Received: 17 February 2022 | Revised: 4 July 2022 | Accepted: 14 July 2022

DOI: 10.1002/widm.1476

OVERVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY

WILEY

On the application of machine learning in astronomy and astrophysics: A text-mining-based scientometric analysis

José-Víctor Rodríguez[1,2] | Ignacio Rodríguez-Rodríguez[3] | Wai Lok Woo[4]

[1]Departamento de Tecnologías de la Información y las Comunicaciones, Universidad Politécnica de Cartagena, Cartagena, Spain

[2]Departamento de Física Teórica y del Cosmos, Universidad de Granada, Granada, Spain

[3]Departamento de Ingeniería de Comunicaciones, BioSIP Group, Universidad de Málaga, Málaga, Spain

[4]Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, UK

Correspondence
José-Víctor Rodríguez, Universidad Politécnica de Cartagena, Departamento de Tecnologías de la Información y las Comunicaciones, E30202, Cartagena, Spain and Universidad de Granada, Departamento de Física Teórica y del Cosmos, E18071, Granada, Spain.
Email: jvictor.rodriguez@upct.es/ jvrodriguez@correo.ugr.es

Edited by: Ryszard Kowalczyk, Associate Editor and Witold Pedrycz, Editor-in-Chief

Abstract

Since the beginning of the 21st century, the fields of astronomy and astrophysics have experienced significant growth at observational and computational levels, leading to the acquisition of increasingly huge volumes of data. In order to process this vast quantity of information, artificial intelligence (AI) techniques are being combined with data mining to detect patterns with the aim of modeling, classifying or predicting the behavior of certain astronomical phenomena or objects. Parallel to the exponential development of the aforementioned techniques, the scientific output related to the application of AI and machine learning (ML) in astronomy and astrophysics has also experienced considerable growth in recent years. Therefore, the increasingly abundant articles make it difficult to monitor this field in terms of which research topics are the most prolific or novel, or which countries or authors are leading them. In this article, a text-mining-based scientometric analysis of scientific documents published over the last three decades on the application of AI and ML in the fields of astronomy and astrophysics is presented. The VOSviewer software and data from the Web of Science (WoS) are used to elucidate the evolution of publications in this research field, their distribution by country (including co-authorship), the most relevant topics addressed, and the most cited elements and most significant co-citations according to publication source and authorship. The obtained results demonstrate how application of AI/ML to the fields of astronomy/astrophysics represents an established and rapidly growing field of research that is crucial to obtaining scientific understanding of the universe.

This article is categorized under:

- Algorithmic Development > Text Mining
- Technologies > Machine Learning
- Application Areas > Science and Technology

KEYWORDS

astronomy, astrophysics, machine learning, scientometrics, text mining

1 | INTRODUCTION

Astronomy and astrophysics rely on the acquisition, storage, and analysis of large datasets. Regardless of whether these are obtained through telescopes, particle detectors, or numerical simulations, they are processed in a way that allows scientific knowledge about the universe to be acquired. Thanks to advances in optical and computer technology, there has been a dramatic growth in the observational and computational aspects of astronomy and astrophysics over the last two decades. This has resulted in extraordinarily large datasets (currently approaching the exabyte scale) that are necessitating new methods of analysis to facilitate scientific discovery (Fayyad et al., 1996), giving rise to the era of so-called Big Data (Brunner et al., 2002; Feigelson & Babu, 2012; Khalil et al., 2019; Zhang & Zhao, 2015). One of these methods is the use of artificial intelligence (AI), which, together with data mining, allows researchers to discern patterns in the data and thus model, classify or predict the behavior of astronomical phenomena and objects (Ball & Brunner, 2010; Borne, 2009; Graham et al., 2013; Ivezic et al., 2014; Wang et al., 2018).

Within AI, the branch of machine learning (ML) has been of extraordinary interest in the astronomical/astrophysical field. ML is the process in which a machine (computer) autonomously learns to identify patterns to make classifications or predictions without having been expressly programmed to do so. Various ML techniques can be applied to astronomy and astrophysics, producing extremely diverse results, including principal component analysis (PCA) for the classification of the spectra of stars (Singh et al., 1998; Xiang et al., 2017), exoplanets (Damiano et al., 2019), galaxies (Conselice, 2006; Kochi et al., 2017), and quasars (Davies et al., 2018; Yip et al., 2004); random forests (RF) for the estimation of photometric redshifts (Carrasco Kind & Brunner, 2013; Mucesh et al., 2021) and the classification of asteroids (Chao et al., 2017); decision trees (DT) for the classification of stellar (Hinners et al., 2017) and galaxy morphologies (Reza, 2021); support vector machines (SVM) for the prediction of solar flares (Qahwaji & Colak, 2007), the detection of gravitational waves (Biswas et al., 2013) and the classification of astronomical objects (Sunitha et al., 2020); artificial neural networks (ANN), which constitute the basis of the ML subfield called deep learning (DL, i.e., structured or hierarchical automatic algorithms that emulate human learning), for the detection of gamma ray bursts (Ball et al., 2004; Singh et al., 2022), the analysis of asteroid composition (de León et al., 2010), pulsar identification (Eatough et al., 2010) and the detection of gravitational lenses (Agnello et al., 2015); and, finally, convolutional/deep neural networks (CNN/DNN), which were developed from ANN, for the classification of transit objects (Gómez et al., 2020) and galaxies (Dieleman et al., 2015), ~~photometric redshifts estimation (Hoyle, 2016)~~, and the identification of the astronomical sources of radio waves (Lukic et al., 2020).

In line with these increasingly powerful techniques, there has been a plethora of scientific output on the topic of AI and ML as applied in astronomy and astrophysics. This growing pool of scientific publications introduces difficulties in the identification of the most novel contributions or prolific research topics or which authors or countries are at the forefront of this research field. A solution to this issue may be offered by scientometrics, a discipline that uses different techniques and tools to study the scientific output on a topic by measuring and analyzing it. A scientometric study can provide a macroscopic view of an extensive set of scientific publications, with methods such as text mining allowing the production of relational maps related to the scientific development of a given line of research. This makes it possible to identify patterns in terms of the most productive scientific journals, the countries with the highest research output or the most prolific authors, and the topics on which recent research has been focusing.

In this respect, despite the recent publication of several very comprehensive review articles on the use of AI/ML techniques in the field of astronomy/astrophysics (Fluke & Jacobs, 2020; Kremer et al., 2017; Longo et al., 2019; Wang et al., 2018), to the best of the authors' knowledge, there is no updated analysis that uses scientometrics to address this issue.

Consequently, this work comprises a text-mining-based scientometric study of the scientific output in the last three decades regarding the application of AI and ML to the fields of astronomy and astrophysics.

2 | MATERIALS AND METHODS

2.1 | Scientometrics and text-mining

Scientometrics concerns the quantitative aspects of scientific information in relation to its production, dissemination and utilization to thoroughly understand the mechanisms and evolutionary drivers underlying scientific research (Chellappandi & Vijayakumar, 2018). As a research technique, it is located in the information and library sciences and employs quantitative methods, for example, text mining, to explore bibliographic data, such as author, publication year, and country of origin (Broadus, 1987). A scientometric analysis can be wielded to produce a representative outline that describes a collection of scientific documents. Prior studies have used scientometrics tools to explore myriad aspects of the scientific literature, from keywords and topics (Thongpapanl, 2012) to authors, institutions, and countries. This information can be evaluated using a wide set of indicators, such as the number of citations or the volume of published articles (Song & Ding, 2014). The current study draws on some of these indicators to present a holistic perspective of the technique, enabling readers to gain an insight into the possible outcomes thanks to its unique advantages as well as the implementation needs.

As yet, there is no consensus on how to best measure the virtues of scientific literature. In general, productivity and influence are the two common ways with which to measure research output (Podsakoff et al., 2008); the former is usually based on the total number of publications, while the latter is captured by the number of citations. However, several other useful scientometric indicators have also been established, often with remarkable results.

2.2 | Software

In the current study, the Java-based VOSviewer software (version 1.6.17), developed at Leiden University, the Netherlands, is utilized to perform the scientometric mapping and text mining (Van Eck & Waltman, 2014). VOSviewer first mines texts to identify the keywords used in a set of research articles, and then generates bibliometric maps, referred to as landscapes, via the visualization of similarities (VOS) mapping technique (Van Eck & Waltman, 2010). By depicting these landscapes in different ways, inferences can be made on the publications' characteristics.

Briefly, using bibliographic data to create graphical maps, VOSviewer can provide a representation of serviceable characteristics, including bibliographic coupling (Kessler, 1963), co-citation (Small, 1973), and co-authorship and the co-occurrence of author keywords.

VOSviewer implements the so-called Leiden algorithm. Notably, the use of the Leiden algorithm to find well-connected clusters in a network means that this technique addresses some of the deficiencies inherent in the Louvain algorithm. In this sense, the Louvain method allows researchers to identify communities within networks. For each community, the algorithm maximizes a modularity score that quantifies how well nodes are assigned to the communities. In other words, it compares the densities of the connections between a community's nodes with those for the same nodes in a random network. As a hierarchical clustering algorithm, the Louvain algorithm brings identified communities into a single node through recursive merging, performing modularity clustering on the condensed graphs (Blondel et al., 2008). Specifically, the Louvain method can identify clusters that have arbitrary connections. Nonetheless, the communities detected by the Leiden algorithm have guaranteed connections. Moreover, the iterative use of the Leiden algorithm enables convergence on a partition wherein all population subsets are optimally locally assigned (Traag et al., 2019).

The abovementioned VOS algorithm used by the VOSviewer for the graph layout was developed by authors Van Eck and Waltman (2007). The low-dimensional visualization provided by VOS provides the highest possible degree of accuracy as objects are located so that the distance between any given object pair well represents their similarity. VOS grounds upon the notion of minimizing the weighted sums of all object pairs' squared Euclidean distances. During the summation, the greater a pair's similarity, the greater their squared distance weight.

2.3 | Indicators

We focused on the indicators outlined in the following to summarize the provenance of the application of AI/ML to the astronomy and astrophysics research fields.

2.3.1 | Production and chronology

The first analyzed characteristic is the quantity of scientific material produced regarding the investigated topic, that is, the application of AI/ML to astronomy and astrophysics. This reveals the most productive countries in this field, which can subsequently be compared at the level of the gross domestic product (GDP). Furthermore, by clarifying how this stream of research has occurred chronologically, we can then estimate whether this can be considered a *hot topic*.

2.3.2 | Topics analysis

Topic mapping analysis unveils the main (occasionally unseen) topics that are mooted by the bibliographic materials in the investigated research field. Hence, through the application of a factual methodology, this analysis allows so-called idle topics (i.e., subtle or obscure) lying within the vast quantity of bibliographic material to be transformed into clear-cut visual illustrations of the subject clusters and their correlations. This tool has already shown great potential in the fields of text mining and scientometrics (Yan et al., 2012). It relies on dissimilarities between probability distributions, that is, how both an individual semantic element and all semantic elements are, respectively, distributed across the group of all topics (Van Eck et al., 2010). If these distributions are highly disparate, this indicates that the semantic element probably links to a specific concept, whereas equivalent distributions suggest that the semantic element is not coupled in this way. The relationships between the keywords are computed according to how often they appear in the articles; in other words, the co-occurrence of two terms in numerous articles proposes that these have a robust relationship. The results of such an analysis allow the grouping of the terms into clusters on a map (Van Eck & Waltman, 2007) via the abovementioned VOS technique.

2.3.3 | Citations and highly cited elements

Examining which articles, authors and journals have the most citations can reveal the key research elements in the application of AI/ML to astronomy and astrophysics.

2.3.4 | Co-citation analysis

In a co-citation analysis, weights are applied corresponding to the co-citation strength, referring to the tendency for certain journals or authors to be co-cited in one article. This bases on the understanding that articles authored by frequently co-referenced researchers tend to deal with corresponding ideas (Boyack & Klavans, 2010). Utilizing the clustering technique of Van Eck (Waltman & Van Eck, 2012), VOSviewer generates a co-citation matrix that depicts the clusters of closely related articles. At least 20 citations are required for a co-citation analysis in order to ensure that the data visualization contains minimal disorder. This work also employed lower thresholds (5, 10, and 15 citations) to identify the optimum level. Co-citation analysis can be conducted at the author level, wherein the co-citation value is calculated according to the correlations between a given author's articles, or the journal level, using the relationships among journals. The co-citation structures unveiled through this technique offer an in-depth insight into the application of AI/ML to the fields of astronomy and astrophysics.

2.3.5 | Overlay visualization

Overlay visualization allows other information types, for example, publication year, to be superimposed onto topic analysis or citation analysis, thereby depicting the trend of a research stream or set of collaborative works. This technique can automatically identify trends, making it one of the most important scientometric tools (Leydesdorff & Bornmann, 2012) as it can unequivocally portray the chronological development of a topic. VOSviewer can hereby produce a map illustrating the associations between specific elements, onto which additional datapoints containing further information (e.g., publication age and citation impact) can be superimposed.

2.4 | Data acquisition

2.4.1 | Source of data

The most extensive bibliographic sources for research data are generally considered to be the Web of Science (WoS) and Scopus databases (Zhu & Liu, 2020). The former, which is both multidisciplinary and selective, comprises numerous specialized indexes. Its main component is its Core Collection (WoS CC), constituting six main citation indexes, namely the Science Citation Index Expanded (SCIE), the Arts and Humanities Citation Index (AandHCI), the Social Science Citation Index (SSCI), the Conference Proceedings' Citation Index (CPCI), the Book Citation Index (BKCI), and recently, the Emerging Sources Citation Index (ESCI) (Ahmad & Batcha, 2019). While Scopus is also multidisciplinary (Baas et al., 2020), it additionally integrates content from several more specialized databases, including Embase, Fluidex, Compendex, World Textile Index, Medline, Geobase, and Biobase and Medline (Valderrama-Zurián et al., 2015). Nonetheless, as the first large-scale international source of bibliographic data, WoS has become the most influential and has been the primary database employed in scientometrics (Li et al., 2018).

Google Scholar (GS) is another potential source of bibliographic information (Halevi et al., 2017), and the free availability of its content on a non-subscription basis is one of its main advantages. Its other advantage is its much broader and deeper scope, even though it lacks clearly defined boundaries. While its advantages give GS an edge over WoS and Scopus, they also diminish its reliability as a bibliographic database. GS is also in particular plagued by reduced stability, transparency, and precision (Orduña Malea et al., 2017). Many other useful data sources exist, including the relatively new and access-free ResearchGate, Microsoft Academic, CrossRef, and OpenCitations (Visser et al., 2021); however, they still have questionable validity.

Based on the reasoning above, the WoS database is used in this study identifying research applying AI/ML to the fields of astronomy and astrophysics.

2.4.2 | Collected data

The extraction of data from WoS was performed on November 1, 2021. The search period was 1989 until the data collection date. This way, by considering all publications given by the search, 2723 relevant articles were found.

For each identified publication, numerous data fields were extracted, such as title, author(s), country of origin, keywords, abstract, journal, and publication date. The search criteria, that is, the key terms used in the search, are presented in Table 1, together with the number of intermediate results of each term (in decreasing order of relevance).

As the table shows, the searches linked concepts from the field of AI/ML with the fields of astronomy or astrophysics to identify relevant records. We hereby availed of most general AI/ML concepts as proposed by a recent study (Shnurenko et al., 2020).

Beyond the general application areas, keywords referring to specific applications and algorithms (Bonaccorso, 2018) were also incorporated; however, these occasionally returned insignificant results when considered in isolation. Keywords that returned innumerable scientific articles (e.g., “dataset” or “clustering”) were excluded as these terms did not inevitably signify the relevance of the article to the topic of interest. Some of the more atypical or highly distinct terms

TABLE 1 Search criteria

Keywords		Number of intermediate results
(Astronomy OR Astrophysics) AND	(Machine Learning OR	775
	Neural Networks OR	524
	Big Data OR	397
	Deep Learning OR	238
	Artificial Intelligence OR	190
	Genetic Algorithm OR	142
	Random Forest OR	129
	Cloud Computing OR	119
	Support Vector Machines OR	84
	Feature Selection OR	47
	Decision Trees OR	36
	K Nearest Neighbors OR	21
	Gradient Boosting OR	12
	Naïve Bayes)	9

gave negligible results when used alone; it should further be noted that these results had already been returned in the previous general searches using key terms such as "artificial intelligence," "machine learning," and "big data."

The data produced by the search were first exported into text format; following this, outliers and incomplete records were removed and the records were prepared for import into VOSviewer.

3 | RESULTS AND DISCUSSION

3.1 | Output and chronological analysis

Figure 1 represents the chronological evolution of the number of publications related to the application of AI/ML in astronomy/astrophysics.

As mentioned in Section 1, AI and ML techniques have experienced exponential development in the last few decades, as can be seen in Figure 2, where the total publications regarding the search criteria of Table 1, without the terms "Astronomy" and "Astrophysics," have been depicted.

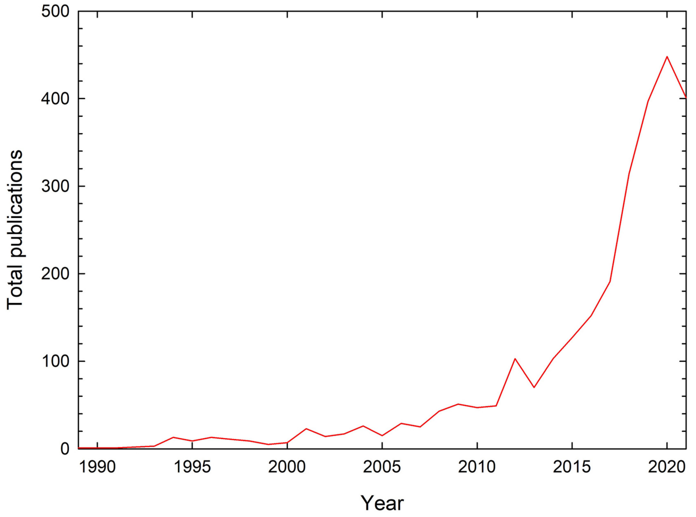

FIGURE 1 Timeline of the number of publications considering the search criteria of Table 1.

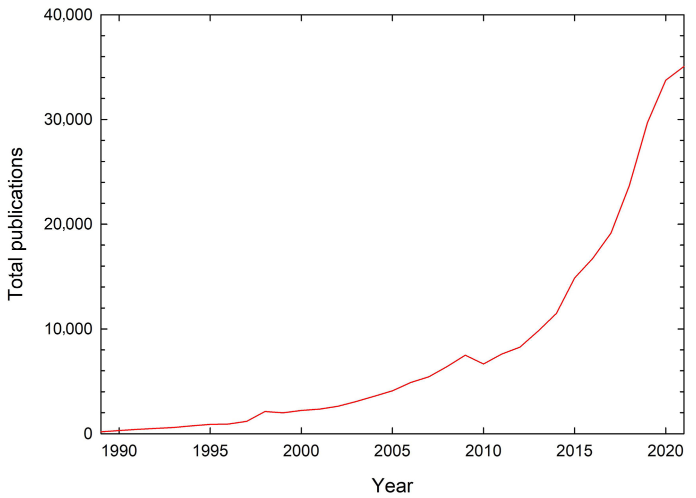

FIGURE 2 Timeline of the number of publications considering the search criteria of Table 1 but without the terms "astronomy" and "astrophysics."

TABLE 2 Top 15 most productive WoS categories

Category	No. of publications
Astronomy Astrophysics	1761
Computer Science Information Systems	253
Optics	236
Engineering Electrical Electronic	232
Computer Science Theory Methods	224
Computer Science Artificial Intelligence	220
Computer Science Interdisciplinary Applications	211
Physics Particles Fields	116
Physics Multidisciplinary	85
Physics Applied	84
Computer Science Software Engineering	65
Computer Science Hardware Architecture	64
Instruments Instrumentation	58
Physics Nuclear	54
Imaging Science Photographic Technology	49

Consequently, as shown in Figure 1, the scientific output related to the application of these techniques in astronomy or astrophysics has also seen dramatic growth (almost proportional), especially after 2005. Specifically, 488 of the 2723 articles found in WoS during the data acquisition stage had been published during 2020; meanwhile, for the year 2021, 401 had been published by the data collection date (November 1, 2021). Furthermore, 1751 articles have been published in just the last 5 years. Meanwhile, 1648 of the 2723 articles were published in open-access journals according to the results from WoS. Unlike the conventional subscription-based model, which requires readers to pay to access scholarly publications, open-access publishing enables research results to be disseminated to a wide audience at no cost to the readers. In this sense, the fact that most of the publications are open access (60%) may indicate that there is a certain trend toward the democratization of knowledge, at least as far as the field of AI/ML applied to astronomy/astrophysics is concerned.

Another interesting analysis that can be performed in relation to the total number of articles published in this line of research is their distribution among the different WoS categories or disciplines. Table 2 lists the 15 categories with the most publications.

As can be seen (and as might be expected), the Astronomy Astrophysics category clearly leads this list, with 1761 publications. However, the table also contains several well-represented fields related to computer science, engineering, and optics. The less productive categories include those related to physics (specific and general) and instrumentation.

The global scientific output in this field according to the number of articles published by each country is depicted as a map in Figure 3.

As Figure 3 clearly shows, scientific research articles related to this topic are prevalently produced in the United States (948), China (483), and the United Kingdom (478), followed by Germany (369). A substantial amount of scientific research on the application of AI/ML to the astronomy and astrophysics fields has also been produced in several other European countries.

Notably, the two countries generating the most research articles on this topic are simultaneously the two largest global economies in terms of GDP, namely the US ($20.94 trillion) and China ($14.72 trillion) (WorldBank, 2021a). To delve deeper into this analysis, Figure 4 classifies the top 15 most productive countries in relation to their incomes.

It emerges that the majority of the 15 countries producing the most publications are high-income nations, as per the World Bank classification (WorldBank, 2021b). The exceptions are the upper-middle-income nations China and Brazil and the lower-middle-income nation India. According to the World Bank classification, the most productive low-income country, regarding the number of publications, is Ethiopia; having published two articles, it achieves a rank of 65.

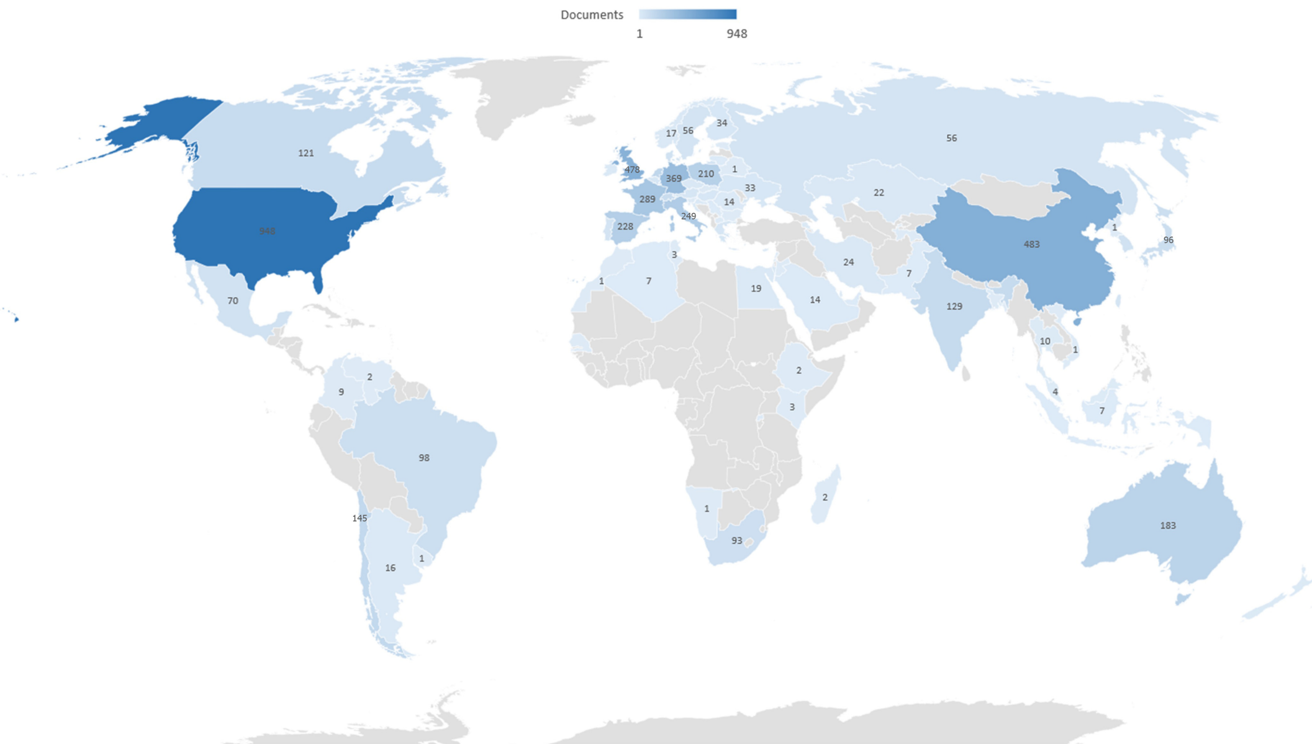

FIGURE 3 Number of publications by country.

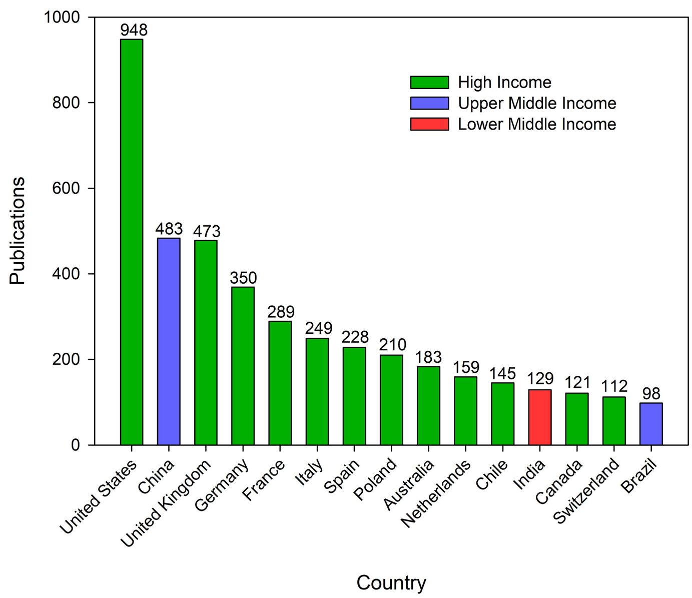

FIGURE 4 Top 15 most productive countries classified according to their incomes.

To compare the data in Figure 4 with the pattern of global scientific production (taking into account all disciplines and subjects), Figure 5 portrays the 15 most productive countries—in terms of published articles—in all scientific fields for the 1996–2020 period (*Source*: Scimago).

In both figures, the first four positions are occupied by the same countries and in the same order. However, while Japan is in fifth place for the scientific production for all categories, it is not among the first 15 for AI/ML in astronomy/astrophysics publications. Russia and South Korea are similarly absent from Figure 4. On the other hand, European countries such as France, Italy, Spain, and the Netherlands appear in both figures, and in the same order. It is worth highlighting that in Figure 5, India occupies a more advanced place than it does in Figure 4, especially given its World Bank classification as a low-income country. Finally, it should be noted that while Chile is not among the 15 countries with the most scientific output, it does appear among the 15 most prolific in the field of AI/ML applied to astronomy/astrophysics. This fact could be explained by the historical astronomical tradition that it enjoys for hosting one of the best places on Earth to observe the sky—the Atacama Desert.

Following the examination of the evolution of scientific output, including its distribution by country, the following presents an analysis of collaborations in the publishing of work related to the application of AI/ML to astronomy/

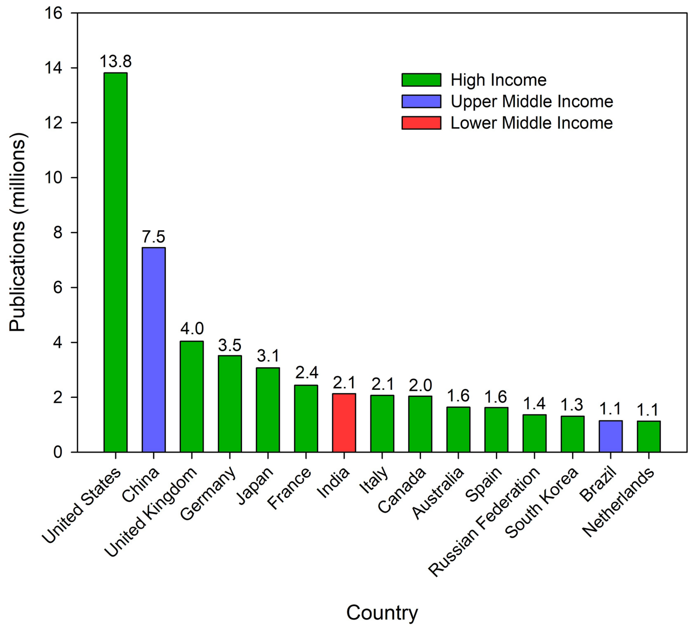

FIGURE 5 Top 15 most productive countries in all scientific fields for the 1996–2020 period (*Source:* Scimago).

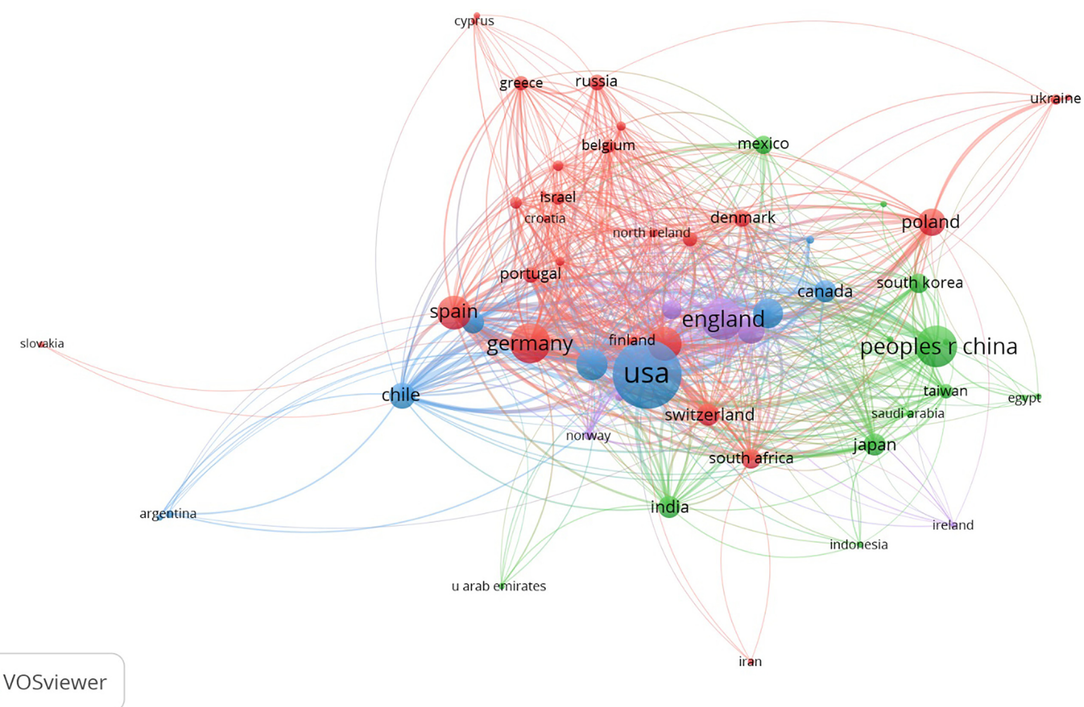

FIGURE 6 Map of co-authorships by country.

astrophysics. To this end, Figure 6 presents a map depicting co-authorship between researchers in different countries (with a minimum of three articles).

The colors represent clusters that are generated according to how often terms co-occur in a country; words that tend to co-occur more often are gathered into clusters. The sphere size reflects how many articles that country has published, and the line thickness represents the level of collaboration. The four countries that stand out as the most relevant are the USA, China, England, and Germany, leading the blue, green, violet, and red clusters, respectively, and showing a great level of collaboration within and between their clusters.

It appears that the co-authorship cluster structure primarily correlates with the respective geographical area and cultural and/or social affinity. In other words, primarily European countries occupy the red cluster (eventually with South Africa), while US authors are in the blue cluster. The violet cluster is associated with the United Kingdom and Ireland, and the green cluster comprises Asian countries and some collaborative efforts with Mexico and India. Based on this, we may assume that when researchers from different countries are co-authors, their collaboration is highly conditioned by factors including geographical closeness and cultural and linguistic similarity. For example, Spain plays a significant role as a research nexus linking Latin America with Europe.

On the other hand, in order to analyze the evolution over the years of the publications by country, Figure 7 is presented.

As can be seen, some countries appear to have more established research, for example, the United States, Mexico, and Poland, as they have bluer circles, meaning average publication dates that are 5 years older or more. Meanwhile, the most recent works come from the Nordic countries, for example, Finland and Norway, as well as countries like Belgium and Portugal; their circles in Figure 7 are yellow, meaning publication dates no more than 2 years ago. Finally, while it is comparatively less time-shifted, the research from Spain, Germany, and other European countries, as well as China and England, are in the 3–4 year range, that is, light blue to green.

3.2 | Topic analysis

VOSviewer was used to perform a topic analysis via semantic analysis, thereby creating a map showing the main subjects of interest, that is, relevant to the application of AI/ML to astronomy and astrophysics, and their relationships. Hereby, 4181 relevant keywords were initially identified in the WoS database; after setting the minimum threshold of

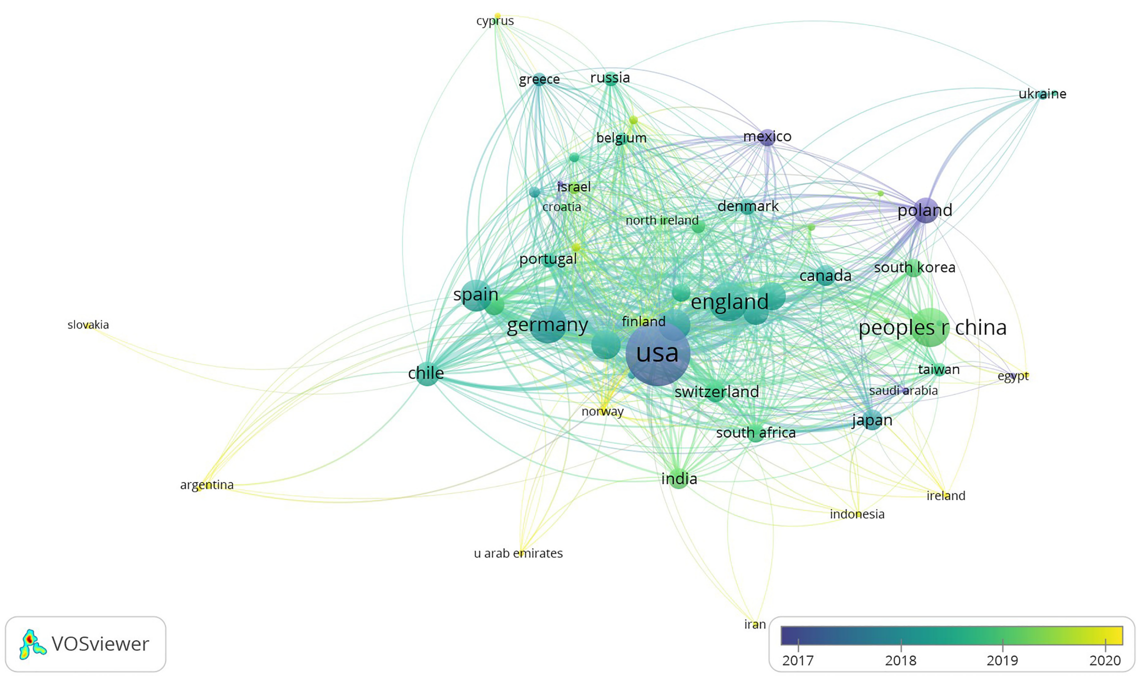

FIGURE 7 Mapping and visualization of the publications by country overlaid with the average publication year.

keyword occurrences to 10, only 157 remained. The clustering process subsequently classified these into five main clusters (i.e., colors) based on the similarities they presented (Figure 8).

The clusters reflect how often certain key terms co-occur, clustering those terms that often co-occur. The circle size demonstrates how often a term occurs. Although this process is automatic, the authors provided additional input in the results' interpretation to better identify certain issues and obtain an improved understanding of what the concept maps show. As this work sought to review almost 2800 documents from WoS, the following gives only the most recent and explanatory examples of the topics identified by the clusters and should thus not be considered an extensive review of all of the possible inputs.

Overall, five clusters were discernable; these are described in the following.

3.2.1 | AI and ML techniques applied to the fields of astronomy and astrophysics (red cluster)

This cluster mainly contains terms referring to the different AI and ML methods that can be applied to astronomy and astrophysics to analyze the datasets. As mentioned in the introduction, five main ML techniques are considered in the astronomical and astrophysical fields, namely PCA, DT, RF, SVM, and ANN. As the works and fields of astronomical/astrophysical research in which these methods are applied have already been introduced in Section 1, the following only expands this list and briefly explains the basis of these techniques.

PCA is a statistical method that simplifies the complexity of sample spaces with many simultaneous dimensions while preserving their information (Jolliffe, 2002; Morrison, 2005). In this way, a large dataset with multiple correlated variables can be described through a few new uncorrelated variables that describe most of the data's variability. One of the first applications of PCA to astronomy was in the 1960s, as a multivariate analysis tool to classify stellar spectra (Deeming, 1964) and subsequently address the morphological classification of galaxies (Whitmore, 1984), detect quasars (Francis et al., 1992), study the sun through the analysis of its magnetic fields and sunspots (Takalo & Mursula, 2018; Zharkova et al., 2012), and estimate the physical structure of the interstellar medium (Gratier et al., 2017).

DT are a class of prediction and classification models that, by recursively separating a given set of data, create logical construction diagrams to represent and categorize a series of successively occurring conditions to solve a problem

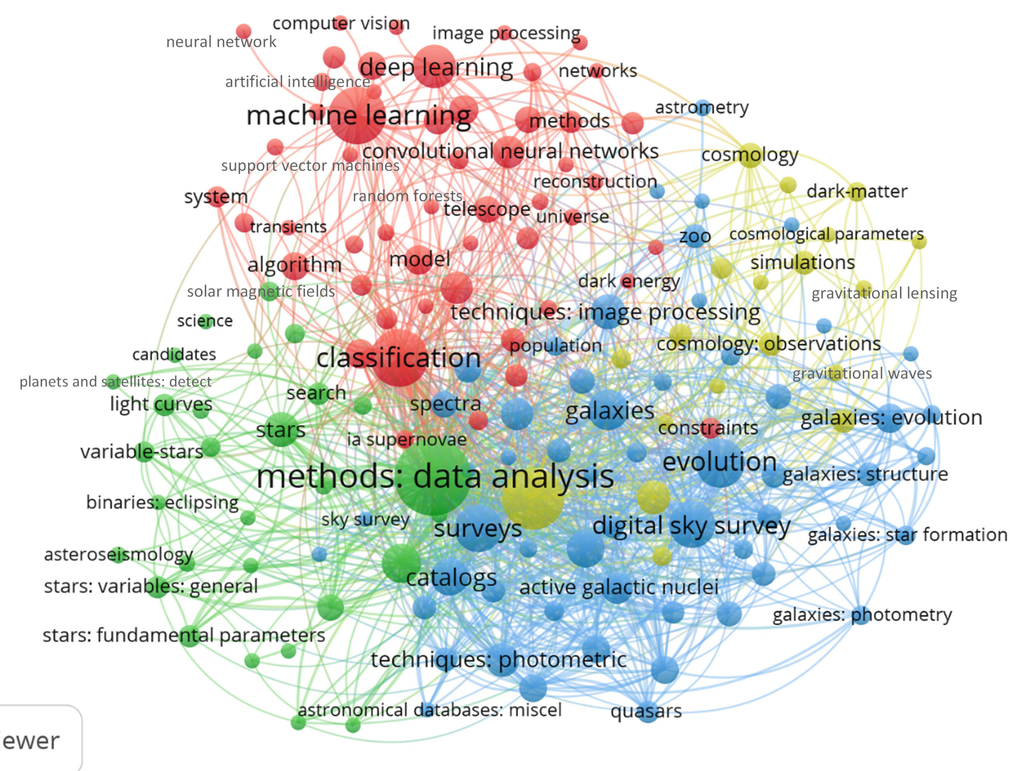

FIGURE 8 Map of topics (co-occurrence of keywords).

(Quinlan, 1986; Rokach & Maimon, 2008). The application of DTs to astronomy began in the 1990s, with tasks such as star-galaxy identification (Weir et al., 1995) to subsequently investigate the morphological classification of galaxies (Kriessler et al., 1998; Owens et al., 1996), predict the generation of solar flares (Ma et al., 2017), and estimate the redshift of quasars (Meshcheryakov et al., 2018). A further development of DT is the adaptive boosting (AdaBoost) method (Freund & Schapire, 1995), which consists of assigning weights to the contribution of each DT in such a way that previously misclassified data are given greater importance; this produces a new classifier that achieves better and more accurate results. AdaBoost has been used in astronomy for photometric redshift estimation (Hoyle et al., 2015), star-galaxy identification (Sevilla-Noarbe & Etayo-Sotos, 2015), star-quasar classification (Viquar et al., 2019), and light curve analysis (Edes-Huyal et al., 2021). An alternative use of DTs is gradient boosting (Friedman, 2001), a model that is formed by a set of individual DTs trained sequentially, wherein each new tree tries to improve the errors of the previous trees. The prediction of a new observation is thereby obtained by adding the predictions of all the individual trees that constitute the model. Gradient boosting has been applied to astronomy in fields such as star-galaxy classification (Morice-Atkinson et al., 2018), the prediction of solar wind (Bailey et al., 2021), and the identification of stars, galaxies, and active galactic nuclei (Golob et al., 2021).

The RF algorithm also arose as an extension of DT (Breiman, 2001) and consists of a group (ensemble) of DTs that are trained with different subsets of the data (a technique called bagging) in such a way that no tree considers all the training data. Thus, each tree is trained with different data samples for the same problem; by combining their results, some errors compensate for others, producing a prediction with better generalization. The use of RF in astronomy and astrophysics has been widespread since the 2000s, including in the analysis of outliers in astronomical images (Baron & Poznanski, 2017), predicting disruptive events in galaxies (French & Zabludoff, 2018), studying black hole growth (Goulding et al., 2018), detecting radio pulses (Pang et al., 2018), and classifying supernova (Ishida et al., 2019).

SVM encompass a classification technique that, given a supplied dataset, tries to find a hyperplane that allows the group to be divided into subsets following a certain decision criterion (Cortes & Vapnik, 1995). The aim is to find the hyperplane that provides the maximum separation between the data subgroups (support vectors). Also used since the 2000s, SVM have been applied in astronomy in areas such as detecting variable stars (Woźniak et al., 2004), estimating photometric redshifts (Wadadekar, 2004), identifying star-galaxies (Fadely et al., 2012), predicting the solar coronal mass (Liu et al., 2018), analyzing the physical properties of galaxies (Hui et al., 2018), and classifying evolved stars (Hernandez et al., 2021).

Finally, ANN, which constitute the basis of DL and were already proposed in the 1950s (Rosenblatt, 1957), can be considered the most powerful learning model within the field of AI and ML. ANNs, inspired by biological brains, comprise a set of units, called artificial neurons, which are connected to transmit signals. Thus, the input information crosses the neural network, where it is subjected to various operations weighted with the corresponding weights, producing output values. These systems learn and train themselves (rather than being explicitly programmed), and excel in areas where detecting solutions or characteristics is difficult to express using conventional programming. In the astronomical/astrophysical field, ANN first made an appearance in the 1990s in applications of adaptive optics for telescopes (Angel et al., 1990) and continued to find use in research fields such as the morphological classification of galaxies (Lahav et al., 1996), estimating photometric redshifts (Collister & Lahav, 2004), the search for cosmic strings (Ciuca & Hernández, 2017), the discovery of hypervelocity stars (Marchetti et al., 2017), identifying pulsars (Bethapudi & Desai, 2018), the analysis of neutron stars (Fujimoto et al., 2018), and predicting the parameters of stellar atmospheres (Azzam et al., 2021). CNN/DNN represent an evolution of ANN, as mentioned in the introduction, and have been used within the field of astronomy/astrophysics to morphologically classify galaxies (Huertas-Company et al., 2015), research transit objects (Gieseke et al., 2017), detect exoplanets (Pearson et al., 2018; Shallue & Vanderburg, 2018), and identify lens (Davies et al., 2019) and gravitational waves (Baltus et al., 2021).

Other terms that appear in this cluster, such as "computer vision" and "image processing," refer to the methods used to process the millions of astronomical images that are produced every day (through, among other instruments, telescopes) and from which it is possible to obtain knowledge on the underlying physical processes (Müller et al., 2018) and determine the classification of the objects that appear (Ma et al., 2019). These methods are usually based on CNN (Dieleman et al., 2015; Hoyle, 2016) due to the power of this technique when classifying (Krizhevsky et al., 2012).

3.2.2 | Application of AI and ML in the field of stellar analysis (green cluster)

This group includes terms related to the study of stars within the scope of AI and ML applications. Two main lines of classification can hereby be distinguished: Photometric and spectral. Examples of AI/ML applications in the first can be found in Dorn-Wallenstein et al. (2021), Ksoll et al. (2018), and Zhang et al. (2018) and, in the second, in Garcia-Dias et al. (2018), Miettinen (2018), and Sharma et al. (2020). The use of AI/ML in astronomy has led to the recent discovery of new star types, including hypervelocity stars (Marchetti et al., 2017), hot sub-dwarfs (Bu et al., 2019), Wolf–Rayet type stars (Morello et al., 2018), and unresolved binaries (Kuntzer & Courbin, 2017). In addition, AI/ML methods have also been used to obtain new knowledge and classification algorithms regarding neutron stars (Fujimoto et al., 2018; Tan et al., 2018).

Furthermore, as can be seen in Figure 8, the prevalent term "variable stars" also appears in the green cluster. These star types provide intense photometric information that is ideal for the application of AI/ML techniques to enable both classification tasks (Hosenie et al., 2019; Naul et al., 2018; Papageorgiou et al., 2018) and the development of new insights (Pashchenko et al., 2018; van Roestel et al., 2018).

Another field related to stellar astrophysics that appears in this group—through the terms "planets and satellites: detect" and "light curves"—is the detection of exoplanets and transit objects. In this regard, ML models have substantially improved the accuracy in identifying candidates for extrasolar planets (Armstrong et al., 2018; Jara-Maldonado et al., 2020; Mislis et al., 2018), determining their possible habitability (Basak et al., 2021; Saha et al., 2018), and evaluating the composition of their atmosphere (Hayes et al., 2020; Márquez-Neila et al., 2018). With regard to transit objects, AI has made it possible to accelerate their identification, especially in real-time applications (Bellm et al., 2019; Djorgovski et al., 2016; Sánchez et al., 2018).

In the area of "solar magnetic fields," AI/ML is used to study the sun, for example, classifying and predicting solar flares (Abduallah et al., 2020; Benvenuto et al., 2018), anticipating coronal mass ejection phenomena (Liu et al., 2018; Tiwari, 2021), and predicting the number of sunspots in the next solar cycle (Panigrahi et al., 2021; Wang et al., 2021). Furthermore, AI/ML can also be applied within the solar system, for example, in the detection of trans-Neptunian objects (Chen et al., 2018; Henghes et al., 2020) or the classification of asteroids (Carruba et al., 2020; Erasmus et al., 2018).

Finally, another field of stellar astronomical research in which AI/ML methods can be applied in this cluster is "asteroseismology" or stellar seismology, where they can facilitate the study of the internal structure of pulsating stars based on their frequency spectra (Hon et al., 2017; Mackereth et al., 2021).

Study of cosmology using AI and ML methods (yellow cluster)

The yellow cluster groups the terms related to the application of AI/ML to cosmology. As demonstrated by the term "simulations," this field of study works with mathematical models that use computer simulations to consider the corresponding initial conditions and gravitational physical mechanisms, generating a large volume of data (generally, relative to evolution processes) that can be used for prediction and analysis tasks (Poczos, 2018; Rodríguez et al., 2018; Villaescusa-Navarro et al., 2021). Other examples of areas of cosmology in which AI/ML find use are the study of dark matter (Bertone et al., 2017; Lucie-Smith et al., 2018; Stephon et al., 2020) and its halos surrounding galaxies (Agarwal et al., 2018; Lucie-Smith et al., 2019; Nadler et al., 2018), dark energy (Arjona & Nesseris, 2020a; Escamilla-Rivera et al., 2020), models of the creation and expansion of the universe (Arjona & Nesseris, 2020b), the cosmic microwave background (CMB) (Arjona, 2020; Mishra & Reddy, 2020), and the total mass of galaxies (considering dark matter) (McLeod et al., 2017). At this point, generative adversarial networks (GAN) are worth mentioning; since they are capable of simplifying, or even eliminating, the use of costly simulations in cosmology, they constitute one of the most promising ML models in astronomy/astrophysics (Diakogiannis et al., 2019).

Another field of application of AI/ML in cosmological analysis, as indicated in this cluster by the term "gravitational lensing," is the study of gravitational lenses. In this sense, beyond the discoveries produced by AI/ML in this field (Mirzoyan et al., 2019; Ostrovski et al., 2017; Teimoorinia et al., 2020), it should be mentioned that one of the obstacles in applying DL in the search for this peculiar astrophysical phenomenon is the scarcity of data with which to carry out the training phase. For this reason, simulated lenses are often used, usually on a galactic scale. This has allowed gravitational lenses to be detected from data originating from various astronomical sources (Magro et al., 2021; Metcalf et al., 2019; Petrillo et al., 2019; Schaefer et al., 2017).

Finally, it is also worth highlighting the scope of application of AI/ML to cosmology through the analysis of gravitational waves, recently detected (Abbott et al., 2018) thanks to the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) (Harry & LIGO, 2010). Through the use of ML, it has been possible to improve detection by discarding noise in a more efficient way (Powell et al., 2017; Vajente et al., 2020).

3.2.3 | Analysis of galaxies by means of AI and ML techniques (blue cluster)

Finally, the blue group refers to terms related to the use of AI/ML methods in the study of galaxies. Indeed, one of the most widespread applications of AI/ML in astronomy has been in galaxy classification (Beck et al., 2018; De Diego et al., 2020; Domínguez Sánchez et al., 2018; Nolte et al., 2019). Notably, automatic galactic classification has undoubtedly been accelerated thanks to the previous (visual) classifier work carried out by numerous volunteers in the Galaxy Zoo project (Lintott et al., 2008) (note the term "zoo" in this cluster).

In addition to classification tasks, there are other applications of AI/ML to the analysis of galaxies, such as estimating their physical properties (Rafieferantsoa et al., 2018; Ucci et al., 2017) ("galaxies: structure"), their rate of star formation (Delli et al., 2019) ("galaxies: star formation"), the evolution of their discs (Forbes et al., 2019) ("galaxies: evolution"), their kinematics (Dawson et al., 2020; de los Rios et al., 2021), and their photometric redshifts (Bilicki et al., 2018; Razim et al., 2021; Zhang et al., 2020) ("galaxies: photometry").

Finally, the presence of the terms "active galactic nuclei" and "quasars" is notable within this group; these refer to the ML astronomical application that addresses the morphological classification of nuclei of active galaxies (Chang et al., 2021; Chen et al., 2020; Ma et al., 2019) and the analysis of gamma-ray emissions arising from these (Fidor & Sitarek, 2021) as well as the detection of quasars (Herle et al., 2020; Jin et al., 2019; Schindler et al., 2017) and blazars (Kang et al., 2019; Mao et al., 2020; Sversut & Neto, 2020).

3.3 | Topic variation over time

Following the topic analysis and the delineation of the clusters and affinities of the topics, Figure 9 presents an overlay visualization that superimposes each topic's average date of publication from 2015 onwards. This sheds light on both the more established lines of research in this field and which topics are receiving the most research attention (*hot topics*, defined as subjects currently of great interest). The blue circles in the figure refer to the start of the five-year period under investigation, while the green and yellow circles are more recent.

This technique was designed to enable observations to be made on how research has evolved over decades; however, in this work, the time span has been limited to 5 years because of the explosive growth in this topic in recent years, as highlighted in Figure 1.

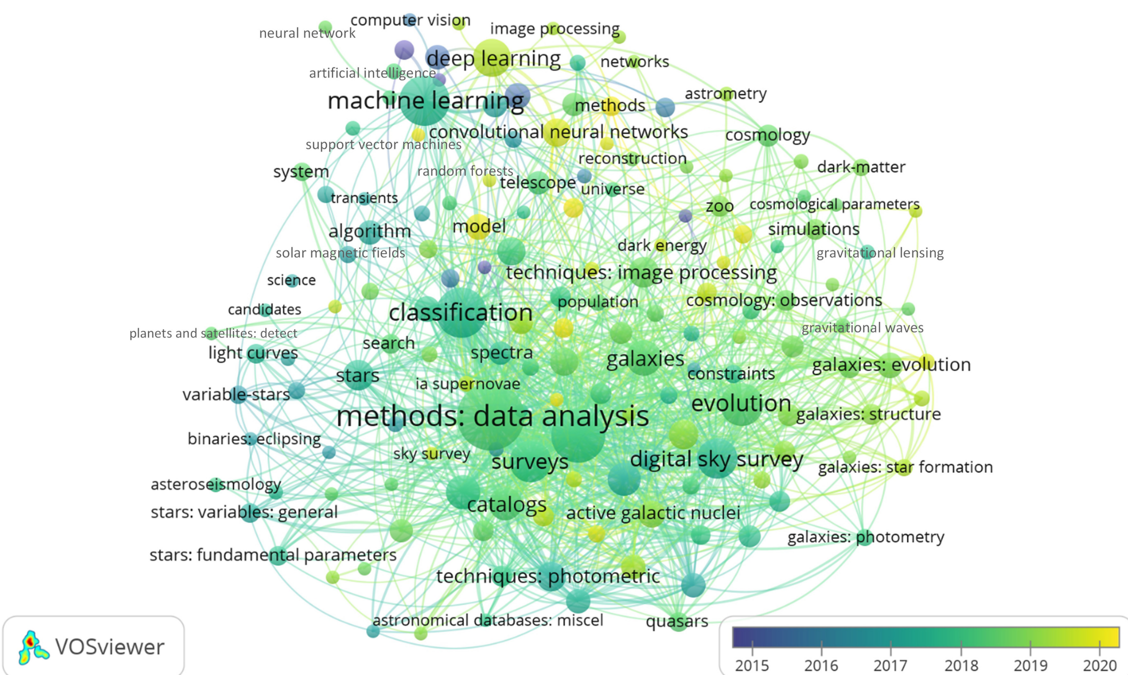

FIGURE 9 Mapping and visualization of the topics overlaid with the average publication year.

As can be seen, the cluster related to AI/ML applications in the analysis of stars (green in Figure 8) seems to be the most established of the four identified in Figure 8 as the topics that form it generally appear bluer, that is, have average publication dates in the first half of 5-year period. It should be noted that the great potential of AI/ML in the astronomical field was first realized in spectral and photometric stellar classification, among others.

On the other hand, the cluster on the use of AI/ML in cosmology (yellow in Figure 8) is the most recent in Figure 9 as the topics comprising it are, in general, green or yellow (within the last 2–3 years). Clearly, AI/ML is providing new methods to analyze the data produced through cosmological simulations to the extent that this line of research is at the forefront of the application of AI/ML in astrophysics. AI/ML thus provides knowledge on the evolution of the universe and its structure, including the study of dark matter and dark energy as well as the interstellar environment.

For its part, the cluster referring to the study of galaxies using AI/ML (the blue group in Figure 8) comprises topics that, while not being the most time-shifted, do represent a range spanning the last 3–4 years (from light blue to yellow). In this sense, research areas such as the application of AI/ML to galactic classification through photometry appear as the most established (bluer), while other fields, such as the analysis of the evolution and structure of galaxies or active galactic nuclei, emerge as pioneers (more green-yellow).

Finally, with topic colors ranging from blue to yellow, the cluster related to the different AI/ML techniques that are applied in the fields of astronomy/astrophysics (the red group in Figure 8) is the most distributed in Figure 9 in terms of publications over the last 5 years. This relates to the progression of the application of the various AI/ML techniques in astronomy and astrophysics, ranging from the most established ("computer vision") to the most innovative ("convolutional neural networks").

To further delve into this analysis, Figure 10 presents a specific graph showing how the prevalence of the five main AI/ML methods applied to astronomy/astrophysics problems has changed over the years (from 2010 to the present).

As can be observed, the use of ANN has been prioritized, experiencing dramatic growth in the last 5 years. Notably, the use of CNN has similarly grown substantially since 2015, when they first emerged (as mentioned above, CNN are an evolution of ANN). Meanwhile, SVM was the next most-used method until 2019, when it was overtaken by RF and later, PCA. It is worth noting that, although PCA is a technique used to reduce the dimensionality of datasets before the application of various ML prediction or classification algorithms, due to its prevalence since the 1960s within astrophysics, such as for stellar spectrum, galaxy or quasar classification, we include the PCA curve in Figure 10, bearing in mind that this curve should be interpreted as the number of publications that use PCA in combination with other—more strictly speaking—ML methods. Finally, DTs have seen discreet yet regular use over the years.

Figure 11 presents a map of the main AI/ML techniques applied to astronomy/astrophysics (the five main ones already mentioned, along with some additional methods based on the keywords) overlaid with the average publication year (since 2011).

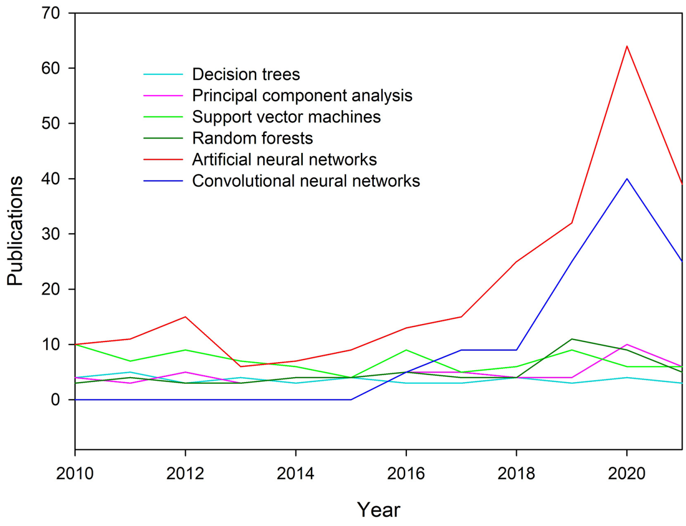

FIGURE 10 Timelines of the publications regarding the main ML/AI methods applied.

CNN, together with innovative methods such as LSTM and Bayesian inference, are the most recent. Next is the group formed by RF, gradient boosting and naive Bayes, followed by that including k-nearest neighbors, maximum likelihood, ANN and PCA. Finally, DT, SVM, and genetic algorithms are the least recently used on the map.

It is worth bearing in mind that the notions presented above reflect a general interpretative view; individual points could also exist in all the thematic clusters associated with topics that emerged earlier or later than those close by. Furthermore, recall that the colors used in Figure 9 are the average publication dates for articles on each topic; in other words, this does not necessarily mean that pioneering or recent publications on a given topic did not exist.

3.4 | Citations and highly cited elements

Based on the journals, authors and articles with the highest number of citations, conclusions can be drawn on which research elements were the most important for the field of AI/ML applied to astronomy and astrophysics. Thus, the evolution of the total number of citations of publications on the application of AI/ML in the fields of astronomy and astrophysics is shown in Figure 12.

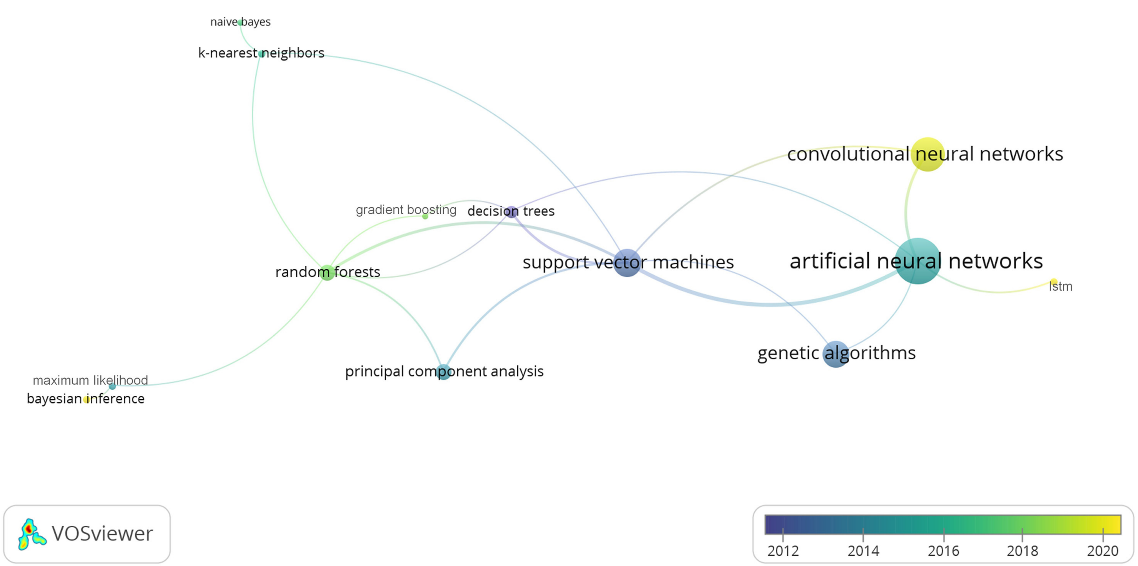

FIGURE 11 Mapping and visualization of the methods overlaid with the average publication year.

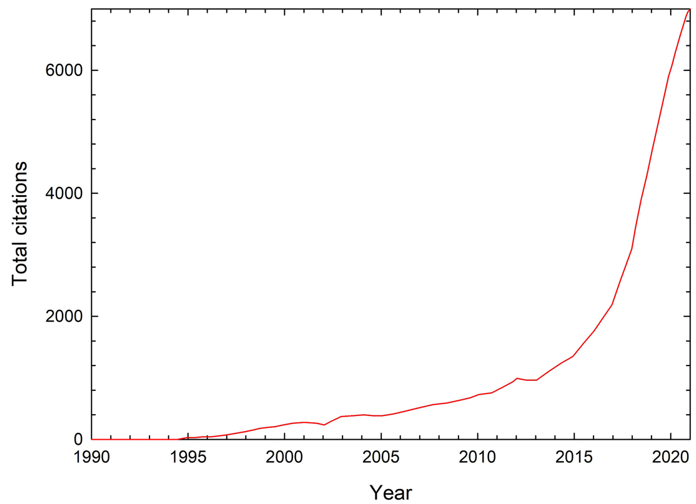

FIGURE 12 Timeline of the total number of citations.

Clearly, the number of citations has skyrocketed in recent decades, especially since 2005. The citations hereby display an exponential behavior that is directly related to the recent dramatic growth in the number of publications in the field under study (as already observed in Figure 1).

The following presents a more detailed study of the citations (by source and author) as well as an analysis of the most cited publications.

3.4.1 | Citation by source

As a preliminary step to studying the citations according to their source, it is interesting to analyze the most productive sources in this field. Table 3 lists the 15 sources with the most publications.

The two journals with the highest number of publications are *Astronomy & Astrophysics* and *Monthly Notices of the Royal Astronomical Society*. Two additional large groups of sources are observed, namely those dedicated to general astronomy and astrophysics, and those focused on the application of computing to astronomy.

Meanwhile, to facilitate a source-based analysis of the citations, the 15 most-cited journals were taken from WoS (Table 4).

Several journal types in certain categories become immediately visible, namely journals dedicated to astronomy and astrophysics, multidisciplinary journals, journals on computing, and journals on the earth and physics. Although the presence of the journal *PLoS Biology* might be surprising, its inclusion in this list is motivated by publications that analyze the relationship between the behavior of ANN and biological neurons. In any case, the first places on the list are occupied by journals specializing in astronomy and astrophysics. It should be noted that *Big Data*, *Little Data*, *No Data*: *Scholarship in the Networked World* is actually a book.

The results show that with 7691 citations, *Astronomy & Astrophysics Supplement Series* was the most cited journal, followed by *Monthly Notices of the Royal Astronomical Society* with 5412 citations and *Astronomy & Astrophysics* with 5358 citations. Notably, while certain journals were very frequently cited by numerous publications on the topic, others produced a vast number of citations, yet were actually cited by far fewer articles.

Lastly, 4 of the 15 selected journals were open access (*Astrophysical Journal*, *Astrophysical Journal Supplement Series*, *PLoS Biology*, *and Scientific Reports*), although some of the others also avail of this publication channel.

TABLE 3 Top 15 most productive sources

Source	No. of publications
Astronomy & Astrophysics	344
Monthly Notices of the Royal Astronomical Society	331
Proceedings of SPIE	211
Astrophysical Journal	158
Astronomy and Computing	83
Astronomical Journal	72
Research in Astronomy and Astrophysics	63
Astronomical Society of the Pacific Conference Series	59
EPJ Web of Conferences	54
Library and Information Services in Astronomy	46
Astrophysical Journal Supplement Series	45
Astrophysics and Space Science	40
Lecture Notes in Computer Science	34
Chapman Hall CRC Data Mining and Knowledge Discovery Series	33
Science China Physics Mechanics Astronomy	33

3.4.2 | Most cited publications

Table 5 lists the 15 publications with the most citations within the scope of AI/ML applications in astronomy and astrophysics.

The article "SExtractor: Software for source extraction," published in the journal *Astronomy & Astrophysics Supplement Series* and presenting software that automatically detects and classifies sources (stars, galaxies, etc.) from

TABLE 4 Top 15 most cited sources

Publication title	Citations
Astronomy & Astrophysics Supplement Series	7691
Monthly Notices of the Royal Astronomical Society	5412
Astronomy & Astrophysics	5358
Astrophysical Journal	2306
Astrophysical Journal Supplement Series	1000
Annals of Applied Statistics	597
PLoS Biology	577
Nature	577
Annual Review of Astronomy and Astrophysics	348
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	327
Big Data, Little Data, No Data: Scholarship in the Networked World	286
Scientific Reports	164
International Journal of Modern Physics D	155
International Journal of Climatology	147
Scientia Iranica	121

TABLE 5 Top 15 most cited publications

Title	Publication date	Citations
SExtractor: Software for source extraction (Bertin & Arnouts, 1996)	Jun 1996	7509
A correlated topic model of science (Blei & Lafferty, 2007)	Jun 2007	597
Big data: Astronomical or genomical? (Stephens et al., 2015)	Jul 2015	558
Genetic algorithms in astronomy and astrophysics (Charbonneau, 1995)	Dec 1995	538
Image reconstruction by domain-transform manifold learning (Zhu et al., 2018)	Mar 2018	473
Adaptive optics for astronomy—principles, performance, and applications (Beckers, 1993)	Jan 1993	348
Rotation-invariant convolutional neural networks for galaxy morphology prediction (Dieleman et al., 2015)	Jun 2015	314
Big data, little data, no data: Scholarship in the networked world (Borgman, 2015)	Jan 2015	286
Imaging the Schwarzschild-radius-scale Structure of M87 with the event horizon telescope using sparse modeling (Akiyama et al., 2017)	Mar 2017	227
A kernel-based feature selection method for SVM With RBF kernel for hyperspectral image classification (Kuo et al., 2014)	Jan 2014	179
A Gaussian process framework for modeling instrumental systematics: application to transmission spectroscopy (Gibson et al., 2011)	Jan 2012	170
Gaia early data release 3 summary of the contents and survey properties (Brown et al., 2021)	Apr 2021	160
Machine learning quantum phases of matter beyond the fermion sign problem (Broecker et al., 2017)	Aug 2017	159
An asteroseismic view of the radius valley: Stripped cores, not born rocky (Eylen et al., 2017)	Oct 2018	156
Data mining and machine learning in astronomy (Ball & Brunner, 2010)	Jul 2010	155

astronomical images, leads the *ranking* very clearly, with 7509 citations. As expected, the first positions of the table tend to be occupied by older publications. Meanwhile, it is possible to identify groups of publications that cover both generic topics of the application of AI/ML techniques to astronomy/astrophysics and more specific questions related to classification/prediction methods or astronomical observation campaigns.

3.5 | Co-citation analysis

The statistical technique of co-citation analysis enables the idle relationships between journals and/or authors to be detected and then visually expressed as co-citation clusters, thereby elucidating the meaning of this information. This method is derived from the assumption that articles by certain co-cited researchers or journals tend to tackle concepts that are similar or at least related. Co-citation refers to the instance when two separate publications are cited by a third.

3.5.1 | Co-citation by source

In Figure 13, the co-citations map (a minimum of 20 co-citations, which led to 237 sources) is shown as a function of the publication source, for articles related to the application of AI/ML to astronomy/astrophysics.

Regarding the co-citations according to their source (generally, a journal), these occur in four clusters, as described in the following.

The green cluster (which has the highest prevalence) refers to generic journals in the field of astronomy and astrophysics, for example, *Monthly Notices of the Royal Astronomical Society* (9543 co-citations) and *Astrophysical Journal* (8843 co-citations), had already appeared in the first positions of Tables 3 and 4, that is, had the highest number of publications and citations, respectively. Based on the size of the items in this group, the main journals in this cluster appear to be the most co-cited options within the research field under analysis.

The red cluster group comprises journals whose scope is the AI/ML field (*Journal of Machine Learning Research*, with 534 co-citations, *Machine Learning*, with 435 co-citations, or *IEEE Transactions on Pattern Analysis and Machine Intelligence*, with 173 co-citations) or its application to astronomy (*Astronomy and Computing*, with 292 co-citations). As a special case in this group, the journal *Nature* appears in red (923 co-citations); as the general

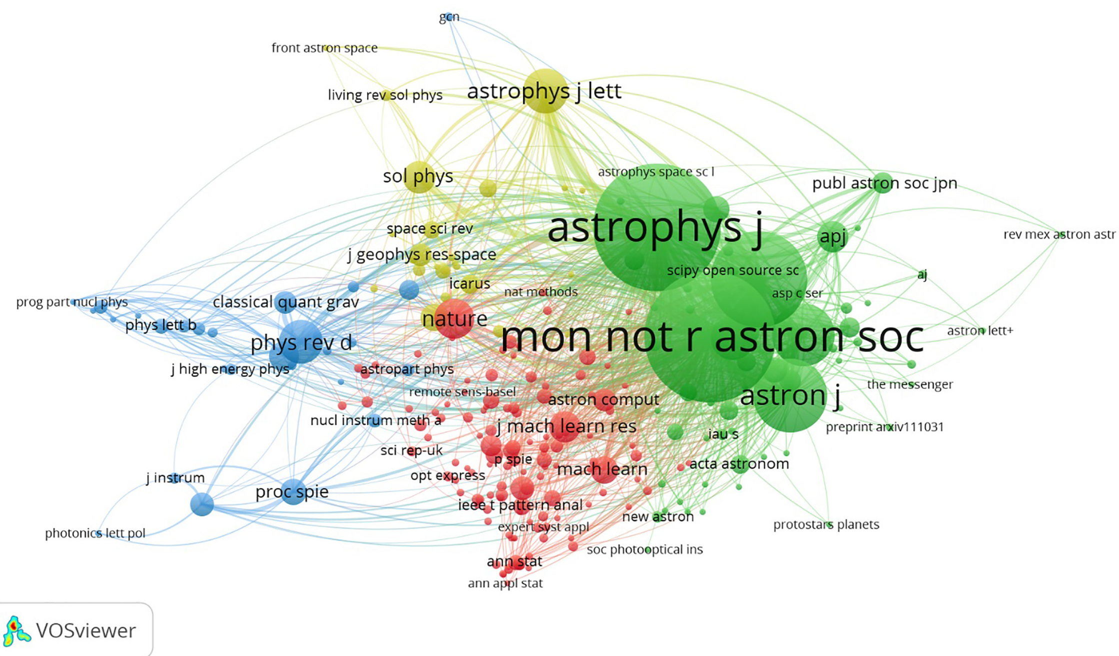

FIGURE 13 Map of co-citation by source.

reference journal within the scientific field, and considering that it is near the center of the map, it could be drawn in the color of any of the clusters.

The blue cluster represents the journals focused on the field of general physics, such as *Physical Review D* (1093 co-citations) and *Physics Letters B* (132 co-citations), or on the physical phenomena underlying astronomy and astrophysics, such as *Proceedings of SPIE* (388 co-citations), *Classical and Quantum Gravity* (302 co-citations), and *Journal of High Energy Physics* (77 co-citations).

Finally, the yellow cluster refers to sources focused on the study of space or the space environment, such as the *Journal of Geophysical Research: Space Physics* (293 co-citations) and *Space Science Reviews* (179 co-citations), or, specifically, on solar astrophysics, such as *Solar Physics* (584 co-citations) or *Living Reviews in Solar Physics* (71 co-citations). It should be noted that this group also contains the journal *The Astrophysical Journal Letters* (1133 co-citations); although it addresses general issues in astrophysics, with regard to co-citations, it seems to be prevalent in questions relating to the space environment or solar astrophysics.

3.5.2 | Co-citations by author

Examining co-citations can clarify the relationships between the authors. Specifically, co-cited authors likely have a connection of some kind, for example through a common subject, discipline, affiliation, or country of origin.

Before analyzing the map of co-citations by author, it is interesting to show the most productive authors in terms of publications related to the application of AI/ML to astronomy/astrophysics. As such, Table 6 contains the 15 authors with the most publications.

The table is led by Professor Ping Guo from Beijing Normal University, China, with 35 publications focused especially on the application of different AI/ML techniques to the classification of astronomical objects and phenomena (Qin et al., 2003; Wang et al., 2016). The next is researcher Yian-Xia Zhang from the CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China, with 33 publications primarily focused on the application of ANN to the estimation of redshifts (Han et al., 2020; Li et al., 2021). In third place is researcher Yong-Heng Zhao, also from the Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China, with 30 publications on the application of AI/ML to the classification of stars (Li et al., 2017; Wang et al., 2020). This highlights the prevalence of Chinese authors in the field of applying AI/ML to astronomy/astrophysics. It should be noted that Professor Giuseppe Longo from the Università

TABLE 6 Top 15 most productive authors

Authors	No. of publications
Guo, P.	35
Zhang, Y. X.	33
Zhao, Y. H.	30
Longo, G.	30
Bertin, E.	25
Brescia, M.	23
Cavouti, S.	22
Kind, M. C.	22
Ishida, E. E. O.	21
Djorgovski, S. G.	20
Smith, M.	20
Yin, Q.	20
Yu, C.	20
Brooks, D.	19
Carretero, J.	19

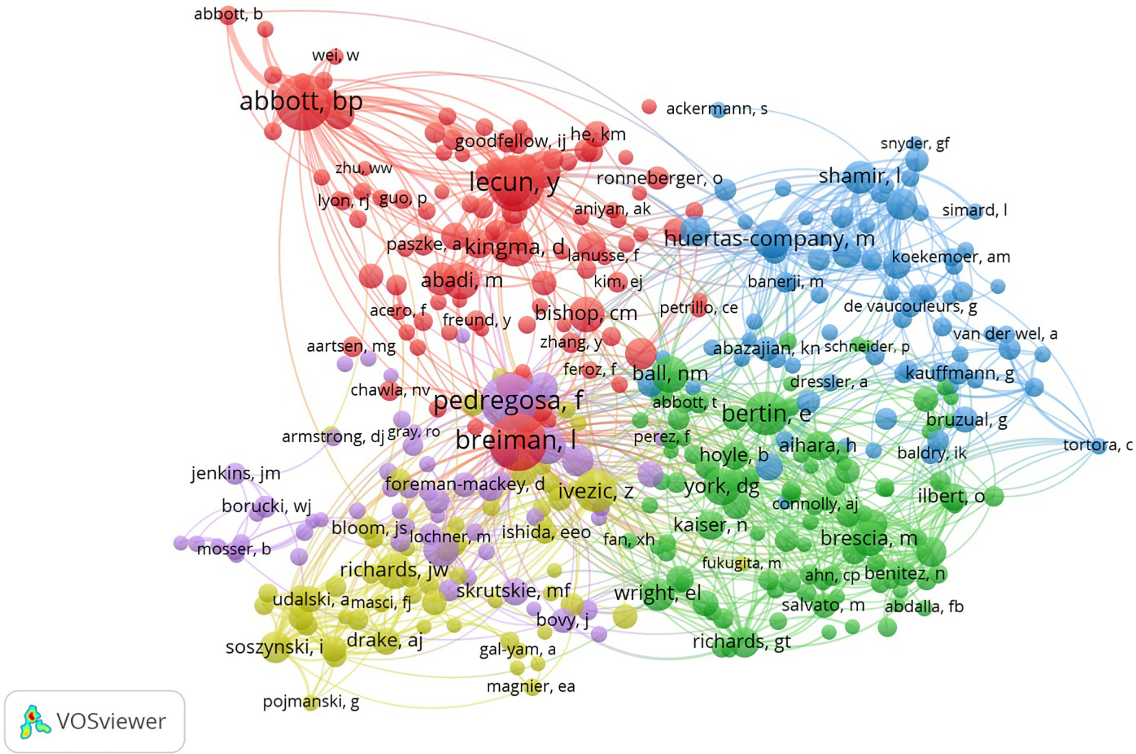

FIGURE 14 Map of co-citations by author.

Federico II, Napoli, Italy, ties for third place, with 30 publications focusing on the use of AI/ML for the estimation of photometric redshifts (Bilicki et al., 2018; Razim et al., 2021).

Figure 14 shows the map of co-citations per author (a minimum of 20 co-citations, which led to 410 authors), in which five clusters can be distinguished; these are described below.

The red cluster comprises authors who focus on the analysis and development of AI/ML techniques that are ultimately applied to the field of astronomy. In this sense, it is worth highlighting researcher Leo Breiman (271 co-citations) from the University of California, who is an RF expert (Breiman, 2001); Professor Yan Lecun (257 co-citations) from the New York University, who specializes in DL (LeCun, 2018); researcher Diederik P. Kingma (123 co-citations) from the University of Amsterdam, who focuses on ANN (Salimans & Kingma, 2016); and Christopher M. Bishop (108 co-citations) from the University of Edinburgh, who specializes in pattern recognition using ML (Bishop, 2010).

The violet group contains authors focused on the application of AI/ML in the search for exoplanets, such as William J. Borucki (43 co-citations) from the California Institute of Technology (Borucki et al., 2010); Jon M. Jenkins (40 co-citations) from the NASA Ames Research Center (Ansdell et al., 2018); and Benoit Mosser (32 co-citations) from the Observatoire de Paris: Laboratory for Space Studies and Instrumentation in Astrophysics (Nielsen et al., 2020).

The yellow cluster includes authors specialized in the use of AI/ML in stellar analysis, such as Joseph W. Richards (106 co-citations) from the University of California (Richards et al., 2011); Professor Igor Soszynski (83 co-citations) of the Astronomical Observatory of the University of Warsaw (Soszyński et al., 2018); and Andrew J. Drake (66 co-citations) from the California Institute of Technology (Mahabal et al., 2017).

The green group includes authors exploring the use of AI/ML to study the physical properties of the space environment and astronomical objects, such as Emmanuel Bertin (162 co-citations) from the Institut d'Astrophysique de Paris (Morgan et al., 2020); Donald G. York (117 co-citations) from the University of Chicago (Fan et al., 2019); and Massimo Brescia (98 co-citations) from the INAF Astronomical Observatory of Capodimonte, Naples (Davide et al., 2017).

Finally, the blue cluster groups authors who investigate the application of AI/ML to study galaxies, such as Marc Huertas-Company (123 co-citations) from de la Université de Paris (Cheng et al., 2021); Lior Shamir (95 co-citations) from Kansas State University Computer Science (Kuminski & Shamir, 2018); and Guinevere Kauffmann (47 co-citations) from the Max Planck Institute for Astrophysics (Grand et al., 2019).

4 | CONCLUSION

This work presented a text-mining-based scientometric analysis of the scientific output in the last three decades regarding the use of AI and ML in the fields of astronomy and astrophysics. Using the VOSviewer software and considering WoS as the data source, the study explored the evolution of publications in this research field, their distribution by country (including co-authorship), the most relevant topics addressed, and the most cited elements—as well as the most significant co-citations—according to the publication source and authorship.

In view of the obtained results, it can be stated that the scientific output on the application of AI/ML to astronomy/astrophysics has experienced exponential growth in recent years, especially since 2005, with the *Astronomy Astrophysics* category of WoS clearly emerging as the most prolific. Furthermore, publications have especially concentrated in the United States, China, and the United Kingdom—three of the strongest global economies. Meanwhile, four countries lead the co-authorship type of collaboration, namely the United States, China, the United Kingdom, and Germany.

Regarding the analysis of the main topics addressed in the scientific output of this field, four main lines of research have been identified, and these are the application of AI/ML techniques to astronomy/astrophysics, the application of AI/ML to analyze stars, the study of cosmology using AI/ML methods, and the analysis of galaxies using AI/ML. The second topic is the most established (stars), while the third is the most recent (cosmology).

The number of citations of publications on the application of AI/ML to astronomy/astrophysics has experienced a marked increase in recent decades, especially since 2005. The journal *Astronomy & Astrophysics Supplement Series* leads this ranking.

Finally, the number of co-citations mapped by source is headed by generic journals in the field of astronomy and astrophysics, namely *Monthly Notices of the Royal Astronomical Society* and *Astrophysical Journal*, while researcher Leo Breiman leads this category according to the author-based mapping.

In short, this work has shown how the application of AI and ML to the fields of astronomy and astrophysics is an already established and yet also significantly growing research field that is crucial to obtaining scientific knowledge about our universe. As AI techniques are in constant development, thus precipitating increasingly rapid and sophisticated methods, the potential for their future use in making new astronomical discoveries seems unimaginable. In light of this, our future research will further pursue the analysis of this fascinating topic through bibliographic and statistical methods. The aim here is to offer a more nuanced understanding of this area concerning how researchers collaborate, how scientific production has progressed, and which research trends are prevalent in this area.

AUTHOR CONTRIBUTIONS

José-Víctor Rodríguez: Conceptualization (lead); data curation (lead); formal analysis (lead); funding acquisition (equal); investigation (lead); methodology (equal); project administration (equal); resources (lead); software (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (lead); writing – review and editing (lead). **Ignacio Rodríguez-Rodríguez:** Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); software (equal); supervision (lead); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Wai Lok Woo:** Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); software (equal); supervision (lead); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal).

CONFLICT OF INTEREST

The authors have no competing interests to declare that are relevant to the content of this article.

DATA AVAILABILITY STATEMENT

All the data is available at Web of Science. https://www.webofknowledge.com.

ORCID

José-Víctor Rodríguez https://orcid.org/0000-0002-3298-6439
Ignacio Rodríguez-Rodríguez https://orcid.org/0000-0002-0118-3406
Wai Lok Woo https://orcid.org/0000-0002-8698-7605

RELATED WIREs ARTICLE

Surveying the reach and maturity of machine learning and artificial intelligence in astronomy

REFERENCES

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., Allam, S., Andrade-Oliveira, F., Annis, J., Avila, S., Banerji, M., Banik, N., Bechtol, K., Bernstein, G., Bernstein, R., Bertin, E., Brooks, D., Buckley-Geer, E., Burke, D., Camacho, H., Carnero Rosell, A., & Zhang, Y. (2018). Dark energy survey year 1 results: Measurement of the baryon acoustic oscillation scale in the distribution of galaxies to redshift 1. *Monthly Notices of the Royal Astronomical Society*, *483*, 4866–4883.

Abduallah, Y., Wang, J. T., Nie, Y., Liu, C., & Wang, H. (2020). DeepSun: Machine-learning as a service for solar flare prediction. *Research in Astronomy and Astrophysics*, *21*, 1–11.

Agarwal, S., Davé, R., & Bassett, B. A. (2018). Painting galaxies into dark matter haloes using machine learning. *Monthly Notices of the Royal Astronomical Society*, *478*, 3410–3422.

Agnello, A., Kelly, B. C., Treu, T., & Marshall, P. J. (2015). Data mining for gravitationally lensed quasars. *Monthly Notices of the Royal Astronomical Society*, *448*, 1446–1462.

Ahmad, M., & Batcha, M. S. (2019). Mapping of publications productivity on journal of documentation 1989–2018: A study based on clarivate analytics–web of science database. *Library Philosophy and Practice (E-Journal)*, *2213*, 2213–2226.

Akiyama, K., Kuramochi, K., Ikeda, S., Fish, V., Tazaki, F., Honma, M., Doeleman, S., Broderick, A., Dexter, J., Mościbrodzka, M., Bouman, K., Chael, A., & Zaizen, M. (2017). Imaging the Schwarzschild-radius-scale structure of M87 with the event horizon telescope using sparse modeling. *The Astrophysical Journal*, *838*, 1. https://doi.org/10.3847/1538-4357/aa6305

Angel, J. R. P., Wizinowich, P., Lloyd-Hart, M., & Sandler, D. (1990). Adaptive optics for array telescopes using neural-network techniques. *Nature*, *348*, 221–224.

Ansdell, M., Ioannou, Y., Osborn, H., Sasdelli, M., Smith, J., Caldwell, D., Jenkins, J., Räissi, C., & Angerhausen, D. (2018). Scientific domain knowledge improves exoplanet transit classification with deep learning. *The Astrophysical Journal*, *869*, L7. https://doi.org/10.3847/2041-8213/aaf23b

Arjona, R. (2020). Machine learning meets the redshift evolution of the CMB temperature. *Journal of Cosmology and Astroparticle Physics*, *2020*, 1–19. https://doi.org/10.1088/1475-7516/2020/08/009

Arjona, R., & Nesseris, S. (2020a). Hints of dark energy anisotropic stress using machine learning. *Journal of Cosmology and Astroparticle Physics*, *2020*, 1–28.

Arjona, R. C., & Nesseris, S. (2020b). What can machine learning tell us about the background expansion of the universe? *Physical Review D*, *101*, 123525.

Armstrong, D. J., Günther, M. N., McCormac, J., Smith, A. M. S., Bayliss, D., Bouchy, F., Burleigh, M., Casewell, S., Eigmüller, P., Gillen, E., Goad, M., Hodgkin, S., Jenkins, J., Louden, T., Métrailler, L., Pollacco, D., Poppenhaeger, K., Queloz, D., Raynard, L., & Wheatley, P. J. (2018). Automatic vetting of planet candidates from ground-based surveys: Machine learning with NGTS. *Monthly Notices of the Royal Astronomical Society*, *478*, 4225–4237.

Azzam, Y. A., Nouh, M. I., & Shaker, A. A. (2021). Prediction of the atmospheric fundamental parameters from stellar spectra using artificial neural network. *NRIAG Journal of Astronomy and Geophysics*, *10*(1), 23–34. https://doi.org/10.1080/20909977.2020.1853012

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386.

Bailey, R. L., Reiss, M. A., Arge, C. N., Möstl, C., Henney, C. J., Owens, M. J., Amerstorfer, U. V., Amerstorfer, T., Weiss, A. J., & Hinterreiter, J. (2021). Using gradient boosting regression to improve ambient solar wind model predictions. *Space Weather*, *19*, e2020SW002673. https://doi.org/10.1029/2020SW002673

Ball, N., & Brunner, R. J. (2010). Data mining and machine learning in astronomy. *International Journal of Modern Physics D*, *19*, 1049–1106. https://doi.org/10.1142/S0218271810017160

Ball, N. M., Loveday, J., Fukugita, M., Nakamura, O., Okamura, S., Brinkmann, J., & Brunner, R. J. (2004). Galaxy types in the Sloan digital sky survey using supervised artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, *348*, 1038–1046.

Baltus, G., Janquart, J., Lopez, M., Reza, A., Caudill, S., & Cudell, J.-R. (2021). Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal. *Physical Review D*, *103*, 102003.

Baron, D., & Poznanski, D. (2017). The weirdest SDSS galaxies: Results from an outlier detection algorithm. *Monthly Notices of the Royal Astronomical Society*, *465*, 4530–4555.

Basak, S., Mathur, A., Theophilus, A. J., Deshpande, G., & Murthy, J. (2021). Habitability classification of exoplanets: A machine learning insight. *European Physical Journal Special Topics*, *230*, 2221–2251. https://doi.org/10.1140/epjs/s11734-021-00203-z

Beck, M. R., Scarlata, C., Fortson, L. F., Lintott, C. J., Simmons, B. D., Galloway, M. A., Willett, K., Dickinson, H., Masters, K., Marshall, P., & Wright, D. (2018). Integrating human and machine intelligence in galaxy morphology classification tasks. *Monthly Notices of the Royal Astronomical Society*, *476*, 5516–5534.

Beckers, J. M. (1993). Adaptive optics for astronomy: Principles, performance, and applications. *Annual Review of Astronomy and Astrophysics*, *31*(1), 13–62.

Bellm, E. C., Kulkarni, S. R., Graham, M. J., Dekany, R., Smith, R. M., Riddle, R., Masci, F. J., Helou, G., Prince, T. A., Adams, S. M., Barbarino, C., Barlow, T., Bauer, J., Beck, R., Belicki, J., Biswas, R., Blagorodnova, N., Bodewits, D., Bolin, B., … Zolkower, J. (2019). The Zwicky transient facility: System overview, performance, and first results. *Publications of the Astronomical Society of the Pacific*, *131*, 018002.

Benvenuto, F., Piana, M., Campi, C., & Massone, A. M. (2018). A hybrid supervised/unsupervised machine learning approach to solar flare prediction. *The Astrophysical Journal*, *853*, 90.

Bertin, E., & Arnouts, S. (1996). SExtractor: Software for source extraction. *Astronomy and Astrophysics, Supplement Series*, *117*(2), 393–404.

Bertone, G., Bozorgnia, N., Kim, J., Liem, S., McCabe, C., Otten, S., & Austri, R. (2017). Identifying WIMP dark matter from particle and astroparticle data. *Journal of Cosmology and Astroparticle Physics.*, *2018*, 1–46. https://doi.org/10.1088/1475-7516/2018/03/026

Bethapudi, S., & Desai, S. (2018). Separation of pulsar signals from noise using supervised machine learning algorithms. *Astronomy and Computing*, *23*, 15–26.

Bilicki, M., Hoekstra, H., Brown, M. J. I., Amaro, V., Blake, C., Cavuoti, S., de Jong, J. T. A., Georgiou, C., Hildebrandt, H., Wolf, C., Amon, A., Brescia, M., Brough, S., Costa-Duarte, M. V., Erben, T., Glazebrook, K., Grado, A., Heymans, C., Jarrett, T., … Wang, L. (2018). Photometric redshifts for the kilo-degree survey - machine-learning analysis with artificial neural networks. *Astronomy & Astrophysics*, *616*, A69.

Bishop, C. M. (2010). *Pattern recognition and machine learning*. Springer.

Biswas, R., Blackburn, L., Cao, J., Essick, R., Hodge, K., Katsavounidis, E., Kim, K., Kim, Y.-M., Le Bigot, E.-O., Lee, C.-H., Oh, J., Oh, S., Son, E., Vaulin, R., Wang, X., & Ye, T. (2013). Application of machine learning algorithms to the study of noise artifacts in gravitational-wave data. *Physical Review D*, *88*, 1–21. https://doi.org/10.1103/PhysRevD.88.062003

Blei, D., & Lafferty, J. (2007). A correlated topic model of Science. *The Annals of Applied Statistics*, *1*, 17–35. https://doi.org/10.1214/07-AOAS114

Blondel, V. D., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*, 10008.

Bonaccorso, G. (2018). *Machine learning algorithms: Popular algorithms for data science and machine learning*. Packt Publishing Ltd.

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. The MIT Press.

Borne, K. (2009). Scientific data mining in astronomy. In H. Kargupta, J. Han, P. S. Yu, R. Motwani, & V. Kumar (Eds.), *Next generation of data mining* (pp. 91–114). Taylor & Francis: CRC Press.

Borucki, W., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W., Devore, E., Dunham, E., Dupree, A., Gautier, T., Geary, J., Gilliland, R., Gould, A., Howell, S., Jenkins, J., Kondo, Y., & Prsa, A. (2010). Kepler planet-detection Mission: Introduction and first results. *Science (New York, N.Y.)*, *327*, 977–980. https://doi.org/10.1126/science.1185402

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for information Science and Technology*, *61*(12), 2389–2404.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Broadus, R. N. (1987). Toward a definition of "bibliometrics". *Scientometrics*, *12*(5-6), 373–379.

Broecker, P., Carrasquilla, J., Melko, R. G., & Trebst, S. (2017). Machine learning quantum phases of matter beyond the fermion sign problem. *Science Reports*, *18*, 8823. https://doi.org/10.1038/s41598-017-09098-0

Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., Biermann, M., Creevey, O. L., Evans, D. W., Eyer, L., Hutton, A., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., … Zwitter, T. (2021). Gaia early data release 3 - summary of the contents and survey properties. *Astronomy & Astrophysics*, *649*, A1. https://doi.org/10.1051/0004-6361/202039657

Brunner, R. J., Djorgovski, S. G., Prince, T. A., & Szalay, A. S. (2002). *Massive datasets in astronomy* (pp. 931–979). Springer US.

Bu, Y., Zeng, J., & Lei, Z. (2019). Searching for hot subdwarf stars from the LAMOST spectra. III. Classifying the hot subdwarf stars from fourth data release of LAMOST using deep learning method. *The Astrophysical Journal*, *886*(2), 128.

Carrasco Kind, M., & Brunner, R. J. (2013). TPZ: Photometric redshift PDFs and ancillary information by using prediction trees and random forests. *Monthly Notices of the Royal Astronomical Society*, *432*, 1483–1501.

Carruba, V., Aljbaae, S., Domingos, R., Lucchini, A., & Furlaneto, P. (2020). Machine learning classification of new asteroid families members. *Monthly Notices of the Royal Astronomical Society*, *496*, 540–549. https://doi.org/10.1093/mnras/staa1463

Chang, Y.-Y., Hsieh, B.-C., Wang, W.-H., Lin, Y.-T., Lim, C.-F., Toba, Y., Zhong, Y., & Chang, S.-Y. (2021). Identifying AGN host galaxies by machine learning with HSC+WISE. *The Astrophysical Journal*, *920*(2), 68.

Chao, H., Yue-hua, M., Zhao, H., & Lu, X.-p. (2017). Spectral classification of asteroids by random Forest. *Chinese Astronomy and Astrophysics*, *41*(4), 549–557.

Charbonneau, P. (1995). Genetic algorithms in astronomy and astrophysics. *The Astrophysical Journal Supplement Series*, *101*, 309. https://doi.org/10.1086/192242

Chellappandi, P., & Vijayakumar, C. S. (2018). Bibliometrics, scientometrics, webometrics/cybermetrics, informetrics and altmetrics: An emerging field in library and information science research. *Shanlax International Journal of Education*, *7*(1), 5–8.

Chen, B., Goto, T., Kim, S., Wang, T., Santos, D. J., Ho, S. C.-C., Hashimoto, T., Poliszczuk, A., Pollo, A., Trippe, S., Miyaji, T., Toba, Y., Malkan, M., Serjeant, S., Pearson, C., Hwang, H., Shim, H., Lu, T., & Matsuhara, H. (2020). An active galactic nucleus recognition model based on deep neural network. *Monthly Notices of the Royal Astronomical Society*, *501*, 3951–3961. https://doi.org/10.1093/mnras/staa3865

Chen, Y.-T., Lin, H.-W., Alexandersen, M., Lehner, M. J., Wang, S.-Y., Wang, J.-H., Yoshida, F., Komiyama, Y., & Miyazaki, S. (2018). Searching for moving objects in HSC-SSP: Pipeline and preliminary results. *Publications of the Astronomical Society of Japan*, *70*, S38.

Cheng, T.-Y., Huertas-Company, M., Conselice, C. J., Aragón-Salamanca, A., Robertson, B. E., & Ramachandra, N. (2021). Beyond the hubble sequence – Exploring galaxy morphology with unsupervised machine learning. *Monthly Notices of the Royal Astronomical Society*, *503*(3), 4446–4465. https://doi.org/10.1093/mnras/stab734

Ciuca, R., & Hernández, O. F. (2017). A Bayesian framework for cosmic string searches in CMB maps. *Journal of Cosmology and Astro-Particle Physics*, *2017*, 28.

Collister, A. A., & Lahav, O. (2004). ANNz: Estimating photometric redshifts using artificial neural networks. *Publications of the Astronomical Society of the Pacific*, *116*, 345–351.

Conselice, C. J. (2006). The fundamental properties of galaxies and a new galaxy classification system. *Monthly Notices of the Royal Astronomical Society*, *373*, 1389–1408.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.

Damiano, M., Micela, G., & Tinetti, G. (2019). A principal component analysis-based method to analyze high-resolution spectroscopic data on exoplanets. *The Astrophysical Journal*, *878*, 153. https://doi.org/10.3847/1538-4357/ab22b2

Davide, E., Molinari, S., Schisano, E., Pestalozzi, M., Pezzuto, S., Merello, M., Noriega-Crespo, A., Moore, T. J. T., Russeil, D., Mottram, J. C., Paladini, R., Strafella, F., Benedettini, M., Bernard, J. P., Di Giorgio, A., Eden, D. J., Fukui, Y., Plume, R., Bally, J., … Yamamoto, H. (2017). The hi-GAL compact source catalogue – I. the physical properties of the clumps in the inner galaxy ($-71\circ.0 < \ell < 67\circ.0$). *Monthly Notices of the Royal Astronomical Society*, *471*(1), 100–143. https://doi.org/10.1093/mnras/stx1357

Davies, A., Serjeant, S., & Bromley, J. (2019). Using convolutional neural networks to identify gravitational lenses in astronomical images. *Monthly Notices of the Royal Astronomical Society*, *487*, 5263–5271. https://doi.org/10.1093/mnras/stz1288

Davies, F. B., Hennawi, J. F., Bañados, E., Simcoe, R. A., Decarli, R., Xiaohui, F., Farina, E. P., Mazzucchelli, C., Rix, H.-W., Venemans, B. P., Walter, F., Wang, F., & Yang, J. (2018). Predicting quasar continua near Lyα with principal component analysis. *The Astrophysical Journal*, *864*(2), 143.

Dawson, J., Davis, T., Gomez, E., Schock, J., Zabel, N., & Williams, T. (2020). Using machine learning to study the kinematics of cold gas in galaxies. *Monthly Notices of the Royal Astronomical Society*, *491*(2), 2506–2519. https://doi.org/10.1093/mnras/stz3097

De Diego, J., Nadolny, J., Bongiovanni, A., Cepa, J., Povic, M., García, A., Padilla-Torres, C., Lara-López, M., Cerviño, M., Pérez-Martínez, R., Alfaro, E., Castañeda, H., Fernández-Lorenzo, M., Gallego, J., González, J., Gonzalez-Serrano, J., Pintos-Castro, I., Sanchez-Portal, M., Cedrés, B., & Bland-Hawthorn, J. (2020). Galaxy classification: Deep learning on the OTELO and COSMOS databases. *Astronomy & Astrophysics*, *638*, A134.

de León, J., Licandro, J., Serra-Ricart, M., Pinilla-Alonso, N., & Campins, H. (2010). Observations, compositional, and physical characterization of near-earth and Mars-crosser asteroids from a spectroscopic survey. *Astronomy & Astrophysics*, *517*, A23.

de los Rios, M., Martinez, H., Coenda, V., Muriel, H., Ruiz, A., Vega-Martínez, C., & Cora, S. (2021). ROGER: Reconstructing orbits of galaxies in extreme regions using machine learning techniques. *Monthly Notices of the Royal Astronomical Society*, *500*(2), 1784–1794. https://doi.org/10.1093/mnras/staa3339

Deeming, T. J. (1964). Stellar spectral classification: I. Application of component analysis. *Monthly Notices of the Royal Astronomical Society*, *127*(6), 493–516.

Delli, V. M., Cavuoti, S., Brescia, M., Longo, G., & Riccio, G. (2019). Stellar formation rates for photometric samples of galaxies using machine learning methods. *Monthly Notices of the Royal Astronomical Society*, *486*, 1377–1391. https://doi.org/10.1093/mnras/stz856

Diakogiannis, F. I., Lewis, G. F., Ibata, R. A., Guglielmo, M., Wilkinson, M. I., & Power, C. (2019). Reliable mass calculation in spherical gravitating systems. *Monthly Notices of the Royal Astronomical Society*, *482*, 3356–3372.

Dieleman, S., Willett, K. W., & Dambre, J. (2015). Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly Notices of the Royal Astronomical Society*, *450*(2), 1441–1459.

Djorgovski, S., Graham, M., Donalek, C., Mahabal, A., Drake, A., Turmon, M., & Fuchs, T. (2016). Real-time data mining of massive data streams from synoptic sky surveys. *Future Generation Computing Systems*, *59*, 95–104.

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. (2018). Improving galaxy morphologies for SDSS with deep learning. *Monthly Notices of the Royal Astronomical Society*, *476*, 3661–3676.

Dorn-Wallenstein, T., Davenport, J., Huppenkothen, D., & Levesque, E. (2021). Photometric classifications of evolved massive stars: Preparing for the era of Webb and Roman with machine learning. *The Astrophysical Journal*, *913*(1), 32.

Eatough, R. P., Molkenthin, N., Kramer, M., Noutsos, A., Keith, M. J., Stappers, B. W., & Lyne, A. G. (2010). Selection of radio pulsar candidates using artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, *407*, 2443–2450.

Edes-Huyal, F., Cataltepe, Z., & Kahya, E. (2021). On the classification and feature relevance of multiband light curves. *The Astronomical Journal*, *161*, 168. https://doi.org/10.3847/1538-3881/abdecf

Erasmus, N., McNeill, A., Mommert, M., Trilling, D. E., Sickafoose, A. A., & van Gend, C. (2018). Taxonomy and light-curve data of 1000 serendipitously observed Main-belt asteroids. *Astrophysical Journal Supplement*, *237*, 19.

Escamilla-Rivera, C., Carvajal, M., & Capozziello, S. (2020). A deep learning approach to cosmological dark energy models. *Journal of Cosmology and Astroparticle Physics.*, *2020*, 1–22. https://doi.org/10.1088/1475-7516/2020/03/008

Eylen, V., Agentoft, C., Lundkvist, M., Kjeldsen, H., Owen, J., Fulton, B., Petigura, E., & Snellen, I. (2017). An asteroseismic view of the radius valley: Stripped cores, not born rocky. *Monthly Notices of the Royal Astronomical Society*, *479*, 1–12. https://doi.org/10.1093/mnras/sty1783

Fadely, R., Hogg, D. W., & Willman, B. (2012). Star-galaxy classification in multi-band optical imaging. *The Astrophysical Journal*, *760*, 15.

Fan, H., Hobbs, L., Dahlstrom, J., Welty, D., York, D. G., Rachford, B., Snow, T., Sonnentrucker, P., Baskes, N., & Gang, Z. (2019). The apache point observatory catalog of optical diffuse interstellar bands. *The Astrophysical Journal*, *878*, 151. https://doi.org/10.3847/1538-4357/ab1b74

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*, 37–53.

Feigelson, E., & Babu, G. (2012). Big data in astronomy. *Significance*, *9*, 22–25.

Fidor, T., & Sitarek, J. (2021). Assessing the capability of random forest to predict the evolution of enhanced gamma-ray states of active galactic nuclei. *Astroparticle Physics*, *2021*, 1–12. https://doi.org/10.1016/j.astropartphys.2021.102625

Fluke, C. J., & Jacobs, C. (2020). Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*, e1349.

Forbes, J., Krumholz, M., & Speagle, J. (2019). Towards a radially-resolved semi-analytic model for the evolution of disc galaxies tuned with machine learning. *Monthly Notices of the Royal Astronomical Society*, *487*(3), 3581–3606. https://doi.org/10.1093/mnras/stz1473

Francis, P. J., Hewett, P. C., Foltz, C. B., & Chaffee, F. H. (1992). An objective classification scheme for QSO spectra. *The Astrophysical Journal*, *398*, 476.

French, K. D., & Zabludoff, A. I. (2018). Identifying tidal disruption events via prior photometric selection of their preferred hosts. *The Astrophysical Journal*, *868*, 99.

Freund, Y., & Schapire, R. E. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. In P. Vitányi (Ed.), *Computational learning theory lecture notes in computer science* (pp. 23–37). Springer.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*, 1189–1232.

Fujimoto, Y., Fukushima, K., & Murase, K. (2018). Methodology study of machine learning for the neutron star equation of state. *Physical Review D*, *98*, 023019.

Garcia-Dias, R., Allende Prieto, C., Sánchez Almeida, J., & Ordovás-Pascual, I. (2018). Machine learning in APOGEE. Unsupervised spectral classification with K-means. *Astronomy & Astrophysics*, *612*, A98.

Gibson, N., Aigrain, S., Roberts, S., Evans, T., Osborne, M., & Pont, F. (2011). A Gaussian process framework for modelling instrumental systematics: Application to transmission spectroscopy. *Monthly Notices of The Royal Astronomical Society*, *419*, 1–13. https://doi.org/10.1111/j.1365-2966.2011.19915.x

Gieseke, F., Bloemen, S., van den Bogaard, C., Heskes, T., Kindler, J., Scalzo, R. A., Ribeiro, V., Roestel, J., Groot, P., Yuan, F., Moller, A., & Tucker, B. E. (2017). Convolutional neural networks for transient candidate vetting in large-scale surveys. *Monthly Notices of the Royal Astronomical Society*, *472*, 3101–3114.

Golob, A., Sawicki, M., Goulding, A. D., & Coupon, J. (2021). Classifying stars, galaxies, and AGNs in CLAUDS + HSC-SSP using gradient boosted decision trees. *Monthly Notices of the Royal Astronomical Society*, *503*(3), 4136–4146. https://doi.org/10.1093/mnras/stab719

Gómez, C., Neira, M., Hernández-Hoyos, M., Arbeláez, P., & Forero-Romero, J.-E. (2020). Classifying image sequences of astronomical transients with deep neural networks. *Monthly Notices of the Royal Astronomical Society*, *499*(3), 3130–3138.

Goulding, A. D., Greene, J. E., Bezanson, R., Greco, J., Johnson, S., Leauthaud, A., Matsuoka, Y., Medezinski, E., & Price-Whelan, A. M. (2018). Galaxy interactions trigger rapid black hole growth: An unprecedented view from the hyper Suprime-cam survey. *Publications of the Astronomical Society of Japan*, *70*, S37.

Graham, M. J., Djorgovski, S. G., Mahabal, A. A., Donalek, C., & Drake, A. J. (2013). Machine-assisted discovery of relationships in astronomy. *Monthly Notices of the Royal Astronomical Society*, *431*, 2371–2384.

Grand, R., Voort, F., Zjupa, J., Fragkoudi, F., Gómez, F., Kauffmann, G., Marinacci, F., Pakmor, R., Springel, V., & White, S. (2019). Gas accretion and galactic fountain flows in the Auriga cosmological simulations: Angular momentum and metal re-distribution. *Monthly Notices of the Royal Astronomical Society*, *490*, 4786–4803. https://doi.org/10.1093/mnras/stz2928

Gratier, P., Bron, E., Gerin, M., Pety, J., Guzman, V., Orkisz, J., Bardeau, S., Goicoechea, J., Petit, F., Liszt, H., Öberg, K., Peretto, N., Roueff, E., Sievers, A., & Tremblin, P. (2017). Dissecting the molecular structure of the Orion B cloud: Insight from principal component analysis. *Astronomy & Astrophysics*, *599*, 1–15. https://doi.org/10.1051/0004-6361/201629847

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *Journal of informetrics*, *11*(3), 823–834.

Han, B., Qiao, L.-N., Chen, J.-L., Zhang, X.-D., Zhang, Y. X., & Zhao, Y. (2020). GeneticKNN: A weighted KNN approach supported by genetic algorithm for photometric redshift estimation of quasars. *Research in Astronomy and Astrophysics*, *21*(1), 17.

Harry, G. M., & LIGO Scientific Collaboration. (2010). Advanced LIGO: The next generation of gravitational wave detectors. *Classical and Quantum Gravity*, *27*, 084006.

Hayes, J., Kerins, E., Awiphan, S., McDonald, I., Morgan, J., Chuanraksasat, P., Komonjinda, S., Sanguansak, N., & Kittara, P. (2020). Optimizing exoplanet atmosphere retrieval using unsupervised machine-learning classification. *Monthly Notices of the Royal Astronomical Society*, *494*, 4492–4508. https://doi.org/10.1093/mnras/staa978

Henghes, B., Lahav, O., Gerdes, D., Lin, H., Morgan, R., Abbott, T., Aguena, M., Allam, S., Annis, J., Avila, S., Bertin, E., Brooks, D., Burke, D., Carnero Rosell, A., Kind, M., Carretero, J., Conselice, C., Costanzi, M., Costa, L., & Wilkinson, R. (2020). Machine learning for searching the dark energy survey for trans-Neptunian objects. *Publications of the Astronomical Society of the Pacific*, *133*, 014501. https://doi.org/10.1088/1538-3873/abcaea

Herle, A., Channegowda, J., & Prabhu, D. (2020). Quasar detection using linear support vector machine with learning from mistakes methodology. In 2020 IEEE international conference on electronics, computing and communication technologies (CONECCT), pp. 1–6. https://doi.org/10.1109/CONECCT50063.2020.9198529.

Hernandez, E. J., Srinivasan, S., & Marshall, J. (2021). Automated classification of evolved stars using machine learning. *Bulletin of the AAS*, *53*(1), 1–2.

Hinners, T., Tat, K., & Thorp, R. (2017). Machine learning techniques for stellar light curve classification. *The Astronomical Journal*, *156*, 1–23. https://doi.org/10.3847/1538-3881/aac16d

Hon, M., Stello, D., & Yu, J. (2017). Deep learning classification in asteroseismology. *Monthly Notices of the Royal Astronomical Society*, *469*(4), 4578–4583. https://doi.org/10.1093/mnras/stx1174

Hosenie, Z., Lyon, R., Stappers, B., & Mootoovaloo, A. (2019). Comparing multi-class, binary and hierarchical machine learning classification schemes for variable stars. *Monthly Notices of the Royal Astronomical Society*, *488*(4), 4858–4872. https://doi.org/10.1093/mnras/stz1999

Hoyle, B. (2016). Measuring photometric redshifts using galaxy images and deep neural networks. *Astronomy and Computing*, *16*, 34–40.

Hoyle, B., Rau, M. M., Zitlau, R., Seitz, S., & Weller, J. (2015). Feature importance for machine learning redshifts applied to SDSS galaxies. *Monthly Notices of the Royal Astronomical Society*, *449*, 1275–1283.

Huertas-Company, M., Gravet, R., Cabrera-Vives, G., Pérez-González, P. G., Kartaltepe, J. S., Barro, G., Bernardi, M., Mei, S., Shankar, F., Dimauro, P., Bell, E. F., Kocevski, D., Koo, D. C., Faber, S. M., & Mcintosh, D. H. (2015). A catalog of visual-like morphologies in the 5 candels fields using deep learning. *The Astrophysical Journal Supplement Series*, *221*, 8.

Hui, J., Aragon, M., Cui, X., & Flegal, J. M. (2018). A machine learning approach to galaxy-LSS classification—I Imprints on halo merger trees. *Monthly Notices of the Royal Astronomical Society*, *475*, 4494–4503.

Ishida, E. E. O., Beck, R., González-Gaitán, S., de Souza, R. S., Krone-Martins, A., Barrett, J. W., Kennamer, N., Vilalta, R., Burgess, J. M., Quint, B., Vitorelli, A. Z., Mahabal, A., & Gangler, E. (2019). COIN collaboration, optimizing spectroscopic follow-up strategies for supernova photometric classification with active learning. *Monthly Notices of the Royal Astronomical Society*, *483*(1), 2–18. https://doi.org/10.1093/mnras/sty3015

Ivezic, Ž., Connelly, A. J., VanderPlas, J. T., & Gray, A. (2014). *Statistics, data mining and machine learning in astronomy*. Princeton University Press.

Jara-Maldonado, M., Alarcon-Aquino, V., Rosas-Romero, R., Starostenko, O., & Ramirez-Cortes, J. M. (2020). Transiting exoplanet discovery using machine learning techniques: A survey. *Earth Science Informatics*, *13*, 573–600. https://doi.org/10.1007/s12145-020-00464-7

Jin, X., Zhang, Y., Zhang, J., Zhao, Y., Wu, X.-B., & Fan, D. (2019). Efficient selection of quasar candidates based on optical and infrared photometric data using machine learning. *Monthly Notices of the Royal Astronomical Society*, *485*(4), 4539–4549. https://doi.org/10.1093/mnras/stz680

Jolliffe, I. (2002). *Principal component analysis*. Springer Verlag.

Kang, S.-J., Fan, J.-H., Mao, W., Wu, Q., Feng, J., & Yin, Y. (2019). Evaluating the optical classification of Fermi BCUs using machine learning. *The Astrophysical Journal*, *872*, 189.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10–25.

Khalil, M., Said, M., Osman, H., Ahmed, B., Ahmed, D., Younis, N., Maher, B., Osama, M., & Ashmawy, M. (2019). Big data in astronomy: From evolution to revolution. *International Journal of Advanced Astronomy*, *7*, 11. https://doi.org/10.14419/ijaa.v7i1.18029

Kochi, C., Nakagawa, T., Isobe, N., Shirahata, M., Yano, K., & Baba, S. (2017). Study of spectral energy distribution of galaxies with principal component analysis. *Publications of The Korean Astronomical Society*, *32*(1), 209–211.

Kremer, J., Stensbo-Smidt, K., Gieseke, F., Pedersen, K. S., & Igel, C. (2017). Big universe, big data: Machine learning and image analysis for astronomy. *IEEE Intelligent Systems*, *32*(2), 16–22. https://doi.org/10.1109/MIS.2017.40

Kriessler, J. R., Han, E. H., Odewahn, S. C., & Beers, T. C. (1998). Automated morphological classification of galaxies and the morphologydensity relation. *American Astronomical Society Meeting Abstracts*, *193*, 38.20.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 25, pp. 1097–1105). Curran Associates, Inc.

Ksoll, V. F., Gouliermis, D. A., Klessen, R. S., Grebel, E. K., Sabbi, E., Anderson, J., Lennon, D., Cignoni, M., de Marchi, G., Smith, L., Tosi, M., & van der Marel, R. P. (2018). Hubble tarantula Treasury project—VI. Identification of pre-main-sequence stars using machine-learning techniques. *Monthly Notices of the Royal Astronomical Society*, *479*, 2389–2414.

Kuminski, E., & Shamir, L. (2018). A hybrid approach to machine learning annotation of large galaxy image databases. *Astronomy and Computing*, *25*, 257–269.

Kuntzer, T., & Courbin, F. (2017). Detecting unresolved binary stars in Euclid VIS images. *Astronomy & Astrophysics*, *606*, A119.

Kuo, B.-C., Ho, H.-H., Li, C.-H., Hung, C.-C., & Taur, J.-S. (2014). A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*, 317–326. https://doi.org/10.1109/JSTARS.2013.2262926

Lahav, O., Nairn, A., Sodré, L., & Storrie-Lombardi, M. C. (1996). Neural computation as a tool for galaxy classification: Methods and examples. *Monthly Notices of the Royal Astronomical Society*, *283*, 207–221.

LeCun, Y. (2018). The Power and limits of deep learning. *Research-Technology Management*, *61*(6), 22–27. https://doi.org/10.1080/08956308.2018.1516928

Leydesdorff, L., & Bornmann, L. (2012). Mapping (USPTO) patent data using overlays to Google maps. *Journal of the American Society for Information Science and Technology*, *63*(7), 1442–1458.

Li, C., Zhang, Y. X., Cui, C., Fan, D., Zhao, Y., Wu, X.-B., Zhang, J.-Y., Han, J., Xu, Y., Tao, Y., Li, S., & He, B. (2021). Photometric redshift estimation of BASS DR3 quasars by machine learning. *Monthly Notices of the Royal Astronomical Society*, *509*, 2289–2303. stab3165. https://doi.org/10.1093/mnras/stab3165

Li, K., Rollins, J., & Yan, E. (2018). Web of science use in published research and review papers 1997–2017: A selective, dynamic, cross-domain, content-based analysis. *Scientometrics*, *115*(1), 1–20.

Li, Y.-B., Luo, A.-L., Du, C.-D., Zuo, F., Wang, M.-X., Gang, Z., Jiang, B., Zhang, H., Liu, C., Qin, L., Wang, R., Du, B., Guo, Y.-X., Wang, B., Han, Z.-W., Xiang, M.-S., Chen, B., Chen, J.-J., & Zhao, Y. H. (2017). Carbon stars identified from LAMOST DR4 using machine learning. *The Astrophysical Journal Supplement Series*, *234*, 1–21. https://doi.org/10.3847/1538-4365/aaa415

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, J., Nichol, R., Szalay, A., Andreescu, D., Murray, P., & Vandenberg, J. (2008). Galaxy zoo: Morphologies derived from visual inspection of galaxies from the Sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, *389*, 1179–1189.

Liu, J., Ye, Y., Shen, C., Wang, Y., & Erdélyi, R. (2018). A new tool for CME arrival time prediction using machine learning algorithms: CAT-PUMA. *The Astrophysical Journal*, *855*, 109.

Longo, G., Merényi, E., & Tino, P. (2019). Foreword to the focus issue on machine intelligence in astronomy and astrophysics. *Publications of the Astronomical Society of the Pacific*, *131*, 100101. https://doi.org/10.1088/1538-3873/ab2743

Lucie-Smith, L., Peiris, H., & Pontzen, A. (2019). An interpretable machine-learning framework for dark matter halo formation. *Monthly Notices of the Royal Astronomical Society*, *490*, 331–342. https://doi.org/10.1093/mnras/stz2599

Lucie-Smith, L., Peiris, H., Pontzen, A., & Lochner, M. (2018). Machine learning cosmological structure formation. *Monthly Notices of the Royal Astronomical Society*, *479*, 3405–3414. https://doi.org/10.1093/mnras/sty1719

Lukic, V., de Gasperin, F., & Brüggen, M. (2020). ConvoSource: Radio-astronomical source-finding with convolutional neural networks. *Galaxies*, *8*, 3. https://doi.org/10.3390/galaxies8010003

Ma, R., Filali-Boubrahimi, S., Hamdi, S. M., & Angryk, R. (2017). Solar flare prediction using multivariate time series decision trees. *IEEE Big Data*, 2569–2578. https://doi.org/10.1109/BigData.2017.8258216

Ma, Z., Xu, H., Zhu, J., Hu, D., Li, W., Shan, C., Zhu, Z., Gu, L., Li, J., Liu, C., & Wu, X. (2019). A machine learning based morphological classification of 14,245 radio AGNs selected from the best-Heckman sample. *Astrophysical Journal Supplement*, *240*, 34.

Mackereth, J., Leung, H. W., Bovy, J., & Miglio, A. (2021). The stellar age revolution with Asteroseismology and machine learning. *Bulletin of the AAS*, *53*(6), 1–2.

Magro, D., Adami, K., DeMarco, A., Riggi, S., & Sciacca, E. (2021). A comparative study of convolutional neural networks for the detection of strong gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, *505*(4), 6155–6165. https://doi.org/10.1093/mnras/stab1635

Mahabal, A., Sheth, K., Gieseke, F., Pai, A., Djorgovski, G., Drake, A., & Graham, M. (2017). Deep-learnt classification of light curves. *IEEE Symposium Series on Computational Intelligence (SSCI)*, *2017*, 1–8. https://doi.org/10.1109/SSCI.2017.8280984

Mao, W., Lv, X., Kang, S., Zhu, K., & Hu, H. (2020). Research and characterization of blazar candidates among the Fermi/LAT4FGL catalog using machine learning. In: 2020 IEEE 9th joint international information technology and artificial intelligence conference (ITAIC), pp. 1916–1920. https://doi.org/10.1109/ITAIC49862.2020.9339103.

Marchetti, T., Rossi, E. M., Kordopatis, G., Brown, A. G. A., Rimoldi, A. S., Youakim, E. K., & Ashley, R. (2017). An artificial neural network to discover hypervelocity stars: Candidates in Gaia DR1/TGAS. *Monthly Notices of the Royal Astronomical Society*, *470*, 1388–1403.

Márquez-Neila, P., Fisher, C., Sznitman, R., & Heng, K. (2018). Supervised machine learning for analysing spectra of exoplanetary atmospheres. *Nature Astronomy*, *2*, 719–724.

McLeod, M., Libeskind, N., Lahav, O., & Hoffman, Y. (2017). Estimating the mass of the local group using machine learning applied to numerical simulations. *Journal of Cosmology and Astro-Particle Physics*, *2017*, 34.

Meshcheryakov, A. V., Glazkova, V. V., Gerasimov, S. V., & Mashechkin, I. V. (2018). Measuring the probabilistic photometric redshifts of X-ray quasars based on the quantile regression of ensembles of decision trees. *Astronomy Letters*, *44*, 735–753.

Metcalf, R. B., Meneghetti, M., Avestruz, C., Bellagamba, F., Bom, C. R., Bertin, E., Cabanac, R., Courbin, F., Davies, A., Decencière, E., Flamary, R., Gavazzi, R., Geiger, M., Hartley, P., Huertas-Company, M., Jackson, N., Jacobs, C., Jullo, E., … Vernardos, G. (2019). The strong gravitational lens finding challenge. *Astronomy & Astrophysics*, *625*, A119.

Miettinen, O. (2018). Protostellar classification using supervised machine learning algorithms. *Astrophysics and Space Science*, *363*, 197.

Mirzoyan, S., Khachatryan, H., Yegorian, G., & Gurzadyan, V. (2019). Machine learning and Kolmogorov analysis to reveal gravitational lenses. *Monthly Notices of the Royal Astronomical Society: Letters*, *489*(1), L32–L36. https://doi.org/10.1093/mnrasl/slz125

Mishra, A., & Reddy, P. (2020). Baryon density extraction and isotropy analysis of cosmic microwave background using deep learning. *Machine Learning: Science and Technology*, *1*, 1–9. https://doi.org/10.1088/2632-2153/abab63

Mislis, D., Pyrzas, S., & Alsubai, K. A. (2018). TSARDI: A machine learning data rejection algorithm for transiting exoplanet light curves. *Monthly Notices of the Royal Astronomical Society*, *481*, 1624–1630.

Morello, G., Morris, P. W., Van Dyk, S. D., Marston, A. P., & Mauerhan, J. C. (2018). Applications of machine-learning algorithms for infrared colour selection of galactic Wolf-Rayet stars. *Monthly Notices of the Royal Astronomical Society*, *473*, 2565–2574.

Morgan, R., Soares-Santos, M., Annis, J., Herner, K., Garcia, A., Palmese, A., Drlica-Wagner, A., Kessler, R., García-Bellido, J., Sherman, T. G., Allam, S., Bechtol, K. C., Bom, C. R., Brout, D., Butler, R. E., Butner, M., Cartier, R., Chen, H., Conselice, C. J., … Tarlé, G. G. (2020). Constraints on the physical properties of GW190814 through simulations based on DECam follow-up observations by the dark energy survey. *The Astrophysical Journal*, *901*(1), 83.

Morice-Atkinson, X., Hoyle, B., & Bacon, D. (2018). Learning from the machine: Interpreting machine learning algorithms for point- and extended-source classification. *Monthly Notices of the Royal Astronomical Society*, *481*(3), 4194–4205.

Morrison, D. (2005). *Multivariate statistical methods* (4th ed.). McGraw-Hill.

Mucesh, S., Hartley, W. G., Palmese, A., Lahav, O., Whiteway, L., Bluck, A. F. L., Alarcon, A., Amon, A., Bechtol, K., Bernstein, G. M., Carnero Rosell, A., Carrasco Kind, M., Choi, A., Eckert, K., Everett, S., Gruen, D., Gruendl, R. A., Harrison, I., Huff, E. M., … Wilkinson, R. D. (2021). A machine learning approach to galaxy properties: Joint redshift–stellar mass probability distributions with random Forest. *Monthly Notices of the Royal Astronomical Society*, *502*(2), 2770–2786.

Müller, A., Hackstein, M., Greiner, M., Frank, P., Bomans, D. J., Dettmar, R.-J., & Enßlin, T. (2018). Sharpening up galactic all-sky maps with complementary data. A machine learning approach. *Astronomy & Astrophysics*, *620*, A64.

Nadler, E. O., Mao, Y.-Y., Wechsler, R. H., Garrison-Kimmel, S., & Wetzel, A. (2018). Modeling the impact of baryons on subhalo populations with machine learning. *The Astrophysical Journal*, *859*, 129.

Naul, B., Bloom, J. S., Pérez, F., & van der Walt, S. (2018). A recurrent neural network for classification of unevenly sampled variable stars. *Nature Astronomy*, *2*, 151–155.

Nielsen, M. B., Ball, W. H., Standing, M. R., Triaud, A. H. M. J., Buzasi, D., Carboneau, L., Stassun, K. G., Kane, S. R., Chaplin, W. J., Bellinger, E. P., Mosser, B., Roxburgh, I. W., Orhan, Z. Ç., Yldz, M., Örtel, S., Vrard, M., Mazumdar, A., Ranadive, P., Deal, M., … Serenelli, A. (2020). Tess asteroseismology of the known planet host star 2 Fornacis. *Astronomy and Astrophysics*, *641*, A25. https://doi.org/10.1051/0004-6361/202037461

Nolte, A., Wang, L., Bilicki, M., Holwerda, B., & Biehl, M. (2019). Galaxy classification: A machine learning analysis of GAMA catalogue data. *Neurocomputing*, *342*, 172–190. https://doi.org/10.1016/j.neucom.2018.12.076

Orduña Malea, E., Martín-Martín, A., & Delgado-López-Cózar, E. (2017). Google scholar as a source for scholarly evaluation: A bibliographic review of database errors. *Revista Española de Documentación Científica*, *40*(4), 1–33.

Ostrovski, F., McMahon, R. G., Connolly, A. J., Lemon, C. A., Auger, M. W., Banerji, M., Hung, J. M., Koposov, S. E., Lidman, C. E., Reed, S. L., Allam, S., Benoit-Lévy, A., Bertin, E., Brooks, D., Buckley-Geer, E., Rosell, A. C., Kind, M. C., Carretero, J., Cunha, C. E., … Walker, A. R. (2017). VDES J2325-5229 a z = 2.7 gravitationally lensed quasar discovered using morphology-independent supervised machine learning. *Monthly Notices of the Royal Astronomical Society*, *465*, 4325–4334.

Owens, E. A., Griffiths, R. E., & Ratnatunga, K. U. (1996). Using oblique decision trees for the morphological classification of galaxies. *Monthly Notices of the Royal Astronomical Society*, *281*, 153–157.

Pang, D., Goseva-Popstojanova, K., Devine, T., & McLaughlin, M. (2018). A novel single-pulse search approach to detection of dispersed radio pulses using clustering and supervised machine learning. *Monthly Notices of the Royal Astronomical Society*, *480*, 3302–3323.

Panigrahi, S., Pattanayak, R. M., Sethy, P. K., & Behera, S. A. (2021). Forecasting of sunspot time series using a hybridization of ARIMA, ETS and SVM methods. *Solar Physics*, *296*, 6. https://doi.org/10.1007/s11207-020-01757-2

Papageorgiou, A., Catelan, M., Christopoulou, P.-E., Drake, A. J., & Djorgovski, S. G. (2018). An updated catalog of 4680 northern eclipsing binaries with Algol-type light-curve morphology in the Catalina sky surveys. *Astrophysical Journal Supplement*, *238*, 4.

Pashchenko, I., Sokolovsky, K., & Gavras, P. (2018). Machine learning search for variable stars. *Monthly Notices of the Royal Astronomical Society*, *475*, 2326–2343. https://doi.org/10.1093/mnras/stx3222

Pearson, K. A., Palafox, L., & Griffith, C. A. (2018). Searching for exoplanets using artificial intelligence. *Monthly Notices of the Royal Astronomical Society*, *474*, 478–491.

Petrillo, C. E., Tortora, C., Vernardos, G., Koopmans, L. V. E., Kleijn, G. V., Bilicki, M., Napolitano, N., Chatterjee, S., Covone, G., Dvornik, A., Erben, T., Getman, F., Giblin, B., Heymans, C., Jong, J. T. A., Kuijken, K., Schneider, P., Shan, H., Spiniello, C., & Wright, A. H. (2019). LinKS: Discovering galaxy-scale strong lenses in the kilo-degree survey using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, *484*, 3879–3896.

Poczos, B. (2018). Machine learning algorithms for matching theories, simulations, and observations in cosmology. United States. https://doi.org/10.2172/1572709.

Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Bachrach, D. G. (2008). Scholarly influence in the field of management: A bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century. *Journal of Management*, *34*(4), 641–720.

Powell, J., Torres-Forné, A., Lynch, R., Trifirò, D., Cuoco, E., Cavaglià, M., Heng, I., & Font, J. A. (2017). Classification methods for noise transients in advanced gravitational-wave detectors II: Performance tests on advanced LIGO data. *Classical and Quantum Gravity*, *34*, 034002.

Qahwaji, R., & Colak, T. (2007). Automatic short-term solar flare prediction using machine learning and sunspot associations. *Solar Physics*, *241*, 195–211.

Qin, D.-M., Guo, P., Hu, Z., & Zhao, Y. (2003). Automated separation of stars and Normal galaxies based on statistical mixture modeling with RBF neural networks. *Chinese Journal of Astronomy and Astrophysics*, *3*, 277. https://doi.org/10.1088/1009-9271/3/3/277

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81–106.

Rafieferantsoa, M., Andrianomena, S., & Davé, R. (2018). Predicting the neutral hydrogen content of galaxies from optical data using machine learning. *Monthly Notices of the Royal Astronomical Society*, *479*, 4509–4525.

Razim, O., Cavuoti, S., Brescia, M., Riccio, G., Salvato, M., & Longo, G. (2021). Improving the reliability of photometric redshift with machine learning. *Monthly Notices of the Royal Astronomical Society*, *507*(4), 5034–5052. https://doi.org/10.1093/mnras/stab2334

Reza, M. (2021). Galaxy morphology classification using automated machine learning. *Astronomy and Computing*, *37*, 100492.

Richards, J. W., Starr, D., Butler, N., Bloom, J., Brewer, J., Crellin-Quick, A., Higgins, J., Kennedy, R., & Rischard, M. (2011). On machine-learned classification of variable stars with sparse and Noisy time-series data. *The Astrophysical Journal*, *733*, 10. https://doi.org/10.1088/0004-637X/733/1/10

Rodríguez, A. C., Kacprzak, T., Lucchi, A., Amara, A., Sgier, R., Fluri, J., Hofmann, T., & Réfrégier, A. (2018). Fast cosmic web simulations with generative adversarial networks. *Computational Astrophysics and Cosmology*, *5*, 4.

Rokach, L., & Maimon, O. (2008). *Data mining with decision trees. Theory and applications*. World Scientific Pub Co Inc. https://doi.org/10.1142/9789812771728_0001

Rosenblatt, F. (1957). *The perceptron—A perceiving and recognizing automaton*. Cornell Aeronautical Laboratory.

Saha, S., Basak, S., Safonova, M., Bora, K., Agrawal, S., Sarkar, P., & Murthy, J. (2018). Theoretical validation of potential habitability via analytical and boosted tree methods: An optimistic study on recently discovered exoplanets. *Astronomy and Computing*, *23*, 141–150.

Salimans, T., & Kingma D. P. (2016). Weight normalization: A simple reparameterization to accelerate training of deep neural networks. Advances in neural information processing systems 29 (NIPS 2016). arXiv:1602.07868.

Sánchez, B., Domínguez, M. J. R., Lares, M., Beroiz, M., Cabral, J. B., Gurovich, S., Quiñones, C., Artola, R., Colazo, C., Schneiter, M., Girardini, C., Tornatore, M., Nilo Castellón, J. L., García Lambas, D., & Díaz, M. C. (2018). Machine learning on difference image analysis: A comparison of methods for transient detection. *Astronomy and Computing*, *28*, 100284.

Schaefer, C., Geiger, M., Kuntzer, T., & Kneib, J.-P. (2017). Deep convolutional neural networks as strong gravitational lens detectors. *Astronomy & Astrophysics*, *611*, 1–9. https://doi.org/10.1051/0004-6361/201731201

Schindler, J.-T., Fan, X., McGreer, I. D., Yang, Q., Wu, J., Jiang, L., & Green, R. (2017). The extremely luminous quasar survey in the SDSS footprint. I. Infrared-based candidate selection. *The Astrophysical Journal*, *851*, 13.

Sevilla-Noarbe, I., & Etayo-Sotos, P. (2015). Effect of training characteristics on object classification: An application using boosted decision trees. *Astronomy and Computing*, *11*, 64–72.

Shallue, C. J., & Vanderburg, A. (2018). Identifying exoplanets with deep learning: A five-planet resonant chain around Kepler-80 and an eighth planet around Kepler-90. *The Astronomical Journal*, *155*, 94.

Sharma, K., Singh, H., Gupta, R., Kembhavi, A., Vaghmare, K., Shi, J., Zhao, Y., Zhang, J., & Wu, Y. (2020). Stellar spectral interpolation using machine learning. *Monthly Notices of the Royal Astronomical Society*, *496*(4), 5002–5016. https://doi.org/10.1093/mnras/staa1809

Shnurenko, I., Murovana, T., & Kushchu, I. (2020). *Artificial Intelligence*. UNESCO Institute for Information Technologies.

Singh, H. P., Gulati, R. K., & Gupta, R. (1998). Stellar spectral classification using principal component analysis and artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, *295*, 312–318.

Singh, K. K., Dhar, V. K., & Meintjes, P. J. (2022). Artificial neural networks for cosmic gamma-ray propagation in the universe. *New Astronomy*, *91*, 101701.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science*, *24*(4), 265–269.

Song, M., & Ding, Y. (2014). Topic modeling: Measuring scholarly impact using a topical lens. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact* (pp. 235–257). Springer.

Soszyński, I., Udalski, A., Szymański, M. K., Wyrzykowski, L., Ulaczyk, K., Poleski, R., Pietrukowicz, P., Kozlowski, S., Skowron, D., Skowron, J., Mróz, P., Rybicki, K., & Iwanek, P. (2018). The OGLE collection of variable stars. Type II cepheids in the magellanic system. *Acta Astronomica*, *68*, 89–109.

Stephens, Z., Lee, S., Faghri, F., Campbell, R., Zhai, C., Efron, M., Iyer, R., Schatz, M., Sinha, S., & Robinson, G. (2015). Big Data: Astronomical or genomical? *PLoS Biology*, *13*, e1002195. https://doi.org/10.1371/journal.pbio.1002195

Stephon, A., Sergei, G., Evan, M. D., Michael, W., & Toomey, E. U. (2020). Deep learning the morphology of dark matter substructure. *The Astrophysical Journal*, *1*, 15.

Sunitha, P., Venkatesh, N., & Atreya, M. (2020). Support vector machine method to identify and classify astronomical objects. *International Journal of Engineering Applied Sciences and Technology*, *5*, 209–215. https://doi.org/10.33564/IJEAST.2020.v05i04.029

Sversut, A. B., & Neto, P. (2020). Machine learning applied to multifrequency data in astrophysics: Blazar classification. *Monthly Notices of the Royal Astronomical Society*, *498*, 1750–1764. https://doi.org/10.1093/mnras/staa2449

Takalo, J., & Mursula, K. (2018). Principal component analysis of sunspot cycle shape. *Astronomy and Astrophysics*, *620*, 1–10. https://doi.org/10.1051/0004-6361/201833924

Tan, C. M., Lyon, R. J., Stappers, B. W., Cooper, S., Hessels, J. W. T., Kondratiev, V. I., Michilli, D., & Sanidas, S. (2018). Ensemble candidate classification for the LOTAAS pulsar survey. *Monthly Notices of the Royal Astronomical Society*, *474*, 4571–4583.

Teimoorinia, H., Toyonaga, R., Fabbro, S., & Bottrell, C. (2020). Comparison of multi-class and binary classification machine learning models in identifying strong gravitational lenses. *Publications of the Astronomical Society of the Pacific*, *132*, 1010.

Thongpapanl, N. T. (2012). The changing landscape of technology and innovation management: An updated ranking of journals in the field. *Technovation*, *32*(5), 257–271.

Tiwari, A. (2021). Predicting arrival time for CMEs: Machine learning and ensemble methods. https://doi.org/10.5194/egusphere-egu21-7661.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, *9*(1), 1–12.

Ucci, G., Ferrara, A., Gallerani, S., & Pallottini, A. (2017). Inferring physical properties of galaxies from their emission-line spectra. *Monthly Notices of the Royal Astronomical Society*, *465*, 1144–1156.

Vajente, G., Huang, Y., Isi, M., Driggers, J. C., Kissel, J. S., Szczepańczyk, M. J., & Vitale, S. (2020). Machine-learning nonstationary noise out of gravitational-wave detectors. *Physical Review D*, *101*, 042003.

Valderrama-Zurián, J. C., Aguilar-Moya, R., Melero-Fuentes, D., & Aleixandre-Benavent, R. (2015). A systematic analysis of duplicate records in Scopus. *Journal of Informetrics*, *9*(3), 570–576.

Van Eck, N., Waltman, L., Noyons, E., & Buter, R. (2010). Automatic term identification for bibliometric mapping. *Scientometrics*, *82*(3), 581–596.

Van Eck, N. J., & Waltman, L. (2007). VOS: A new method for visualizing similarities between objects. In H.-J. Lenz & R. Decker (Eds.), *Advances in data analysis* (pp. 299–306). Springer.

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538.

Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact* (pp. 285–320). Springer.

van Roestel, J., Kupfer, T., Ruiz-Carmona, R., Groot, P. J., Prince, T. A., Burdge, K., Laher, R., Shupe, D. L., & Bellm, E. (2018). Discovery of 36 eclipsing EL CVn binaries found by the Palomar transient factory. *Monthly Notices of the Royal Astronomical Society*, *475*, 2560–2590.

Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., Spergel, D. N., Somerville, S., Davé, R., Pillepich, A., Hernquist, L., Nelson, D., Torrey, P., Narayanan, D., Li, Y., Philcox, O. H., La Torre, V., Maria Delgado, A., Ho, S., Hassan, S., Burkhart, B., Wadekar, D., … Bryan, G. L. (2021). The CAMELS project: Cosmology and astrophysics with machine-learning simulations. *The Astrophysical Journal*, *915*, 1–34.

Viquar, M., Basak, S., Dasgupta, A., Agrawal, S., & Saha, S. (2019). Machine learning in astronomy: A case study in quasar-star classification. In A. Abraham, P. Dutta, J. Mandal, A. Bhattacharya, & S. Dutta (Eds.), *Emerging Technologies in Data Mining and Information Security. Advances in intelligent systems and computing* (Vol. 814). Springer. https://doi.org/10.1007/978-981-13-1501-5_72

Visser, M., Van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, Crossref, and Microsoft academic. *Quantitative Science Studies*, *2*(1), 20–41.

Wadadekar, Y. (2004). Estimating photometric redshifts using support vector machines. *Publications of the Astronomical Society of the Pacific*, *117*, 79.

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392.

Wang, K., Guo, P., & Luo, A.-L. (2016). A new automated spectral feature extraction method and its application in spectral classification and defective spectra recovery. *Monthly Notices of the Royal Astronomical Society*, *465*, 4311–4324. https://doi.org/10.1093/mnras/stw2894

Wang, K., Guo, P., Yu, F., Duan, L., Wang, Y., & Du, H. (2018). Computational intelligence in astronomy: A survey. *International Journal of Computational Intelligence Systems*, *11*, 575. https://doi.org/10.2991/ijcis.11.1.43

Wang, Q.-J., Li, J.-C., & Guo, L.-Q. (2021). Solar cycle prediction using a long short-term memory deep learning model. *Research in Astronomy and Astrophysics*, *21*, 12. https://doi.org/10.1088/1674-4527/21/1/12

Wang, R., Luo, A.-L., Chen, J.-J., Hou, W., Zhang, S., Zhao, Y. H., Li, X., & Hou, Y.-H. (2020). SPCANet: Stellar parameters and chemical abundances network for LAMOST-II medium resolution survey. *The Astrophysical Journal*, *891*, 23. https://doi.org/10.3847/1538-4357/ab6dea

Weir, N., Fayyad, U. M., Djorgovski, S. G., & Roden, J. (1995). The SKICAT system for processing and analyzing digital imaging sky surveys. *Publications of the Astronomical Society of the Pacific*, *107*, 1243.

Whitmore, B. C. (1984). An objective classification system for spiral galaxies. I. the two dominant dimensions. *The Astrophysical Journal*, *278*, 61.

WorldBank. (2021a). World Bank Group. In: Country Income Classifications (Internet). http://www.worldbank.org/

WorldBank. (2021b). World Bank Group. In: Gross Domestic Product 2021 (Internet). http://www.worldbank.org/

Woźniak, P. R., Williams, S. J., Vestrand, W. T., & Gupta, V. (2004). Identifying red variables in the northern sky variability survey. *The Astronomical Journal*, *128*, 2965–2976.

Xiang, M.-S., Liu, X.-W., Shi, J.-R., Yuan, H.-B., Huang, Y., Luo, A.-L., Zhang, H.-W., Zhao, Y.-H., Zhang, J.-N., Ren, J.-J., Chen, B.-Q., Wang, C., Li, J., Huo, Z.-Y., Zhang, W., Wang, J.-L., Zhang, Y., Hou, Y.-H., & Wang, Y.-F. (2017). Estimating stellar atmospheric parameters, absolute magnitudes and elemental abundances from the LAMOST spectra with kernel-based principal component analysis. *Monthly Notices of the Royal Astronomical Society*, *464*(3), 3657–3678.

Yan, E., Ding, Y., & Jacob, E. K. (2012). Overlaying communities and topics: An analysis on publication networks. *Scientometrics*, *90*(2), 499–513.

Yip, C. W., Connolly, A. J., Berk, D. E. V., Ma, Z., Frieman, J. A., SubbaRao, M., Szalay, A. S., Richards, G. T., Hall, P. B., Schneider, D. P., Hopkins, A. M., Trump, J., & Brinkmann, J. (2004). Spectral classification of quasars in the Sloan digital sky survey: Eigenspectra, redshift, and luminosity effects. *The Astronomical Journal*, *128*, 2603.

Zhang, J., Zhang, Y., & Zhao, Y. (2018). Imbalanced learning for RR Lyrae stars based on SDSS and GALEX databases. *The Astronomical Journal*, *155*, 108.

Zhang, K., Schlegel, D., Andrews, B., Comparat, J., Schäfer, C., Mata, J., Kneib, J.-P., & Yan, R. (2020). Machine learning classifiers for intermediate redshift emission line galaxies. *The Astrophysical Journal*, *883*(1), 63.

Zhang, Y., & Zhao, Y. (2015). Astronomy in the big data era. *Data Science Journal*, *14*, 11.

Zharkova, V., Shepherd, S., & Zharkov, S. (2012). Principal component analysis of background and sunspot magnetic field variations during solar cycles 21–23. *Monthly Notices of the Royal Astronomical Society*, *424*, 2943–2953. https://doi.org/10.1111/j.1365-2966.2012.21436.x

Zhu, B., Liu, J. Z., Rosen, B., & Rosen, M. (2018). Image reconstruction by domain transform manifold learning. *Nature*, *555*, 487–492. https://doi.org/10.1038/nature25988

Zhu, J., & Liu, W. (2020). A tale of two databases: The use of web of science and scopus in academic papers. *Scientometrics*, *123*, A100.

How to cite this article: Rodríguez, J.-V., Rodríguez-Rodríguez, I., & Woo, W. L. (2022). On the application of machine learning in astronomy and astrophysics: A text-mining-based scientometric analysis. *WIREs Data Mining and Knowledge Discovery*, *12*(5), e1476. https://doi.org/10.1002/widm.1476