J. Innovative Eng. Nat. Sci., vol. 4, no. 2, pp. 642-657, 2024.
http://doi.org/10.61112/jiens.1475948

JIENS
Original Paper
Journal of Innovative Engineering and Natural Science
(Yenilikçi Mühendislik ve Doğa Bilimleri Dergisi)
https://dergipark.org.tr/en/pub/jiens

How to combine ML and MCDM techniques: an extended bibliometric analysis

Mehmet Asaf Düzen[a] İsmail Buğra Bölükbaşı[a,*] and Eyüp Çalık[a]

[a]*Department of Industrial Engineering, Yalova University, 77200, Yalova, Türkiye.*

ARTICLE INFO

Article history:
Received 30 April 2024
Received in revised form 26 June 2024
Accepted 22 July 2024
Available online

Keywords:
Machine learning
Multi criteria decision making
Bibliometric analysis
VOSviewer

ABSTRACT

Machine Learning (ML) and Multi Criteria Decision Making (MCDM) are popular methods that have recently been widely used in many different fields. Due to the increasing use of these two methods together, there is a need for a bibliometric analysis in this area. In this study, an extended author-developed bibliometric analysis was performed on 1189 publications retrieved from the Web of Science (WoS) and Scopus databases between January 2000 and April 2024. In the initial bibliometric analysis, as a generic part, the VOSviewer program was used to make the data meaningful. In particular, the analysis was carried out according to years and relationships related to the keyword analysis. In addition, the most frequently used keywords were identified, and the direction of the trend was determined. During the initial bibliometric analysis, 308 publications were analysed, with 297 publications retrieved from the WoS database and 11 publications from Scopus. The study distinguishes itself from the existing literature by establishing new models and categories as an extended part of bibliometric analysis. Using these models and categories, we sought to answer questions about how researchers use ML and MCDM together and in what direction these methods are evolving. In this context, the distribution of models and categories in different research areas and their changes over the years were analysed. This study provides researchers with a comprehensive perspective on the various combination possibilities when integrating ML and MCDM techniques.

I. INTRODUCTION

The selection of the most appropriate alternative among multiple alternatives according to the specified criteria is referred to as decision-making. In recent years, the importance of hybrid methods for solving decision-making problems has increased. The combination of Machine Learning (ML) and Multi-Criteria Decision-Making (MCDM) methods is often used in the application of hybrid methods. ML is a component of artificial intelligence with a wide range of applications, although it is often associated with the data mining of large databases [1]. ML has attracted attention for solving highly complex problems over the last few decades [2]. Many ML algorithms have been applied in various studies. Choosing the best classification algorithm for a given dataset is important for accurate prediction. There is no single algorithm or model that can attain optimal performance for a particular problem domain [3]. Because of this, many ML algorithms have been applied in various studies and choosing the best classification algorithm for a given dataset is important for accurate prediction. Therefore, to solve a problem, multiple ML algorithms are used. The most appropriate algorithm is found by ranking the results obtained from different algorithms, to make the studies more effective and useful. Mean squared error, root mean square error, accuracy, sensitivity, and precision rates are used when comparing algorithm performances [4-5]. Besides ML, MCDM methods are also widely used in the decision-making process. The aim of MCDM is to identify the optimal option by considering multiple criteria during the selection process [6]. Therefore, combining both methods yields more effective results.

*Corresponding author. Tel.: +90-226-815-5361; e-mail: ismail.bolukbasi@yalova.edu.tr

Although ML provides the ability to learn from large data sets, it is difficult to process, organize and select important features. MCDM methods are used to determine which features are important and to process the data in a more meaningful way. Chowdhury et al. [7] integrated ML and MCDM methods to detect COVID-19 from cough sound in their study. Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) and Entropy methods were used in MCDM methods for criteria features, and 9 ML methods such as Random Forest (RF) and Multilayer Perceptron (MLP) were applied to classify the data.

MCDM methods can consider multiple criteria and provide a more precise indication of preference. Additionally, the utilization of MCDM techniques within ML methods can enhance their decision-making abilities [8]. Choudhary et al. [9] performed a study to compare the performance of two ML methods, Extreme Gradient Boosting (XGB) and Extreme Random Trees (ET), for the determination of groundwater potential zones using TOPSIS, one of the MCDM methods. Mustapha et al. [10] employed the Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) technique to conduct a robust comparison of the performance of five different ML methods in their breast cancer screening study. As a result, more optimal, accurate and comprehensive results are obtained by using ML and MCDM methods together.

ML methods are data-driven, while MCDM methods such as Analytic Hierarchy Process (AHP) and Stepwise Weight Assessment Ratio Analysis (SWARA) are knowledge-driven, so they are compared by using them separately. This ensures that the subjective judgments in MCDM are compared with real values. Gudiyangada et al. [11] applied AHP and Analytic Network Process (ANP) methods from MCDM methods, RF and Support Vector Machine (SVM) methods from ML methods and compared the results of all methods. Viviani & Pasi [12] conducted a comparison of correctly classifying User Generated Content (UGC) MCDM method, in addition to the ML methods commonly used in the literature. The analysis showed that the MCDM method provides results comparable to the ML methods. The mentioned studies indicate that ML and MCDM methods are used together because they have complementary advantages. Çalık [12] evaluated the use of these two methods together in employee selection and found that four different approaches emerged. According to these findings, the first approach is to use ML methods first and then MCDM methods. The second approach involves the application of MCDM methods first and then the integration of ML methods. The third approach refers to an approach in which ML performance is ranked by using of MCDM methods. Finally, the fourth approach represents a methodology in which ML and MCDM methods are compared separately. However, Çalık [13], examined the integration of ML and MCDM methods only in the employee selection problem through these four models without any bibliometric analysis. In addition, Liao et al. [14] summarized the four challenges of applying MCDM today and investigated the use of ML methods for MCDM in criteria extraction, criteria interaction, parameter identification, and integrated solution problems through bibliometric analysis. Although this approach allowed the identification of the benefits that ML techniques can provide for MCDM, they focused on the single role of ML in MCDM in the topic of combining these two methods and addressed it in specific domains such as business management, industrial engineering, sustainable development, and emergency management. Considering these two studies, it should be noted that they do not fully investigate the integration of ML and MCDM methods in all aspects and their research areas are limited to specific fields. Despite the prevalence of studies using both ML and MCDM methods, there is no other study in the literature that provides a comprehensive bibliometric analysis of their combined use. In this context, to answer our research question on how ML and MCDM methods can be combined and to fill the gap in this field, this study presents an extended bibliometric analysis with proposed models and categories of these methods without limiting any research area. This research is expected to shed light on the trend of combining ML and MCDM, in which research areas its use is increasing and how it has changed over the years.

The remainder of the study is organized as follows: Section 2 provides a detailed outline of the bibliometric analysis, including an extended version proposed as the research methodology. Section 3 presents the data visualized with VOSviewer and the results of the extended bibliometric analysis. Finally, the conclusion section summarizes the general findings and implications of the research with recommendations for future work.

II. TEORETICAL METHOD

In this chapter, the definition and scope of bibliometric analysis are first explained. Then, bibliometric analysis is treated in two steps. In the first step, the processes of initial bibliometric analysis are explained as a generic part, and in the second step, the extended bibliometric analysis is discussed. A flow chart of the methodology of the study is shown in Figure 1.

2.1 Definition and scope of bibliometric analysis

Bibliometric analysis can be defined as a systematic, popular, and effective method that allows the analysis and inference of studies in the literature based on certain characteristics through a large-scale search using mathematical and statistical techniques [15-17]. There are several types of bibliometric analysis, including keyword analysis, content analysis, country analysis, citation analysis, publication year analysis, journal analysis, and author analysis [18]. Based on the literature review, citation analysis and content analysis were the most widely used [19]. Databases such as Scopus, WoS, and PubMed were used for the bibliometric analysis. The WoS database was chosen for this study because it contains the oldest and most comprehensive record of citation indexes and is a useful analytical tool [20].

2.2 Initial version of bibliometric analysis

A review of the literature shows that the bibliometric analysis process is generally categorized under the headings of identification, screening, relevance, evaluation and included [21-22]. The initial bibliometric analysis is a generic step and in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guideline [23], this study was conducted in three phases: identification, screening, and compliance. These phases are illustrated in Figure 1. The query utilized for this research is presented in the identification category.

The search query used in WoS is as follows: TS = (“machine learning” OR “artificial learning” OR “data mining” OR “ML” OR “deep learning”) AND TS= (“multi criteria” OR “multi objective decision making” OR “multi attribute” OR “multi-criteria” OR “multi-attribute” OR “multi-objective decision making” OR “MCDM” OR “MADM” OR “MODM”). As a result of this query, 1189 publications were obtained. When conducting research, it is expected that the research question fully addresses the purpose of the study. Therefore, it is crucial to select relevant statements that do not deviate from the objective of the research. In this query, the term "multi objective decision making" is used. When the term "multi objective" is added to the query instead of "multi objective decision making", the number of publications increases from 1189 to 4266. Although the number of publications increased, the query lost its relevance as it went beyond the research area. For this reason, the term "multi objective decision making" is used in the query. In the screening and eligibility process, 297 studies were obtained by excluding 892 publications that were repetitive, irrelevant, non-English, and used only ML or MCDM methods from 1189 publications. 11 articles not found in the WoS database query but retrieved from the Scopus database were added. The finalized 308 publications are indicated in the included.

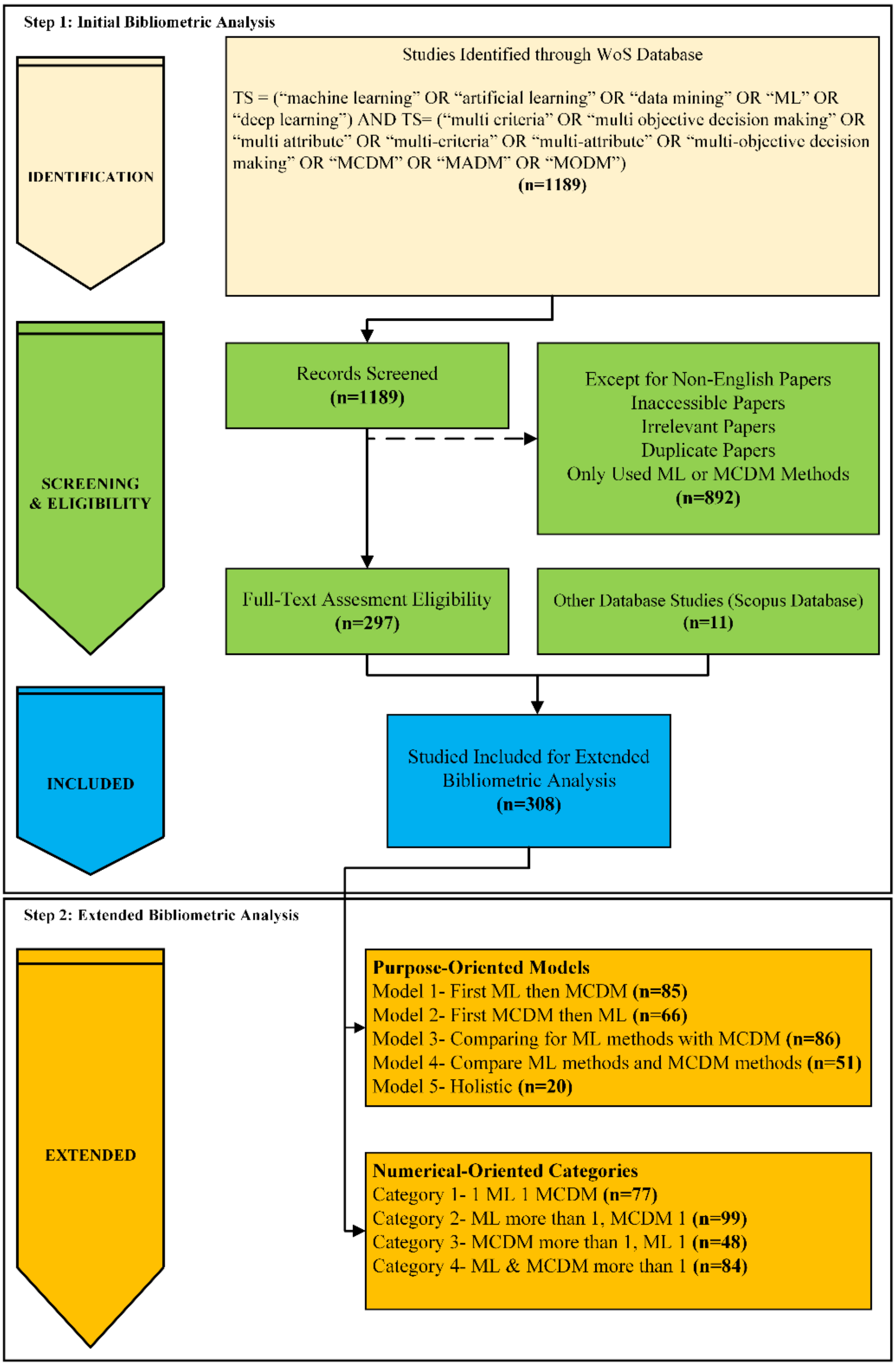

Figure 1. Flowchart of bibliometric analysis

The search query used in WoS is as follows: TS = (“machine learning” OR “artificial learning” OR “data mining” OR “ML” OR “deep learning”) AND TS= (“multi criteria” OR “multi objective decision making” OR “multi attribute” OR “multi-criteria” OR “multi-attribute” OR “multi-objective decision making” OR “MCDM” OR “MADM” OR “MODM”). As a result of this query, 1189 publications were obtained. When conducting research, it is expected that the research question fully addresses the purpose of the study. Therefore, it is crucial to select relevant statements that do not deviate from the objective of the research. In this query, the term "multi objective decision making" is used. When the term "multi objective" is added to the query instead of "multi objective decision making", the number of publications increases from 1189 to 4266. Although the number of publications increased, the query lost its relevance as it went beyond the research area. For this reason, the term "multi objective decision making" is used in the query. In the screening and eligibility process, 297 studies were obtained by excluding 892 publications that were repetitive, irrelevant, non-English, and used only ML or MCDM methods from 1189 publications. 11 articles not found in the WoS database query but retrieved from the Scopus database were added. The finalized 308 publications are indicated in the included.

2.3 Extended version of bibliometric analysis

Extended bibliometric analysis is the step in which purpose-oriented models are developed about the research topic and numerical-oriented categories are created and analyzed. This study differs from the studies in the literature in this regard. The purpose-oriented models are categorized into five models based on the combination of ML and MCDM methods. These models, their definitions and number of publications are described in Table 1.

Numerical-oriented models are classified into four categories based on the frequency of using MCDM and ML approaches. These categories, their definitions and number of publications are described in Table 2.

Table 1. Models and Definitions

Models	Definitions	Publications
Model 1 (First ML then MCDM)	First, ML methods are used to obtain outputs and solve the problem. Subsequently, these outputs serve as input in MCDM methods.	85
Model 2 (First MCDM then ML)	First, MCDM methods are employed to obtain outputs and solve the problem. These outputs are then utilized as input for ML methods.	66
Model 3 (Comparing ML methods with MCDM)	The performance of ML methods is compared using MCDM methods.	86
Model 4 (Comparing ML methods and MCDM methods)	MCDM and ML techniques were applied to the problem, and their results were compared.	51
Model 5 (Holistic)	ML and MCDM methods are used in holistic relationships. This model offers two distinct approaches: ML-MCDM-ML, and MCDM-ML-MCDM. The first approach leverages the ML method as input for the MCDM method, and then utilizes the output obtained from the MCDM method as input for the ML method. In the alternative approach, the MCDM method is first utilized as an input for ML. Then, the results obtained by the ML approach are used as inputs within the MCDM methodology.	20

Table 2. Categories and Definitions

Categories	Definitions	Publications
Category 1 (1 ML, 1 MCDM)	Only 1 MCDM method and only 1 ML method were used in the studies in the literature.	77
Category 2 (ML more than 1, MCDM 1)	While different ML methods (k-NN, LR, RF etc.) are used, only 1 MCDM method is used.	99
Category 3 (MCDM more than 1, ML 1)	While different MCDM methods (AHP, TOPSIS, SAW, etc.) were used, only 1 ML method was used.	48
Category 4 (ML & MCDM more than 1)	More than 1 MCDM (AHP, TOPSIS, SAW etc.) and more than 1 ML method (k-NN, LR, RF etc.) were used.	84

III. RESULTS AND DISCUSSIONS

This section presents the data visualized with VOSviewer and the results of the extended bibliometric analysis. A total of 1189 publications were evaluated using keyword analysis to identify clusters of studies and new research trends. Many packages with different features are used for data visualization, such as VOSviewer, BibExcel, Bibliometrix, Pajek and SciMAT. VOSviewer is an excellent information mapping tool for scientific landscapes using network visualization and density visualization [24]. In this study, the VOSviewer package program was preferred because it offers the possibility of analysis suitable for the desired situation. Two analyses were performed in VOSviewer: the relationship with keywords and the change of keywords over the years. Bibliometric data were generated from WoS and imported into VOSviewer. A visualization for keyword analysis was created from the data file. In the program, co-occurrence and author keywords were selected for analysis and a full count was requested. While 166 out of 3960 keywords meet the threshold, there are on average 4 keywords per article. The keyword analysis by relation is shown in Figure 2.

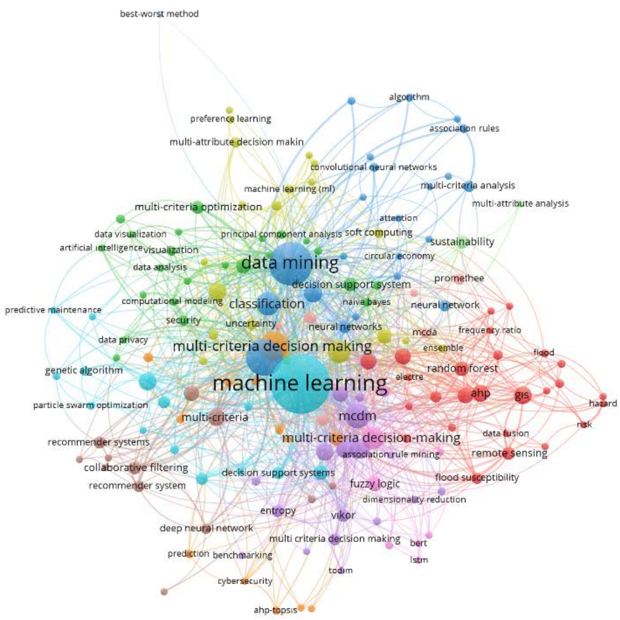

Figure 2. Keyword analysis by relation [Data source: WoS, Search date: 23 April 2024]

The cluster size represents the frequency of occurrence of a term. Therefore, a larger cluster represents a higher frequency of an occurrence. The distance between keywords indicates the relevance between them. That is, shorter distance indicates stronger relevance and longer distance indicates weaker relevance. Each cluster is marked with a different color. The most important and largest cluster, machine learning, is colored blue. This cluster is closely related to 3 highly cited references such as deep learning, data mining and multi-criteria decision making. The 166 items are grouped into 12 classes. A map of the clusters of keyword items by average year of publication is shown in Figure 3.

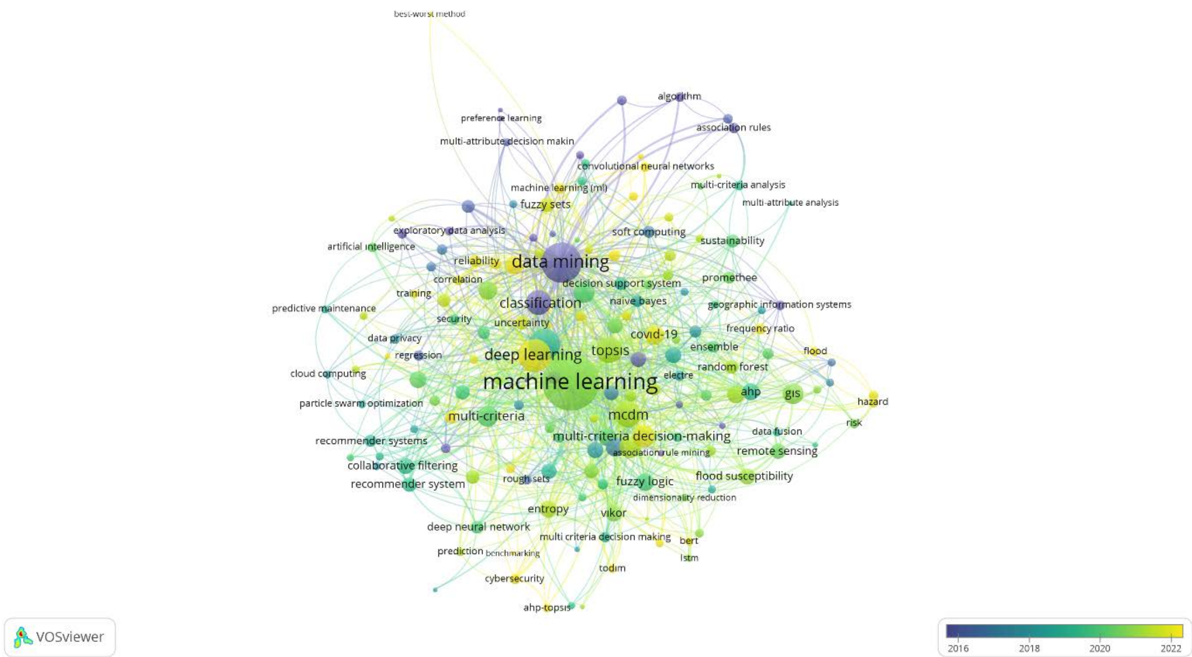

Figure 3. Keyword analysis by publication year [Data source: WoS, Search date: 23 April 2024]

Keyword items in 2022 are shown in yellow, and those in 2014 are shown in purple. The colors indicate the average publication year of the terms and their evolution over time in topics such as machine learning and deep learning. It can be observed that studies on data mining and classification were intensive between 2014-2016, machine learning and artificial intelligence were focused on between 2016-2020, and the interest in deep learning and multi-criteria decision-making have increased since 2020. The top 10 most frequently used words in the field of ML and MCDM are listed as shown in Table 3, which shows that he most frequently used keywords in the analyzed publications are "Machine learning", "Data Mining", "Multi-Criteria Decision Making" and "Deep learning". In general, studies on machine learning were the most popular topic.

Table 3. Top 10 keywords [Data source: WoS, Search date: 23 April 2024]

ID	Keyword	Occurrences	Occurrences (%)	Total link strength
1	machine learning	239	31	344
2	data mining	127	16	171
3	multi criteria decision making	114	15	194
4	deep learning	92	12	117
5	mcdm	44	6	73
6	classification	38	5	68
7	topsis	35	5	75
8	ahp	35	5	65
9	artificial intelligence	27	3	52
10	decision making	22	3	38

An extended bibliometric analysis was performed according to the flowchart in Figure 1 in the Methodology section. Accordingly, the 308 publications obtained from the initial bibliometric analysis were evaluated according to publication types, changes by years, research areas, journals, models, and categories.

Figure 4 shows that 308 publications fall into different publication categories. The analysis of the data shows that 83% of these works were classified as "Article", 9% as "Proceeding Paper", 6% as "Article; Early Access", 1% as "Review" and 1% as "Article; Proceedings Paper". Analyzing these results, the highest rate is for "Article". The ratio of articles to total publications is markedly higher than that of conference proceedings, suggesting that the research domain has reached a point of relative maturity.

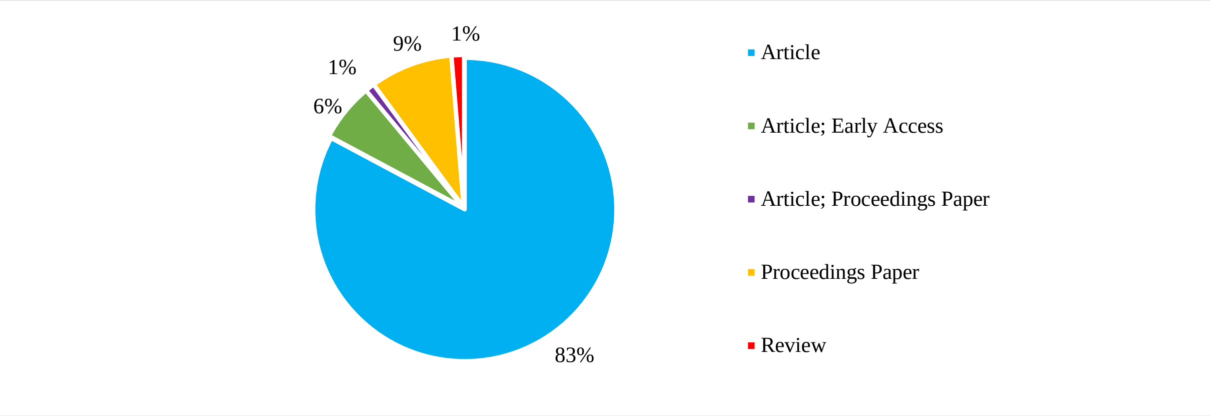

Figure 4. Classification of documents in terms publication types [Data source: WoS, Search date: 23 April 2024]

It was found that 308 papers integrated the use of ML and MCDM techniques. According to the results, there were relatively few studies integrating ML and MCDM approaches between 2000 and 2017. However, since 2018, there has been a notable increase in publications. A significant increase was observed between 2020-2023. As can be seen from Figure 5 in the study, the data used for 2024 only includes values until April 23, 2024, and the amount it will reach at the end of the year is unknown. This increase in hybrid methods suggests that the purpose and timing of this study is also important in terms of providing a perspective for researchers.

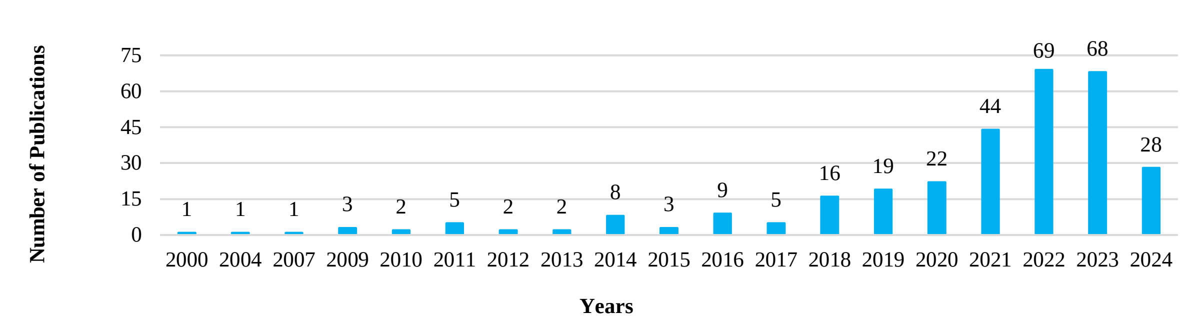

Figure 5. The number of publications from 2000 to 2024 that used both ML and MCDM techniques [Data source: WoS, Search date: 23 April 2024]

The top 10 most cited references are displayed in Table 4. Khosravi et. al [25] study on flood susceptibility related to "Engineering" is the most cited study with 367 citations. Furthermore, when all 10 studies are taken into consideration, the field of "Environmental Sciences and Ecology" has the highest number of research areas, accounting for 4 out of 10 studies. In terms of the model and category classification detailed in Table 1 and Table 2, a review of the most cited studies reveals that model 4, which compares the use of ML and MCDM methods, stands out as a purpose-oriented model. In contrast, model 5, which has never been used due to its complexity. Additionally, model 2, which uses ML and MCDM methods sequentially, has never been used, and model 3 was rarely used. Although there is a balanced distribution in the number-oriented categories, category 4, in which ML and MCDM methods are used more than once, has never been employed.

Table 4. Top 10 most cited references [Data source: WoS, Search date: 23 April 2024]

Titles	Authors	Citations	Year	Research Areas	Model	Category
-A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods	Khosravi et. al [25]	367	2019	Engineering	4	2
-Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests	Aertsen et. al [26]	309	2010	Environmental Sciences & Ecology	3	3
-Understanding commuting patterns using transit smart card data	Ma et. al [27]	236	2017	Business & Economics	1	1
-Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory	Nachappa et. al [28]	164	2020	Engineering; Geology	4	2
-GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naive Bayes tree, bivariate statistics and logistic regression: A case of Topla basin, Slovakia	Ali et. al [29]	156	2020	Biodiversity & Conservation	4	2
-Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques	Costache et. al [30]	139	2019	Environmental Sciences & Ecology	4	3
-GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches	Arabameri et. al [31]	136	2019	Environmental Sciences & Ecology	4	2
-Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis	Rahman et. al [32]	131	2019	Environmental Sciences & Ecology	4	3
-An incident information management framework based on data integration, data mining, and multi-criteria decision making	Peng et. al [33]	120	2011	Computer Science	1	1
-Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels	Ahani et. al [34]	119	2019	Business & Economics	1	1

Figure 6 illustrates the journals with the highest frequency of published ML and MCDM studies. It demonstrates that the majority of articles published in this field were published in “Expert Systems with Applications”. After that, "Applied Soft Computing" was the journal with the second highest number of publications. In addition, it was observed that the other journals have a similar number of publications and focus on practical applications.

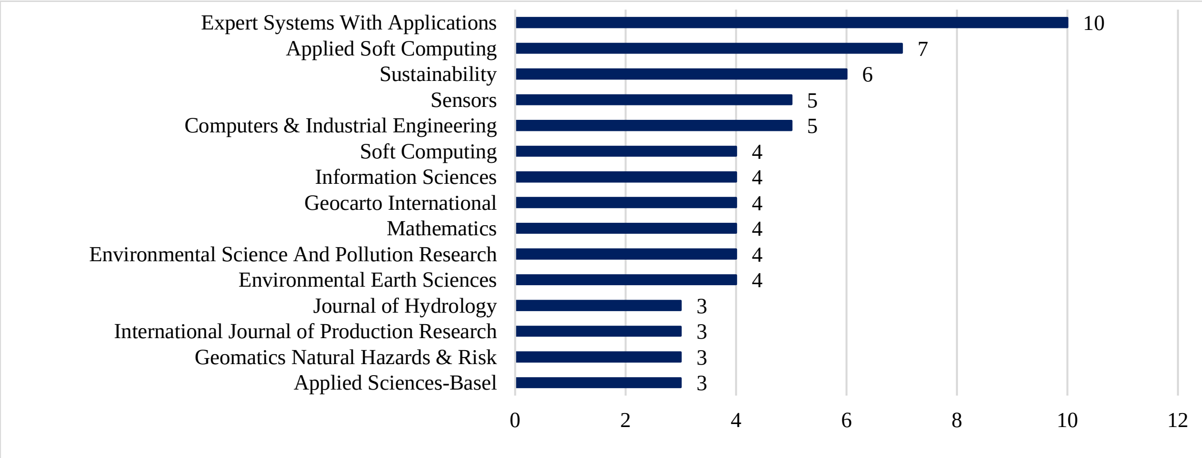

Figure 6. The top 15 journals that feature the combined use of ML and MCDM methods from 2000 to 2024 [Data source: WoS, Search date: 23 April 2024]

Figure 7 shows the top 10 research areas that have used ML and MCDM together in the last decade. According to the graph, 112 papers were published in "Computer Science", which ranked first. This was followed by "Engineering" in second place with 39 publications, and "Environmental Sciences & Ecology" in third place with 32 publications. Finally, "Operations Research & Management Science" ranked last with only four publications. Since ML is more closely related to computer science, it is expected that computer science would be the first among these research fields.

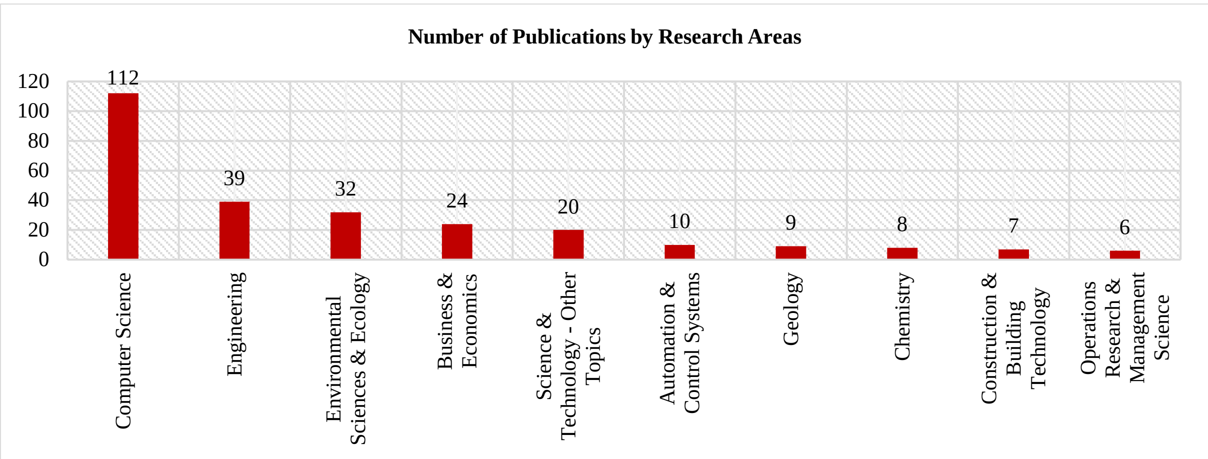

Figure 7. The top 10 research areas that feature the combined use of ML and MCDM methods from 2000 to 2024. [Data source: WoS, Search date: 23 April 2024]

To this point, we have presented results from the generic part of bibliometric analysis regarding publication types, changes over the years, research areas, and journal. Now, the results are presented according to models and categories with reference to the novel step of the research. Figure 8 illustrates that Model 1 and Model 3 were the most used models among the 308 publications, with a percentage of 28%. Model 2 is in second place with 22%, while Model 5 is the least used with only 6%. The fact that Model 5 is the least used can be explained by the fact that this approach requires more challenges in the implementation phases compared to other models.

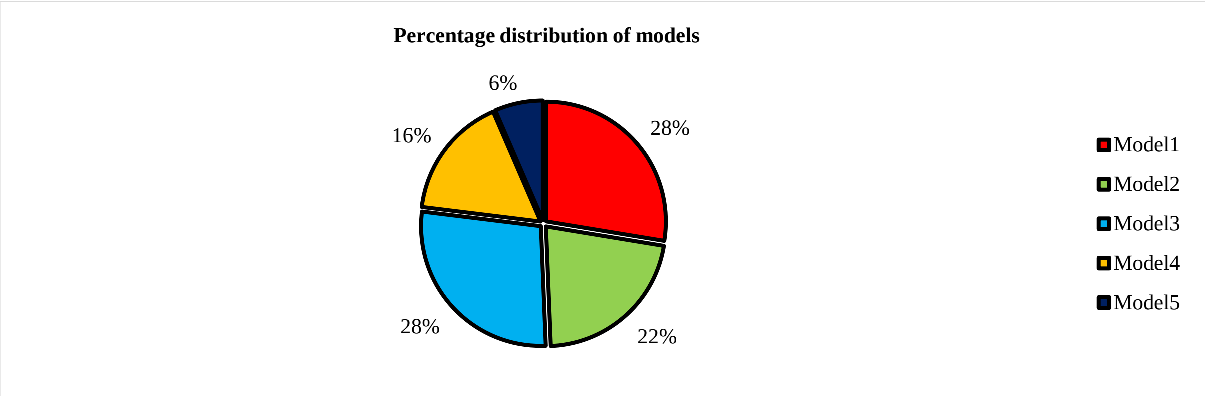

Figure 8. Percentage distribution of models from 2000 to 2024 [Data source: WoS, Search date: 23 April 2024]

Figure 9 shows the most common research areas where ML and MCDM methods are used together. Looking at the models applied in these domains, Model 1, Model 2, and Model 3 are preferred in all research fields. Notably, Model 4 was not used in the Business & Economics and Science & Technology- Other Topics research areas. Examining these five fields as given in Figure 9, the number of times the models are used varies. It can be said that all the models used in "Engineering" are distributed in a balanced way. In Computer Science, all models were used more than the models used in other research fields. It was observed that researchers clearly use Model 3 the most.

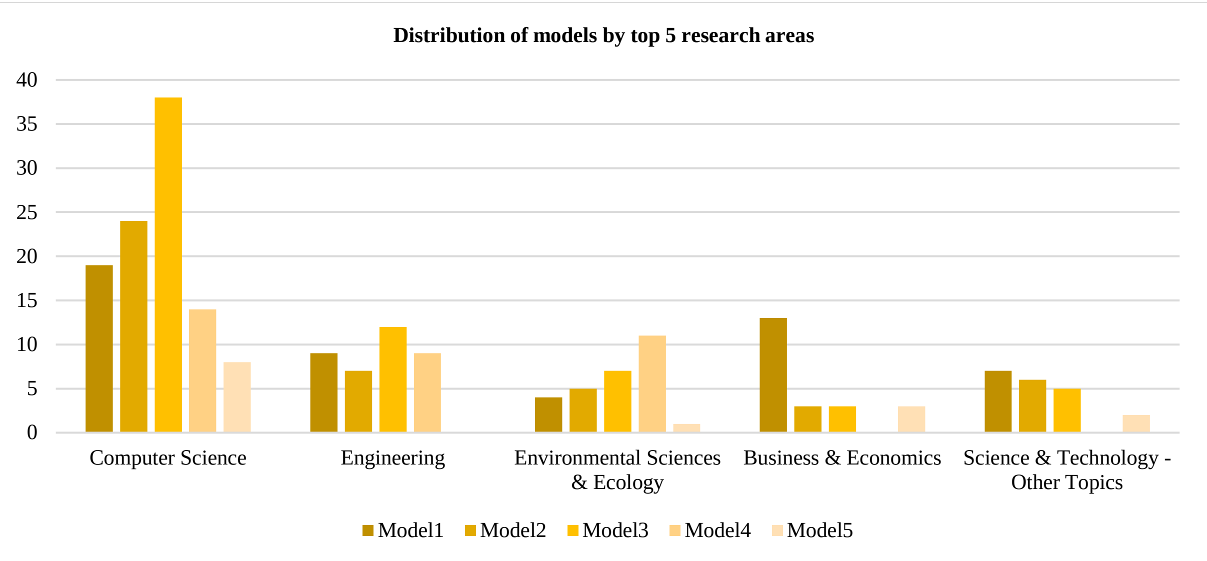

Figure 9. Distribution of models used from 2000 to 2024 by top 5 research areas [Data source: WoS, Search date: 23 April 2024]

The distribution of proposed models over the last decade is shown graphically in Figure 10. Although some models were not used in certain years before 2018, we found that all models were used every year after 2018. After 2020, there is a notable increase in the use of Models 1. In addition, the use of Model 3 increases significantly starting from 2021. The output for the year 2022 shows that, with the exception of Model 5, the number of uses of the other models is close to each other. The low usage of Model 5 is due to its complexity as well as its recent popularity due to increased interest from researchers. Model 3 is clearly more prominent in 2024 compared to other models.

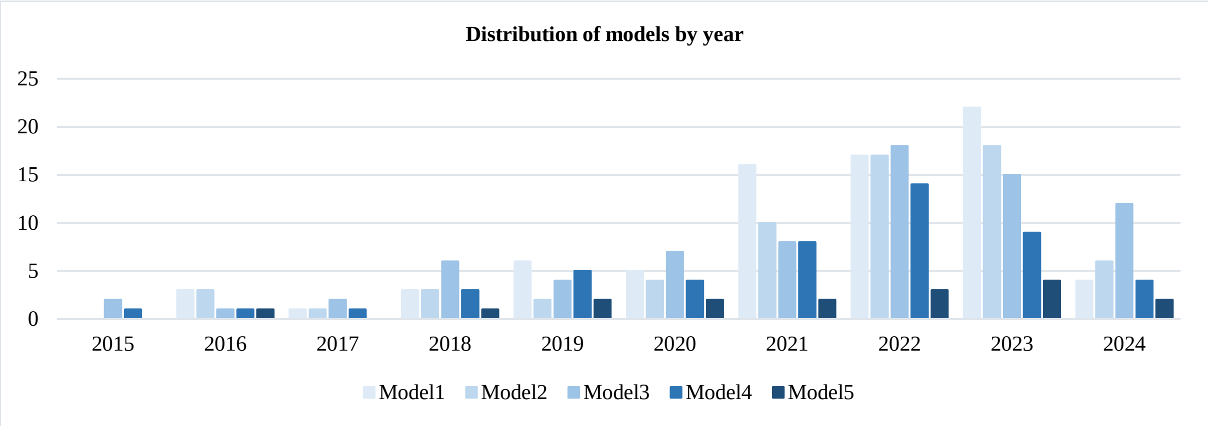

Figure 10. Distribution of models in the last decade [Data source: WoS, Search date: 23 April 2024]

Within the scope of bibliometric analysis, inferences have been made on models up to this point. After that, the analyses related to the categories will be explained in the rest of the study. Figure 11 illustrates that Category 2 was the most used model among the 308 publications, with a percentage of 37%. Category 4 was in second place with 25%, while Category 3 was the least used with only 14%. Considering that the traditional use of ML methods is greater than 1, it is expected that categories 2 and 4 are overused and Category 3 is underused. The proportions of categories where both methods were used once or more (Category 1 and Category 4) are close and balanced.

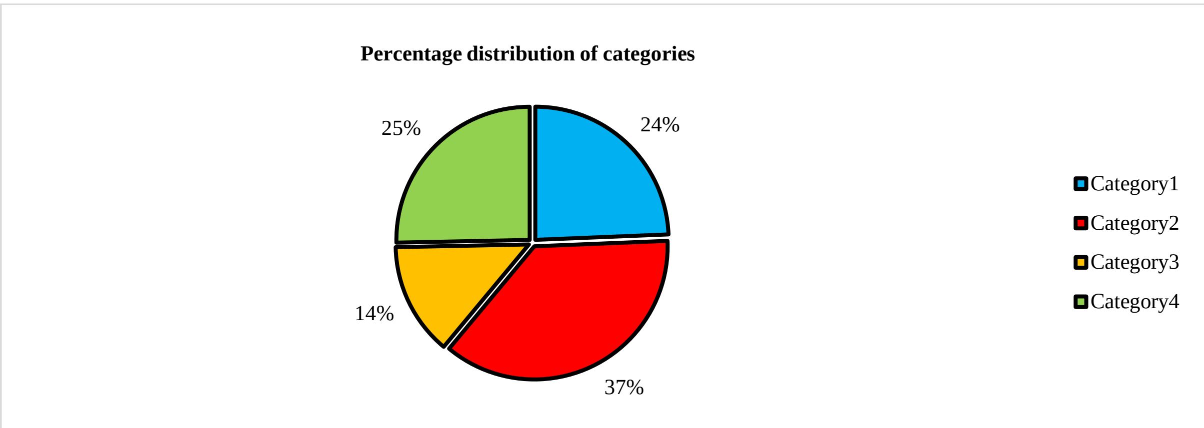

Figure 11. Percentage distribution of categories from 2000 to 2024 [Data source: WoS, Search date: 23 April 2024]

Figure 12 shows the most common research areas where ML and MCDM methods are used together. The graph shows that all categories were used in all research fields. Category 2 was the most preferred category in "Computer Science", while it was the least preferred category in "Business & Economics". There are differences in the distribution of categories by field. In particular, the use of Category 2 is prominent in "Computer Science" and "Environmental Science & Ecology", while Category 4 is prominent in "Engineering". The distribution of categories across research fields indicates an unbalanced distribution.

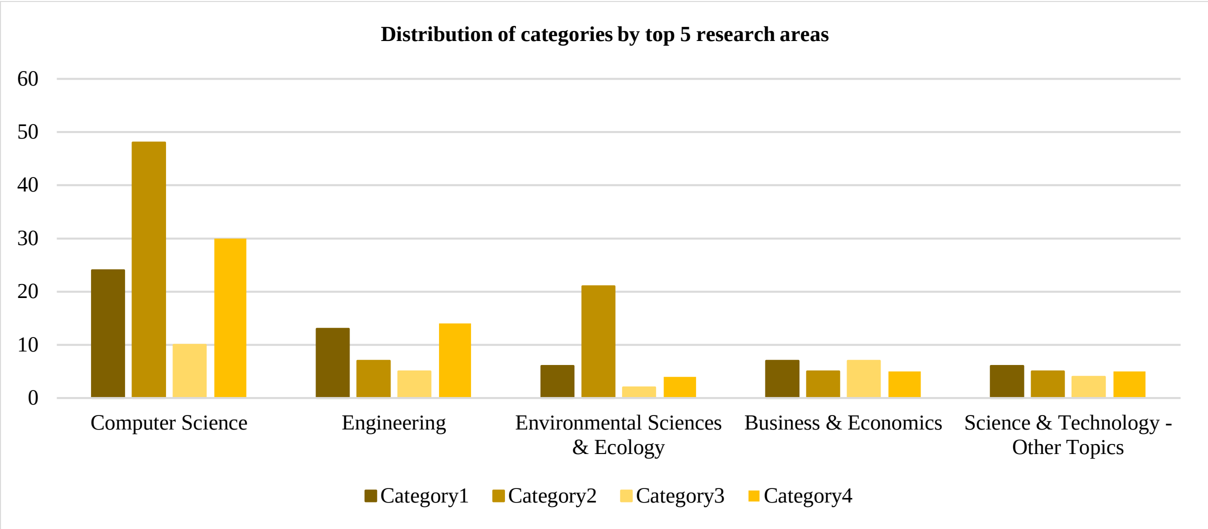

Figure 12. Distribution of categories used from 2000 to 2024 by top 5 research areas [Data source: WoS, Search date: 23 April 2024]

The distribution of categories used in research over the past decade is shown in Figure 13. Although some categories were not used in certain years prior to 2018, we found that all categories are used every year after 2018. In 2019 and beyond, there is a steady increase in the use of Category 2, while there are significant increases in the use of other categories. In Category 3, the data indicates a consistent and gradual increase between 2018 and 2023. However, the data used for 2024 only includes values until April 23, 2024, the numerical representation lags significantly behind the other categories.

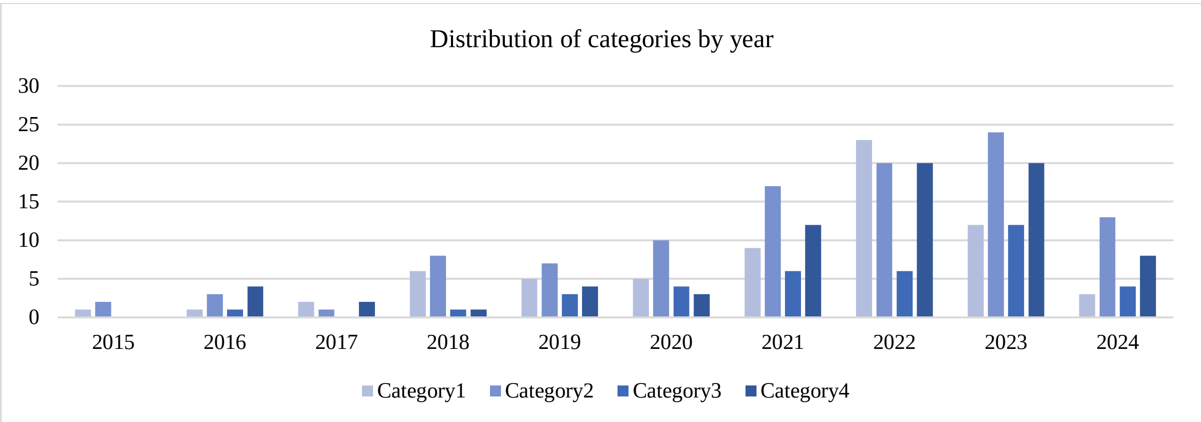

Figure 13. Distribution of categories in the last decade [Data source: WoS, Search date: 23 April 2024]

IV. CONCLUSIONS

In this study, we conducted two step bibliometric analyses: an initial analysis and an extended analysis mentioned in the methodology section for the combination of ML and MCDM methods and evaluated the results. During the initial step of bibliometric analysis, publications were searched in WoS and Scopus databases and VOSviewer was used to visualize the 1189 publications by obtained by query in Figure 1. Keyword analysis was performed using VOSviewer to provide insight into the potential future direction of the study. Accordingly, it was observed that the keywords ML, Data Mining and MCDM were the most frequently used in the publications. Since 2014, the topic of Data Mining was prominent in the early years then ML and MCDM have become prominent, while the interest in Deep Learning has increased in recent years as expected. This can be attributed to the increasing interest in artificial intelligence and the widespread adoption of deep learning, particularly in recent years. After completing the processes of identification, screening, eligibility, and inclusion, an extended bibliometric analysis on 308 publications were conducted. This novel approach, including purpose-oriented models and numerical-oriented categories, distinguishes our research from other bibliometric studies in the literature. The use of a combination of ML and MCDM methods has increased significantly over the years from 2020 to 2023 as a general result of this analysis. This increase in the usage of hybrid methods underscores the significance of the study's purpose in terms of providing a viewpoint for researchers. According to the highest frequency of publication, it was found that the journal "Expert Systems with Applications", which is particularly popular in the field of decision making, had the highest number of publications. Furthermore, the other journals have approximately the similar number of publications and are oriented towards the field of application. In terms of research field, the field of "Computer Science" was observed to have the highest number of studies, which is to be expected given the recent prominence of complex techniques such as deep learning. It is concluded from the research results that researchers aiming to integrate ML and MCDM methods should design their research in a way that is compatible with the field of computer science and include researchers from this field in their research teams when necessary.

With respect to the purpose-oriented models, Models 1 and 3 were generally the most frequently used, while Model 5 was the least frequently used, although it has steadily increased over the years. This discrepancy in model usage can be attributed to differences in implementation complexity. Despite the scarcity of Model 5 applications among researchers, the growth of model 5 over the years highlights the potential for further research in the application of this model. Moreover, respecting the numerical-oriented categories, Category 2 was used the most, while Category 3 was used the least, which seems to be related to how the researchers construct their research objectives. In addition, the fact that there is no distinct distribution of categories by research area and lack of observable patterns over the years implies that researchers may use all categories with various approaches in each area. With respect to the purpose-driven models, Models 1 and 3 were generally the most frequently used, while Model 5 was the least frequently used, although it has steadily increased over the years. This discrepancy in model usage can be attributed to differences in implementation complexity. Although the scarcity of Model 5 applications among researchers, the growth of model 5 over the years highlights the potential for further research in the application of this model. Furthermore, respecting the numerical-oriented categories, Category 2 was used the most, while Category 3 was used the least, which seems to be related to how the researchers construct their research objectives. Additionally, the fact that there is no distinct distribution of categories by research area and lack of observable patterns over the years implies that researchers may use all categories with various approaches in each area. In addition to the valuable insights yielded by this study, researchers can conduct association rule analyses to ascertain which machine learning and MCDM methods are utilized in conjunction, thereby obtaining a perspective that extends beyond bibliometric analysis.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their informative remarks and ideas, which assisted in advancing our research.

REFERENCES

1. Alpaydin E (2014) Introduction to machine learning. The MIT Press, Cambridge
2. Sandeep MS, Tiprak K, Kaewunruen S, Pheinsusom P, Pansuk W (2023) Shear strength prediction of reinforced concrete beams using machine learning. Structures 47:1196–1211. https://doi.org/10.1016/j.istruc.2022.11.140
3. Kou G, Wu W (2014) An analytic hierarchy model for classification algorithms selection in credit risk analysis. Math Prob in Eng. https://doi.org/10.1155/2014/297563
4. Ali R, Lee S, Chung TC (2017) Accurate multi-criteria decision making methodology for recommending machine learning algorithm. Exp Syst with Appl 71,257–278. https://doi.org/10.1016/j.eswa.2016.11.034
5. Golmohammadi D, Zhao L, Dreyfus D (2023) Using machine learning techniques to reduce uncertainty for outpatient appointment scheduling practices in outpatient clinics. Omega 120:102907. https://doi.org/10.1016/j.omega.2023.102907
6. Taherdoost H, Madanchian M (2023) Multi-criteria decision making (MCDM) methods and concepts. Encyclopedia 3(1):77–87. https://doi.org/10.3390/encyclopedia3010006
7. Chowdhury NK, Kabir MA, Rahman MM, Islam SMS (2022). Machine learning for detecting COVID-19 from cough sounds: An ensemble-based MCDM method. Comp in Bio and Med 145:105405. https://doi.org/10.1016/j.compbiomed.2022.105405
8. Yilmaz,I, Adem A, Dağdeviren M (2023) A machine learning-integrated multi-criteria decision-making approach based on consensus for selection of energy storage locations. J of Energy Storage 69:107941. https://doi.org/10.1016/j.est.2023.107941
9. Choudhary S, Pingale SM, Khare D (2022) Delineation of groundwater potential zones of upper Godavari sub-basin of India using bi-variate, MCDM and advanced machine learning algorithms. Geo Inter 37(27): 15063–15093. https://doi.org/10.1080/10106049.2022.2093992
10. Mustapha MT, Ozsahin DU, Ozsahin I, Uzun B (2022) Breast cancer screening based on supervised learning and multi-criteria decision-making. Diagnostics 12(6), 1326. https://doi.org/10.3390/diagnostics12061326
11. Nachappa TG, Piralilou ST, Gholamnia K, Ghorbanzadeh O, Rahmati O, Blaschke T (2020) Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using dempster shafer theory. J of Hydro, 590, 125275. Elsevier B.V. https://doi.org/10.1016/j.jhydrol.2020.125275
12. Viviani M, Pasi G (2017) A multi-criteria decision making approach for the assessment of information credibility in social media. 11th International Workshop, Naples, Italy, Dec. 19-21.
13. Çalık E (2021) İşgören seçiminde makine öğrenmesi ve çok kriterli karar verme yöntemlerinin birlikte kullanımının değerlendirilmesi. J of Turk Stud 16:1483–1494. https://doi.org/10.7827/turkishstudies.51593
14. Liao H, He Y, Wu X, Wu Z, Bausys R (2023) Reimagining multi-criterion decision making by data-driven methods based on machine learning: A literature review. Information Fusion. https://doi.org/10.1016/j.inffus.2023.101970
15. Andres C, Verastegui A, Elizabet N, Gilvonio C, Esthefani M, Flores R (2023) A bibliometrics study of plants, animals, bacteria, algae and technologies that reduce, filter and eliminate microplastics from planet earth, ecological solutions for the environment. Deci Sci Letters 12:773–782. https://doi.org/10.5267/dsl.2023.6.004
16. Fernández JMM, Moreno JJG, González EPV, Iglesias GA (2022) Bibliometric analysis of the application of artificial intelligence techniques to the management of innovation projects. Appl Sci 12(22). https://doi.org/10.3390/app122211743
17. Pritchard A (1969) Statistical bibliography or bibliometrics? J of Doc. 25(4):348.
18. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM (2021) How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research 133:285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

19. Gao C, Sun M, Geng Y, Wu R, Chen W (2016) A bibliometric analysis based review on wind power price. Applied Energy 182:602–612. https://doi.org/10.1016/j.apenergy.2016.08.144
20. Ellegaard O, Wallin JA (2015) The bibliometric analysis of scholarly production: How great is the impact? Scientometrics 105(3):1809–1831. https://doi.org/10.1007/s11192-015-1645-z
21. Demir G, Chatterjee P, Pamucar D (2024) Sensitivity analysis in multi-criteria decision making: A state-of-the-art research perspective using bibliometric analysis. Exp Syst with Appl 237, 121660. https://doi.org/10.1016/j.eswa.2023.121660
22. Riahi Y, Saikouk T, Gunasekaran A, Badraoui I (2021) Artificial intelligence applications in supply chain: A descriptive bibliometric analysis and future research directions. Exp Syst with Appl 173, 114702. https://doi.org/10.1016/j.eswa.2021.114702
23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372. https://doi.org/10.1136/bmj.n71
24. Xu Z, Ge Z, Wang X, Skare M (2021) Bibliometric analysis of technology adoption literature published from 1997 to 2020. Tech For and Soc Cha 170. https://doi.org/10.1016/j.techfore.2021.120896
25. Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Prakash I (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. J of Hydro 573:311-323. https://doi.org/10.1016/j.jhydrol.2019.03.073
26. Aertsen W, Kint V, Orshoven JV, Özkan K, Muys B (2010) Comparison and ranking of different modelling techniques for prediction of site index in mediterranean mountain forests. Eco Model 221(8):1119-1130. https://doi.org/10.1016/j.ecolmodel.2010.01.007
27. Ma X, Liu C, Wen H, Wang Y, Wu YJ (2017) Understanding commuting patterns using transit smart card data. J of Trans Geo 58:135-145. https://doi.org/10.7307/ptt.v32i1.3052
28. Nachappa TG, Piralilou ST, Gholamnia K, Ghorbanzadeh O, Rahmati O, Blaschke T (2020) Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using dempster shafer theory. J of Hydro 590:125275. https://doi.org/10.1016/j.jhydrol.2020.125275
29. Ali SA, Parvin F, Pham QB, Vojtek M, Vojteková J, Costache R, Ghorbani MA (2020) GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: a case of Topľa basin, Slovakia. Eco Ind 117(106620). https://doi.org/10.1016/j.ecolind.2020.106620
30. Costache R, Pham QB, Sharifi E, Linh NTT, Abba SI, Vojtek M, Khoi, DN (2019) Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques. Remote Sensing 12(1):106. https://doi.org/10.3390/rs12010106
31. Arabameri A, Rezaei K, Cerda A, Lombardo L, Comino JR(2019) GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. Sci of the Total Env 658:160-177. https://doi.org/10.1016/j.scitotenv.2018.12.115
32. Rahman M, Ningsheng C, Islam MM, Dewan A, Iqbal J, Washakh RMA, Shufeng T (2019) Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis. Earth Syst and Env 3:585-601. https://doi.org/10.1007/s41748-019-00123-y
33. Peng Y, Zhang Y, Tang Y, Li S (2011) An incident information management framework based on data integration, data mining, and multi-criteria decision making. Deci Supp Syst 51(2):316-327. https://doi.org/10.1016/j.dss.2010.11.025
34. Ahani A, Nilashi M, Yadegaridehkordi E, Sanzogni L, Tarik AR, Knox K, Ibrahim O (2019) Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. J of Ret and Cons Serv 51:331-343. https://doi.org/10.1016/j.jretconser.2019.06.014