust framework for evaluating scientific output and research evolution [17]. These analytical approaches, enhanced by recent developments in bibliometric tools, enable sophisticated examination of large-scale scientific data, offering comprehensive insights into research dynamics [18]. While existing studies have provided valuable insights into specific aspects of feature selection, such as optimization algorithms in healthcare applications [19] and microarray research trends [20], there remains a critical need for a comprehensive understanding of this rapidly evolving field.

Our study addresses this research gap through a comprehensive bibliometric analysis of feature selection literature spanning the past decade. This analysis aims to map the field's intellectual structure, identify emerging research frontiers, and provide evidence-based insights to guide future investigations. The paper is structured as follows: First, we present a detailed description of our bibliometric methodology, encompassing data collection processes and analytical tools. Subsequently, we present our findings, followed by an in-depth discussion of the results. The paper concludes with a synthesis of key insights, while acknowledging methodological limitations and proposing future research directions in this domain.

2. METHOD

Despite the variety of literature databases available, such as Dimensions [21], Lens, PubMed [22], and Cochrane Library. For this study, Scopus and Web of Science were chosen as they are the most comprehensive and commonly referenced multidisciplinary databases [23]. Additionally, based on suggestions [24] and recommendations [23], these two selected databases complement each other, especially when comparing various fields, institutions, countries, or languages. Therefore, this study will focus on merging the two main databases: Scopus and Web of Science.

The bibliometric data collection process took place on September 25, 2024. All searches were conducted on the same day to prevent bias from daily database updates. The research study aimed to identify keywords present in the articles' titles, keywords, or abstracts. The following topical query was performed: ("Feature selection" OR "Dimensionality reduction" OR "Attribute selection") AND ("Supervised learning" OR "Unsupervised learning " OR “Machine learning” OR “Deep learning”) AND ("Statistical methods" OR "PCA" OR “Principal component analysis” OR "Correlation-based" OR "Chi-square" OR "Mutual information" OR "ANOVA" OR "Variance threshold"). This resulted in 2088 documents in Scopus and 1,328 documents in Web of Science. Figure 1 shows the identification of studies via databases.

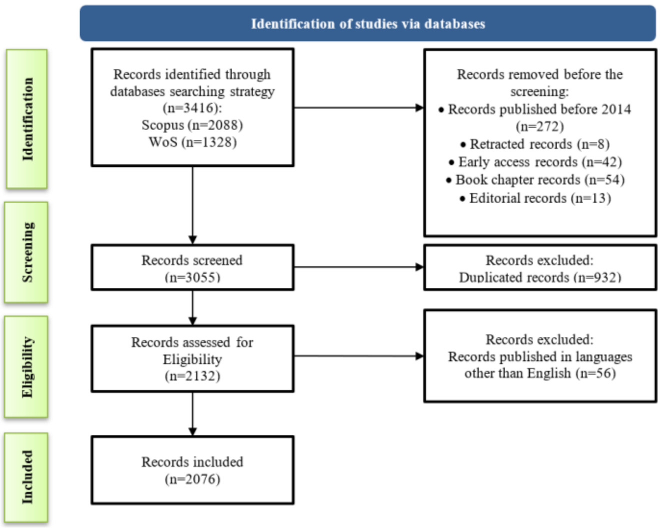

Figure 1. Identification of studies via databases

The PRISMA flowchart [25], as depicted in Figure 1, was utilized to select only 2,079 eligible papers. Initially, all document types: articles, proceeding papers, and reviews were included. Then, records published outside the period of 2014-2024 were removed. The latest version of R (4.4.1) [26] was used, along with the bibliometric package, to merge, screen, analyze, and map the downloaded bibliographic data [27]. A ".xlsx" Excel file was uploaded to the Biblioshiny, a web-based graphical user interface included in bibliometric [28]. Data analysis was supported by downloading Excel (.csv) and image (.png) files, which were processed in line with the study's objectives.

3. RESULTS AND DISCUSSION

In this section, we present the findings of our bibliometric analysis, utilizing two distinct approaches. First, we conducted a descriptive analysis to gather statistical information. Next, we employed scientific mapping to explore the conceptual structure, focusing on key terms and the social network among authors who have contributed to the development of feature selection techniques. This section presents detailed statistics on yearly research output, average citations per year, and identifies the most productive authors, institutions, countries, publications, and keywords. It also sheds light on the most frequently cited sources and authors.

3.1. Descriptive analysis

3.1.1. Key retrieved information

Table 1 provides an overview of the main bibliometric indicators associated with feature selection techniques across various domains of artificial intelligence. The table summarizes information such as document types, publication sources, authorship statistics, keyword frequencies, and patterns of research collaboration. According to the data, 2,079 documents have been published on this topic, the majority being journal articles (1,365) and conference papers (539). Notably, the annual growth rate is negative (-27.8%), suggesting a decline in scientific output on feature selection techniques in recent years. This trend will be examined more closely in the following sections to better understand its implications.

3.1.2. Annual scientific production over the years

Figure 2 illustrates the annual scientific production over the years. Starting from 2014, a steady increase in output is observed until 2020, followed by a moderate rise through 2021 and 2022, reaching a peak in 2023. This sustained growth may be attributed to the increasing complexity of feature selection challenges in machine learning and other artificial intelligence applications, as well as to the dedicated efforts of researchers. However, this upward trend is followed by a noticeable decline in 2024, as also noted in Table 1.

Table 1. Main statistical information

Category	Description	Results
Main information about data	Timespan	2014:2025
	Sources (journals and books)	1,091
	Documents	2,079
	Annual growth rate %	-27.8
	Document average age	2.61
	Average citations per doc	16.32
	References	16328
Document contents	Keywords plus (ID)	11144
	Author's keywords (DE)	5055
Authors	Authors	6,122
	Authors of single-authored docs	65
Authors collaboration	Single-authored docs	110
	Co-Authors per doc	4.18
	International co-authorships %	4.521
Document types	Article	1,365
	Article article	9
	Article conference paper	4
	Article review	1
	Article; early access	7
	Article; proceedings paper	3
	Article; retracted publication	4
	Conference paper	539
	Conference paper article	2
	Conference review	40
	Conference review article	1
	Proceeding's paper	34
	Review	69
	Review article	1

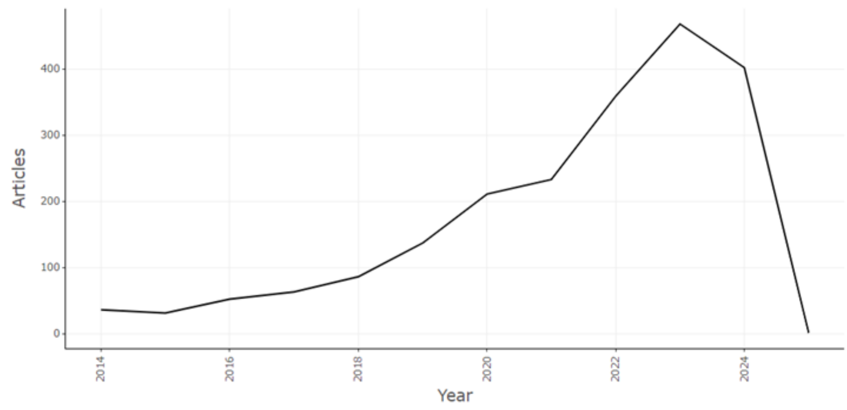

Figure 2. Annual scientific production over the years

3.1.3. Most relevant sources and affiliations

The analysis of publication sources and author affiliations offers insight into the key venues and institutions contributing to feature selection research. Figure 3 displays the ten most relevant sources, with IEEE Access leading the list with 60 published documents. Table 2 presents the most productive institutions, revealing a concentration of research efforts in Asia, particularly China and Saudi Arabia. These findings reflect the geographical and institutional dynamics shaping current research trends in the field.

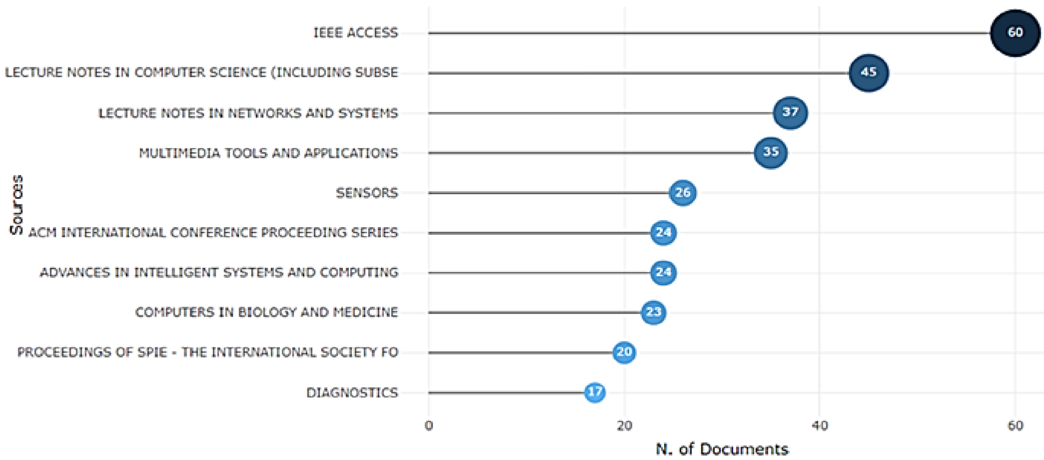

Figure 3. Most relevant sources

Table 2. Affiliations

Affiliation	Articles
Northwestern Polytechnical University	25
King Saud University	24
University of California	20
Comsats University Islamabad	18
Princess Nourah Bint Abdulrahman University	18
Shenzhen University	17
Guangdong University of Technology	15
Najran University	15
Xinjiang University	14

3.1.4. Statistics by country

Tables 3 and 4 summarize the scientific output by country in the field of feature selection techniques. China, India, and the USA lead in publication volume, with 696, 514, and 330 documents, respectively. At the regional level, Asia accounts for the majority of contributions (1,437 publications), reflecting the continent's growing investment in artificial intelligence research. The Americas, mainly represented by the USA and Canada, contribute a total of 844 documents. These figures highlight Asia's central role in driving global research output, while also illustrating the strong presence of North America in the field.

Table 3. Countries' productivity

Region	Frequency
China	696
India	514
USA	330
Canada	95
Saudi Arabia	95
UK	84
Iran	70
Australia	62
South Korea	62
Spain	55

Table 4. Most cited countries

Country	Total citations	Average article citations
China	8751	21.10
USA	3632	24.70
India	3252	9.90
United Kingdom	1801	40.90
Korea	1495	28.80
Saudi Arabia	992	18.00
Italy	876	38.10
Iran	824	24.20
Spain	721	34.30
Turkey	608	21.00

Figure 4 presents the research productivity trends of five countries—Canada, China, India, Saudi Arabia, and the USA. The results show a steady rise in publication output for China, India, and the USA between 2014 and 2024. As the study concluded in September 2024, data for the final months of the year are not included, leaving the continuation of this trend uncertain. By contrast, Saudi Arabia and Canada display only moderate growth in their research output over the same period.

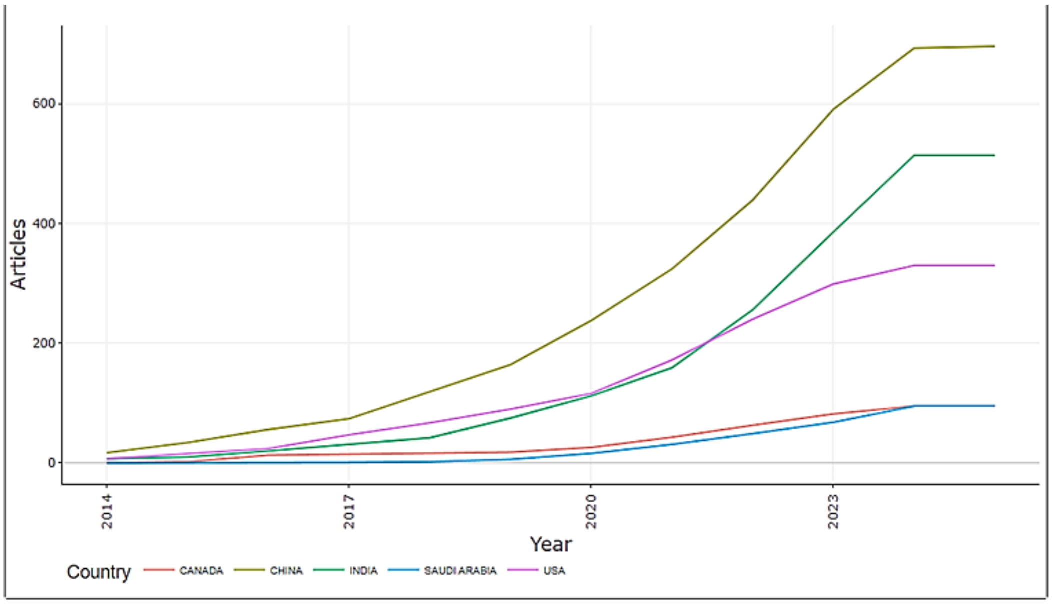

Figure 4. Countries' production over time

3.1.5. Author productivity through Lotka's law

Lotka's law was applied to analyze author productivity. Proposed by Alfred James Lotka in 1926 [29], this statistical principle describes how scientific output is distributed among authors: a small proportion accounts for the majority of publications, while most authors contribute only a few works. This law can be mathematically formulated as in (1).

$$N(x) = \frac{N(1)}{x^n} \quad (1)$$

Where N(x) represent the number of authors who have published a total of x papers, N(1) number of authors who have published just one paper and n is typically around 2. However subsequent research has shown that n can vary depending on the field of discipline. Figure 5 illustrates that, in the field of feature selection, most authors have contributed only one or two publications, while very few have published extensively. The observed distribution follows Lotka's law, indicating that a minority of authors contribute a disproportionately large number of studies.

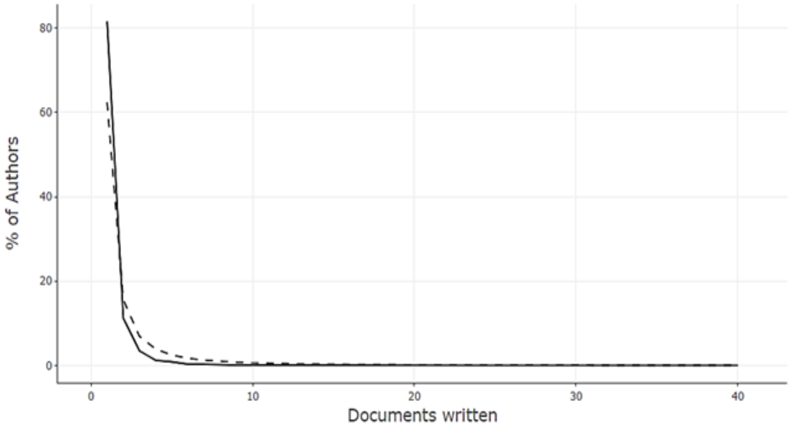

Figure 5. Distribution of author productivity based on Lotka's law

3.1.6. Statistics by documents

Among the 2,079 documents related to feature selection techniques, this subsection highlights the ten most globally cited papers, as shown in Table 5. These publications span diverse domains such as remote sensing, image processing, bioinformatics, and medical diagnostics, reflecting the interdisciplinary nature of feature selection research. Chen *et al.* [30] discussed deep learning-based classification for hyperspectral data, using a hybrid approach to extract high-level features, and has received 1,949 global citations, illustrating its long-standing influence. Babenko *et al.* [31] explored neural codes for image retrieval, employing principal component analysis for feature extraction, and has accumulated 950 citations. Although most of these papers are not among the most recent, their consistently high citation rates underscore their foundational role and ongoing relevance in the development of feature selection methodologies.

Figure 6 presents the keywords most commonly appearing in publications on feature selection techniques. "deep learning" is the most common keyword, appearing 11% of the time, suggesting that many of these techniques are applied in deep learning contexts. This is followed by "principal component analysis", "dimensionality reduction", "learning system" and "feature extraction".

Table 5. Top ten global cited documents in feature selection

Paper	Total citations	Total citations per year
Chen *et al.* [30]	1949	177.18
Babenko *et al.* [31]	950	86.36
Chen *et al.* [32]	934	93.40
Wang *et al.* [33]	663	73.67
Gromski *et al.* [34]	583	58.30
Nanni *et al.* [35]	413	51.63
Valletta *et al.* [36]	333	41.63
Sharma and Sharma [37]	332	47.43
Chang *et al.* [38]	331	47.29
Bhandary *et al.* [39]	326	65.20

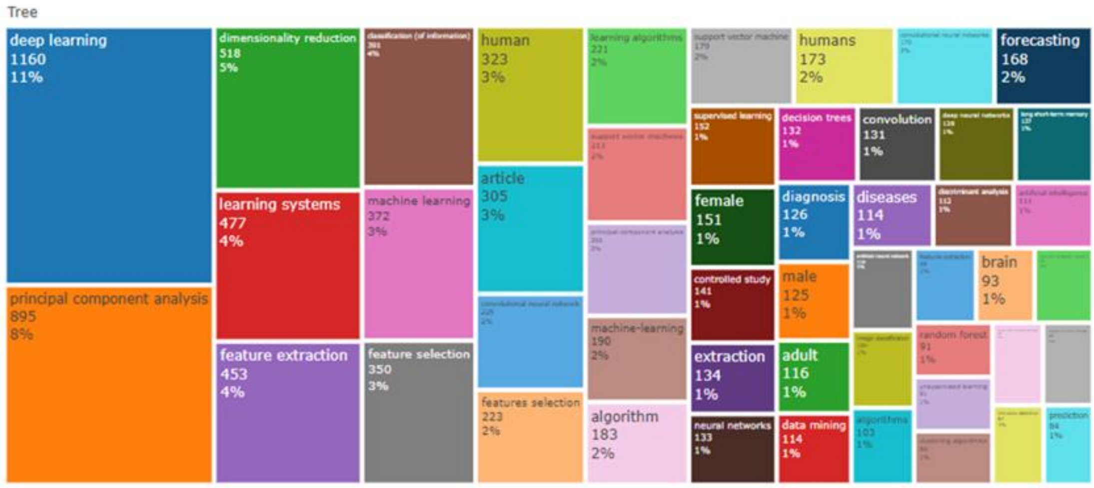

Figure 6. Tree map of the most frequently used keywords in feature selection research

3.2. Conceptual structure: network approach

We use the network approach as a conceptual structure to visualize the relationships between key concepts, themes, or topics within a body of literature on feature selection techniques. It helps us examine how concepts are linked through co-occurrence in publications. This approach typically involves creating networks based on the frequency with which certain keywords or terms appear together in articles.

Figure 7 represents a clustering of the co-occurrence network based on keywords plus. The map highlights two distinct research areas: i) deep learning, feature extraction, and selection techniques, and ii) machine learning techniques, particularly in human-related studies and clinical research. The visualization provides a conceptual understanding of how these two areas are developed in the literature, with deep learning and principal component analysis playing prominent roles in feature selection and dimensionality reduction.

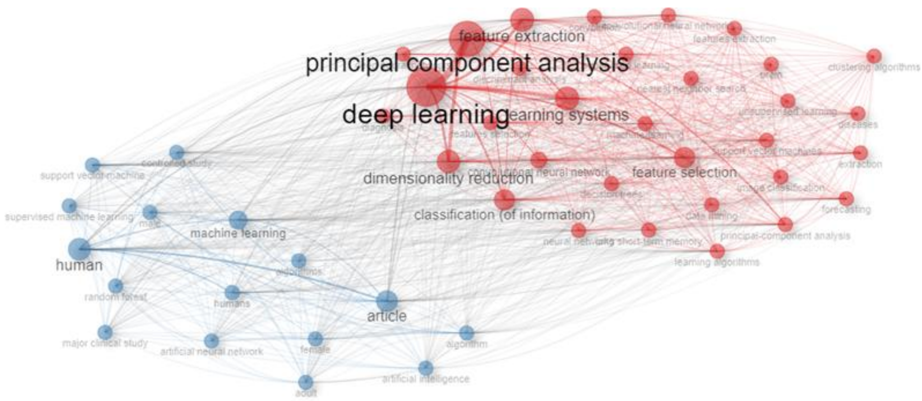

Figure 7. Co-occurrence network: keywords plus

Figure 8 represents a thematic map. It categorizes different research topics or themes based on two key metrics: centrality and density. In this graph, the x-axis (relevance degree/centrality) measures how central or important a theme is within the broader research field, while the y-axis (development degree/density) indicates how developed a theme is within the literature.

Deep learning, machine learning, and feature selection are the most central and well-developed themes, making them critical to the research field. Principal component analysis, convolutional neural networks (CNN), and support vector machine (SVM) are foundational topics but are still in development and require further attention. Deep belief networks and manifold learning are highly developed but niche areas, indicating that they are important within specific research subfields but less influential overall.

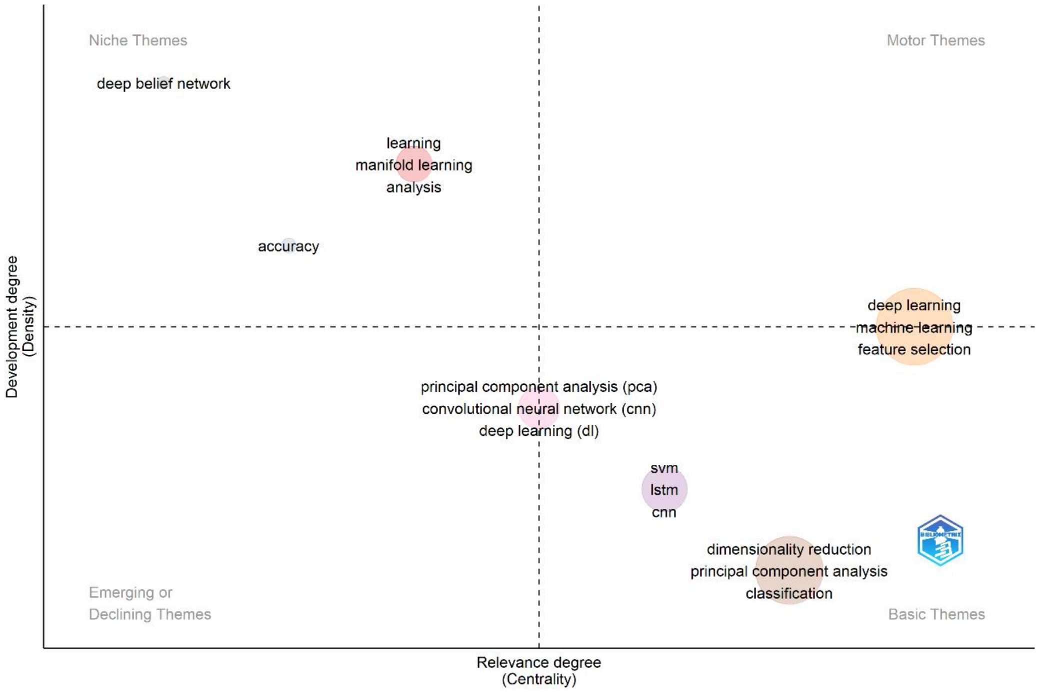

Figure 8. Thematic map: authors' keywords

3.3. Social structure

This section examines the collaboration networks of authors, institutions, and countries involved in feature selection research. In the three diagrams presented, nodes correspond to authors, institutions, or countries, and the edges illustrate their cooperative links. Figure 9 illustrates clusters of co-authors working together, with authors Wang [33], [40]–[45], Wang [46], [47], and Wang [33], [48]–[50] having the largest number of collaborators in producing documents on feature selection techniques.

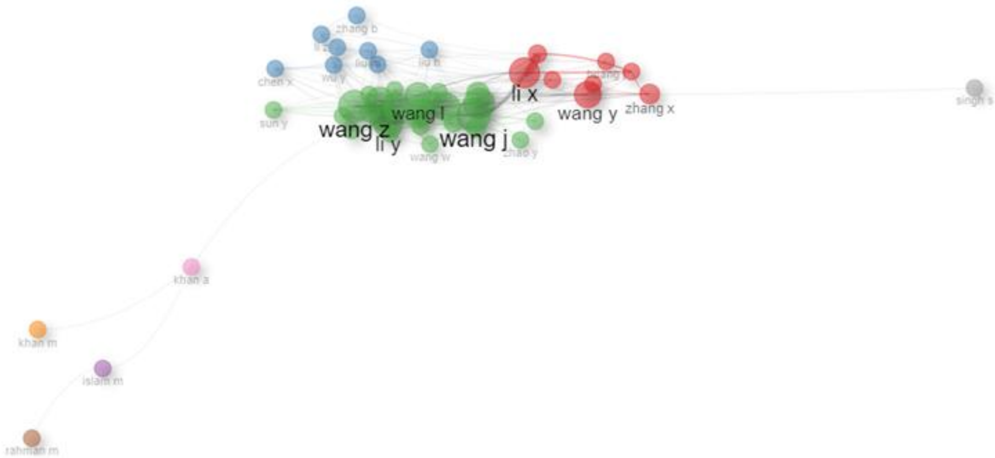

Figure 9. Authors' collaboration network

Figure 10 highlights that China has the highest level of collaboration, as well as the largest number of published documents, consistent with previous statistics. In terms of affiliations, Figure 11 maps the co-affiliation network of institutions. Institutions like COMSATS University Islamabad, King Saud University, and Shenzhen University demonstrate strong collaboration among their affiliated authors.

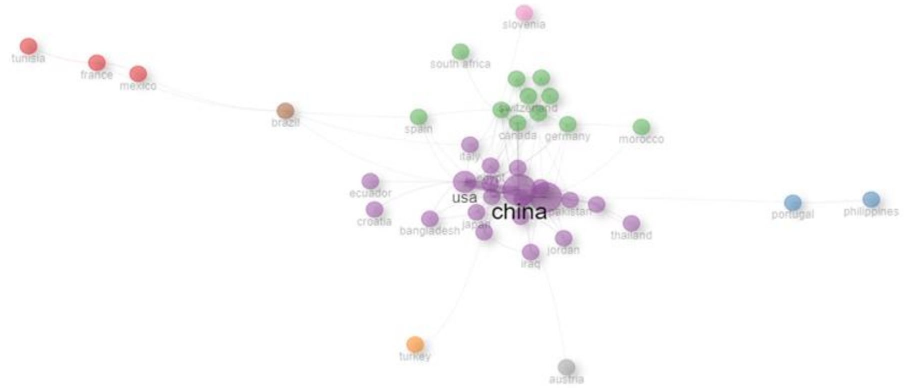

Figure 10. Countries' collaboration network

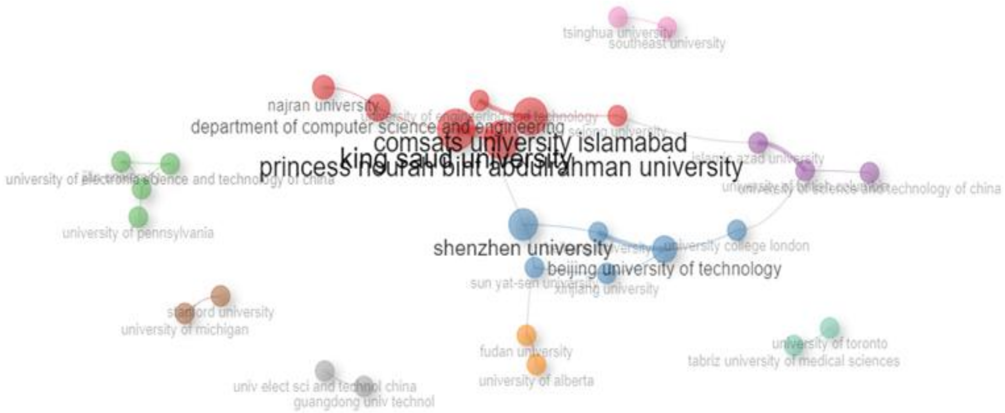

Figure 11. Institutions' collaboration network

4. CONCLUSION

In this study, we conducted a detailed bibliometric analysis to examine the development and impact of feature selection techniques within the fields of machine learning and artificial intelligence. Our analysis drew from 2,079 documents published between 2014 and 2024 in the Scopus and Web of Science databases. To our knowledge, this is one of the most comprehensive bibliometric studies on feature selection, providing insights into key emerging trends and collaborations across diverse research domains. Our findings indicate a substantial increase in research output from 2014 to 2023, reflecting the growing importance of feature selection techniques in addressing challenges associated with high-dimensional data. Countries such as China, India, and the USA emerged as leading contributors to the field, highlighting their significant role in advancing feature selection methodologies. While this study provides valuable insights, it is essential to recognize that the specific keywords used to retrieve data may have influenced our results. Future research could refine and expand the scope by incorporating alternative keywords. Additionally, since this analysis relied on Scopus and Web of Science data, subsequent studies could broaden the perspective by including other databases to capture a more extensive range of contributions to feature selection research.

ACKNOWLEDGEMENTS

The support for this research is provided by the Ministry of Higher Education, Scientific Research, and Innovation, as well as the Digital Development Agency (DDA) and the National Center for Scientific and Technical Research (CNRST) of Morocco, under the Smart DLSP Project - Al Khawarizmi IA-Program.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Oumaima Semmar	✓	✓		✓	✓				✓	✓			✓	
Wissal El Habti	✓	✓	✓			✓		✓	✓					
Donalson Wilson	✓		✓		✓				✓	✓	✓			
Abdellah Azmani	✓			✓			✓			✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The bibliometric data supporting the findings of this study were obtained from the Scopus and Web of Science databases, which are subject to subscription access. The derived dataset generated after screening, cleaning, and merging these records is available from the corresponding author [WEH] upon reasonable request.

REFERENCES

[1] C. F. Worel, "The curse of dimensionality: when too many features break your machine learning model," *SSRN*, 2025, doi: 10.2139/ssrn.5206704.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed., New York, United States: Springer, 2009, doi: 10.1007/978-0-387-84858-7.

[3] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[4] C. Yutong and H. Jin, "A bibliometric and visualization analysis of artificial intelligence and machine learning," in *2024 IEEE 4th International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*, IEEE, 2024, pp. 219–222, doi: 10.1109/ICEIB61477.2024.10602604.

[5] J. Li *et al.*, "Feature selection: a data perspective," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, 2018, doi: 10.1145/3136625.

[6] D. Yang and J. Hu, "A detection method of oil content for maize kernels based on CARS feature selection and deep sparse autoencoder feature extraction," *Industrial Crops and Products*, vol. 222, 2024, doi: 10.1016/j.indcrop.2024.119464.

[7] H. Zhou, P. Zheng, J. Dong, J. Liu, and Y. Nakanishi, "Interpretable feature selection and deep learning for short-term probabilistic PV power forecasting in buildings using local monitoring data," *Applied Energy*, vol. 376, 2024, doi: 10.1016/j.apenergy.2024.124271.

[8] X. Jin, B. Wei, L. Deng, S. Yang, J. Zheng, and F. Wang, "An adaptive pyramid PSO for high-dimensional feature selection," *Expert Systems with Applications*, vol. 257, 2024, doi: 10.1016/j.eswa.2024.125084.

[9] L. K. Kumari and B. N. Jagadesh, "An adaptive teaching learning based optimization technique for feature selection to classify mammogram medical images in breast cancer detection," *International Journal of System Assurance Engineering and Management*, vol. 15, no. 1, pp. 35–48, 2024, doi: 10.1007/s13198-021-01598-7.

[10] H. Yang *et al.*, "Enhancing metastatic colorectal cancer prediction through advanced feature selection and machine learning techniques," *International Immunopharmacology*, vol. 142, 2024, doi: 10.1016/j.intimp.2024.113033.

[11] A. Taha, B. Cosgrave, and S. Mckeever, "Using feature selection with machine learning for generation of insurance insights," *Applied Sciences*, vol. 12, no. 6, 2022, doi: 10.3390/app12063209.

[12] M. Fernandes, A. Canito, V. B.-Canedo, L. Conceição, I. Praça, and G. Marreiros, "Data analysis and feature selection for predictive maintenance: a case-study in the metallurgic industry," *International Journal of Information Management*, vol. 46, pp. 252–262, Jun. 2019, doi: 10.1016/j.ijinfomgt.2018.10.006.

[13] X. Liu *et al.*, "Adapting feature selection algorithms for the classification of Chinese texts," *Systems*, vol. 11, no. 9, 2023, doi: 10.3390/systems11090483.

[14] E. M. Maseno and Z. Wang, "Hybrid wrapper feature selection method based on genetic algorithm and extreme learning machine for intrusion detection," *Journal of Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00887-9.

[15] J. W. Chen and J. Dhahbi, "Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods," *Scientific Reports*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-92725-8.

[16] A. Kandiri, R. Ghiasi, M. Nogal, and R. Teixeira, "Travel time prediction for an intelligent transportation system based on a data-driven feature selection method considering temporal correlation," *Transportation Engineering*, vol. 18, 2024, doi: 10.1016/j.treng.2024.100272.

[17] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: an overview and guidelines," *Journal of Business Research*, vol. 133, pp. 285–296, 2021, doi: 10.1016/j.jbusres.2021.04.070.

[18] R. Kumar, "Bibliometric analysis: comprehensive insights into tools, techniques, applications, and solutions for research excellence," *Spectrum of Engineering and Management Sciences*, vol. 3, no. 1, pp. 45–62, 2025, doi: 10.31181/sems31202535k.

[19] R. Joshi, H. Gadikta, S. Kharat, S. Mandal, and K. Kadam, "Bibliometric of feature selection using optimization techniques bibliometric of feature selection using optimization techniques in healthcare using Scopus and Web of Science databases," *Library Philosophy and Practice (e-journal)*, pp. 1-36, 2021.

[20] G. A. S. Ribeiro, R. M. Barbosa, M. C. Reis, and N. L. Costa, "From bibliometrics to text mining: exploring feature selection methods in microarray research," *Communications in Statistics - Simulation and Computation*, vol. 54, no. 8, pp. 3017–3033, 2025, doi: 10.1080/03610918.2024.2331083.

[21] D. W. Hook, S. J. Porter, and C. Herzog, "Dimensions: building context for search and evaluation," *Frontiers in Research Metrics and Analytics*, vol. 3, doi: 10.3389/frma.2018.00023.

[22] E. W. Sayers *et al.*, "Database resources of the national center for biotechnology information," *Nucleic Acids Research*, vol. 50, no. D1, pp. D20–D26, 2022, doi: 10.1093/nar/gkab1112.

[23] P. Mongeon and A. P.-Hus, "The journal coverage of Web of Science and Scopus: a comparative analysis," *Scientometrics*, vol. 106, no. 1, pp. 213–228, 2016, doi: 10.1007/s11192-015-1765-5.

[24] M. I. E. Fernández, P. L. Barbosa, and A. P. Guerrero, "Web of Science vs. Scopus: un estudio cuantitativo en ingeniería química," *Anales de Documentación*, vol. 13, pp. 159–175, 2010.

[25] P. Tugwell and D. Tovey, "PRISMA 2020," *Journal of Clinical Epidemiology*, vol. 134, pp. A5–A6, 2021, doi: 10.1016/j.jclinepi.2021.04.008.

[26] P. L. de Micheaux, R. Drouilhet, and B. Liquet, *The R software: fundamentals of programming and statistical analysis*, New York, United States: Springer, 2013, doi: 10.1007/978-1-4614-9020-3.

[27] H. Derviş, "Bibliometric analysis using bibliometrix an R package," *Journal of Scientometric Research*, vol. 8, no. 3, pp. 156–160, 2020, doi: 10.5530/jscires.8.3.32.

[28] M. Aria and C. Cuccurullo, "Bibliometrix : an R-tool for comprehensive science mapping analysis," *Journal of Informetrics*, vol. 11, no. 4, pp. 959–975, 2017, doi: 10.1016/j.joi.2017.08.007.

[29] A. J. Lotka, "The frequency distribution of scientific productivity," *Journal of the Washington Academy of Sciences*, vol. 16, no. 13, pp. 317–323, 1926.

[30] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014, doi: 10.1109/JSTARS.2014.2329330.

[31] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in *Computer Vision – ECCV 2014*, Cham, Switzerland: Springer, 2014, pp. 584–599, doi: 10.1007/978-3-319-10590-1_38.

[32] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2381–2392, 2015, doi: 10.1109/JSTARS.2015.2388577.

[33] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, 2016, doi: 10.1016/j.neucom.2015.08.104.

[34] P. S. Gromski *et al.*, "A tutorial review: metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding," *Analytica Chimica Acta*, vol. 879, pp. 10–23, Jun. 2015, doi: 10.1016/j.aca.2015.02.012.

[35] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. non-handcrafted features for computer vision classification," *Pattern Recognition*, vol. 71, pp. 158–172, 2017, doi: 10.1016/j.patcog.2017.05.025.

[36] J. J. Valletta, C. Torney, M. Kings, A. Thornton, and J. Madden, "Applications of machine learning in animal behaviour studies," *Animal Behaviour*, vol. 124, pp. 203–220, 2017, doi: 10.1016/j.anbehav.2016.12.005.

[37] S. Sharma and D. Sharma, "Intelligently applying artificial intelligence in chemoinformatics," *Current Topics in Medicinal Chemistry*, vol. 18, no. 20, pp. 1804–1826, 2018, doi: 10.2174/1568026619666181120150938.

[38] P. Chang *et al.*, "Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas," *American Journal of Neuroradiology*, vol. 39, no. 7, pp. 1201–1207, 2018, doi: 10.3174/ajnr.A5667.

[39] A. Bhandary *et al.*, "Deep-learning framework to detect lung abnormality – a study with chest X-ray and lung CT scan images," *Pattern Recognition Letters*, vol. 129, pp. 271–278, 2020, doi: 10.1016/j.patrec.2019.11.013.

[40] Z. Wang, C. Chen, J. Li, F. Wan, Y. Sun, and H. Wang, "ST-CapsNet: linking spatial and temporal attention with capsule network for P300 detection improvement," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 991–1000, 2023, doi: 10.1109/TNSRE.2023.3237319.

[41] Z. Wang, H. Chen, S. Yang, X. Luo, D. Li, and J. Wang, "A lightweight intrusion detection method for IoT based on deep learning and dynamic quantization," *PeerJ Computer Science*, vol. 9, 2023, doi: 10.7717/peerj-cs.1569.

[42] Z. Wang, C. Hu, W. Liu, X. Zhou, and X. Zhao, "EEG-based high-performance depression state recognition," *Frontiers in Neuroscience*, vol. 17, 2024, doi: 10.3389/fnins.2023.1301214.

[43] Z. Wang, J. Guan, X. Wang, W. Wang, L. Xing, and F. Alharbi, "The danger of minimum exposures: understanding cross-app information leaks on iOS through multi-side-channel learning," in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA: ACM, 2023, pp. 281–295, doi: 10.1145/3576915.3616655.

[44] Z. Wang, K. Zhang, and B. Wang, "Detection of COVID-19 cases based on deep learning with X-ray images," *Electronics*, vol. 11, no. 21, 2022, doi: 10.3390/electronics11213511.

[45] Z. Wang, C. Zou, and W. Cai, "Small sample classification of hyperspectral remote sensing images based on sequential joint deeping learning model," *IEEE Access*, vol. 8, pp. 71353–71363, 2020, doi: 10.1109/ACCESS.2020.2986267.

[46] J. Wang, X. Dai, H. Luo, C. Yan, G. Zhang, and J. Luo, "MI_DenseNetCAM: a novel pan-cancer classification and prediction method based on mutual information and deep learning model," *Frontiers in Genetics*, vol. 12, 2021, doi: 10.3389/fgene.2021.670232.

[47] J. Wang, Y. Wu, B. Li, Z. Yang, and F. Nie, "Fast anchor graph preserving projections," *Pattern Recognition*, vol. 146, 2024, doi: 10.1016/j.patcog.2023.109996.

[48] Y. Wang, X. Wang, M. M. Ariffin, M. Abolfathi, A. Alqhatani, and L. Almutairi, "Attack detection analysis in software-defined networks using various machine learning method," *Computers and Electrical Engineering*, vol. 108, 2023, doi: 10.1016/j.compeleceng.2023.108655.

[49] Y. Wang, "Unsupervised representative feature selection algorithm based on information entropy and relevance analysis," *IEEE Access*, vol. 6, pp. 45317–45324, 2018, doi: 10.1109/ACCESS.2018.2863752.

[50] Y. Wang, Q. Chen, H. Shao, R. Zhang, and H. Shen, "Generating bulk RNA-Seq gene expression data based on generative deep learning models and utilizing it for data augmentation," *Computers in Biology and Medicine*, vol. 169, 2024, doi: 10.1016/j.compbiomed.2023.107828.

BIOGRAPHIES OF AUTHORS

Oumaima Semmar holds a bachelor's degree in Mathematics and Applications, followed by a Master's degree in Applied Analysis and Statistical Engineering. She is currently a Ph.D. student in Data Science and Artificial Intelligence at the Intelligent Automation and Bio Med Genomics Laboratory (IABL), Faculty of Science and Technology of Tangier, Abdelmalek Essaadi University. She is focusing her research on the transformative potential of artificial intelligence in the insurance industry. She aims to bridge the gap between theoretical mathematical frameworks and practical AI solutions for insurance challenges. She can be contacted at email: oumaima.semmar1@etu.uae.ac.ma.

Wissal El Habti earned a bachelor's degree in Statistical Engineering, followed by a master's degree in Applied Analysis and Statistical Engineering. She is currently pursuing a Ph.D. in Data Science and Artificial Intelligence at Abdelmalek Essaadi University, within the Intelligent Automation and Bio Med Genomics Laboratory (IABL), Faculty of Science and Technology, Tangier, Morocco. Her research focuses on the application of artificial intelligence in oncology, with an emphasis on deep learning techniques such as convolutional neural networks (CNNs), vision transformers (ViTs), and generative adversarial networks (GANs). By leveraging these advanced models alongside statistical methodologies, she aims to develop innovative tools to improve cancer diagnosis, prediction, and treatment, contributing to the advancement of precision medicine. She can be contacted at email: wissal.elhabti@etu.uae.ac.ma.

Donalson Wilson holds a Bachelor of Statistical Engineering, a Master of Applied Analysis and Statistical Engineering. Currently, he is a Ph.D. student in Data Science and Artificial Intelligence at Abdelmalek Essaadi University, Faculty of Science and Technology of Tangier, Intelligent Automation and Bio Med Genomics Laboratory (IABL), Morocco. He can be contacted at email: donalson.wilson@etu.uae.ac.ma.

Dr. Abdellah Azmani obtained his doctorate in industrial computing, dynamic systems modeling, and artificial intelligence at the University of Sciences and Technologies of Lille in 1991. He worked as a Professor at the Ecole Centrale de Lille, France, and at the Institute of Computer and Industrial Engineering from Lens, France. He is a professor in the Department of Computer Science at the Faculty of Sciences and Technologies of Tangier, Morocco. He is a founding member and director of the Intelligent Automation and BioMedGenomics Laboratory. He has directed numerous doctoral theses and scientific research projects. Its work focuses particularly on the development of generic and disruptive concepts promoting the integration of artificial intelligence and digital transformation, through the creation of decision-making IT solutions for public and private organizations. He can be contacted at email: a.azmani@uae.ac.ma.