Maritime Policy & Management

The flagship journal of international shipping and port research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmpm20

Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions

Ziaul Haque Munim , Mariia Dushenko , Veronica Jaramillo Jimenez , Mohammad Hassan Shakil & Marius Imset

MARITIME POLICY & MANAGEMENT
2020, VOL. 47, NO. 5, 577–597
https://doi.org/10.1080/03088839.2020.1788731

Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions

Ziaul Haque Munim[a], Mariia Dushenko[a], Veronica Jaramillo Jimenez[a], Mohammad Hassan Shakil[b] and Marius Imset[a]

[a]Faculty of Technology, Natural and Maritime Sciences, University of South-Eastern Norway, Horten, Norway; [b]Taylor's Business School, Taylor's University, Subang Jaya, Malaysia

ABSTRACT

This study provides a bibliometric review of 279 studies on the applications of big data and artificial intelligence (AI) in the maritime industry, published in 214 academic outlets, authored by 842 scholars. We extracted bibliographical data from the Web of Science database and analysed it using the Bibliometrix tool in R software. Based on citation analysis metrics, we revealed the most influential articles, journals, authors and institutions. Using the bibliographic coupling methodology, we identified four underlying research clusters: (1) digital transformation in maritime industry, (2) applications of big data from AIS, (3) energy efficiency and (4) predictive analytics. We analysed these clusters in detail and extracted future research questions. Besides, we present research collaboration networks on the institution and author level.

KEYWORDS

Big data; artificial intelligence; bibliometrix; literature review; machine learning; shipping

1. Introduction

Big data and artificial intelligence (AI) are crucial components of data-driven decision-making in most industries (Liang and Liu 2018). The maritime industry is one of the oldest and traditional industries to still rely more on intuition than on data, due to the vast size of network and planning problems (Brouer, Karsten, and Pisinger 2016). Big data and AI have received considerable attention in recent years, through a number of publications, and some scholars have portrayed the concept of 'big data' as hype (d'Amore, Baggio, and Valdani 2015). The term big data is typically used to denote large amounts of data. With the recent burst of data volume, researchers have been continuously scrutinising novel techniques for analysing big data (Franks 2012). A branch of these techniques is now integrated into the concept of 'AI'.

AI research initially aimed to mimic human decision-making by utilising a large volume of data using machines. Nowadays, AI is capable of doing things that were impossible a decade ago. For example, sophisticated AI systems introduce autonomous ships, which can operate independently without human interaction, and the error rate is lower than that of human-operated ships. AI is gradually transforming the traditional operational process of the maritime industry. Consequently, the amount of research on the application of big data and AI has increased significantly since 2012 (Liang and Liu 2018). Following this trend, data-centric innovative technologies and new business models are being developed (Munim 2019). This transformation is reshaping the maritime industry, providing new opportunities to improve productivity, efficiency and sustainability (Heilig, Lalla-Ruiz, and Voß 2017).

CONTACT Ziaul Haque Munim ziaul.h.munim@usn.no Faculty of Technology, Natural and Maritime Sciences, University of South-Eastern Norway, Horten 3184, Norway

Studies on the synthesis of big data application in maritime are rare, which has created a gap in the academic literature due to the importance of big data and AI in maritime operations (Yang et al. 2019; Mirović, Miličević, and Obradović 2018). Big data- and AI-enhanced maritime operations can contribute to the economic and environmental aspects of the maritime business (Sanchez-Gonzalez et al. 2019). Maritime trade accounts for approximately 80% of world trade (UNCTAD 2018) and the industry faces many challenges due to its vastness (Brouer, Karsten, and Pisinger 2016) as well as continuously evolving regulatory requirements (Lee, Kwon, and Ruan 2019). Big data and AI offer viable solutions to some of these challenges. For example, data about ship performance and navigation systems can help shipping firms monitor vessels' performance and take necessary steps to improve the operational efficiency of the vessels (Mirović, Miličević, and Ines 2018). The industry generates large amounts of data that, if appropriately utilised in decision-making, can improve maritime safety, reduce environmental impacts and minimise cost.

To the best of our knowledge, in the maritime context there have been two review studies on big data (Yang et al. 2019; Mirović, Miličević, and Ines 2018) and two on digitalisation (Sanchez-Gonzalez et al. 2019; Fruth and Teuteberg 2017). The present study is more comprehensive than previous studies in terms of quality and spread of included studies that use big data and AI in the maritime context. For instance, Yang et al. (2019) reviewed studies that use only automatic identification systems (AIS) data. Unlike the present study, Mirović, Miličević, and Obradović (2018) did not follow a systematic approach to literature selection, which can lead to biased findings. Fruth and Teuteberg (2017) and Sanchez-Gonzalez et al. (2019) explicitly focused on digitalisation, although both used big data in their keyword search. While Fruth and Teuteberg (2017) did not include AI aspects in the maritime domain, Sanchez-Gonzalez et al. (2019) literature search process is rather abstract and may not be reproducible.

Unlike previous review studies, the literature search process in the present study was robust, transparent and reproducible. We review published studies that deal with big data and AI applications within the maritime context to map the conceptual structure of the field and identify future research avenues. Hence, we address four research objectives. The first is to find the existence of the big data and AI research in maritime as a standalone research domain. The second is to identify the key journals, articles, institutions and authors within this research domain and find the collaborative network of universities and authors. The third is to map the conceptual structure of big data and AI research in maritime by identifying and exploring underlying research clusters. The final objective is to extract and present the avenues for future research.

The findings of this study have several academic and industry implications. For the scholars and practitioners interested in big data and AI research in maritime, it provides a comprehensive overview of the research domain that introduce readers with the key studies, authors, universities, concepts and methods. Maritime firms and regulatory authorities can use the identified concepts and methods to enhance coordination among major players, optimise resource use, and improve environmental performance and navigational safety.

The remainder of this study is as structured as follows. Section 2 describes the research methodology. Section 3 presents the bibliometric analysis and findings. Section 4 exhibits research clusters, sub-clusters and future research directions. Finally, Section 5 concludes with a critical reflection.

2. Research methodology

Bibliometric analysis is a research methodology that has been widely applied in the field of library and information science studies and uses statistical tools to analyse published academic studies (Liang and Liu 2018). Bibliometrics includes several descriptive statistics of citation data, and network analysis of authors, journals, universities, countries and keywords based on citations and frequency analysis techniques. It supports the identification of research clusters, provides insights into current research interests and reveals trends for emerging topics in a field. We used a four-step approach (see Figure 1), starting with (1) data collection through systematic literature search, comprehensive evaluation of the field by (2a) bibliometric citation analysis and (2b) network analysis aiming to identify publication trends, the most influential journals, studies, institutions and authors, as well as collaborations and relationships. The next steps were (3) bibliographic coupling to identify research clusters and (4a) cluster analysis to map the sub-cluster system. The final step (4b) synthesises findings and discovers potential research directions.

2.1. Bibliography data extraction

The basis for our review is a collection of bibliography data from the most renowned academic database—ISI Web of Science (WoS). To collect relevant bibliography data, we perform a *keyword search in the WoS database* in October, 2019. Table 1 shows the detailed keyword search process using the Boolean function. The final keyword (i.e. Step 8) includes two parts. The first part ('artificial intelligence' OR 'big data' OR 'business intelligence' OR 'data analytics' OR

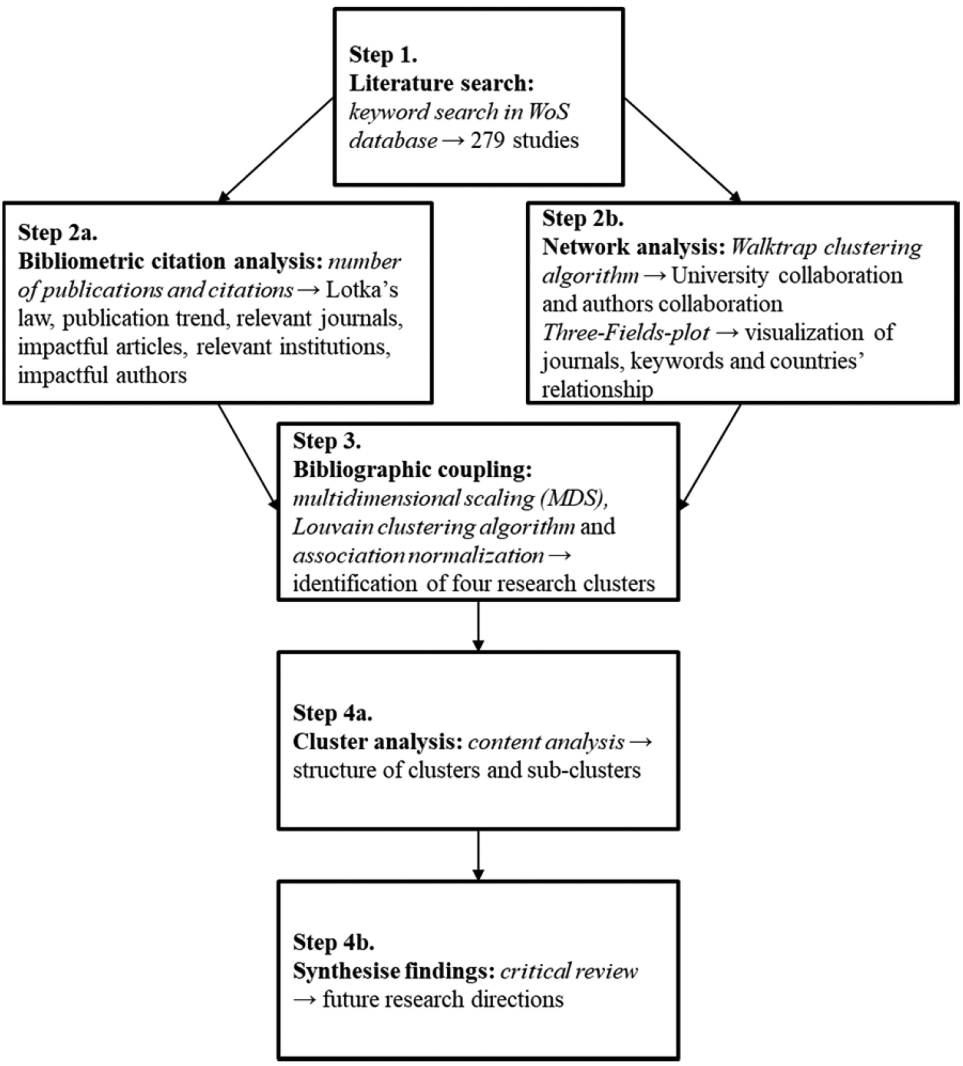

Figure 1. Methodological flowchart for bibliometric review.

Table 1. Keyword search in WoS.

Step	Keyword search	#Articles
1	'big data' AND 'maritime'	76
2	'artificial intelligence' AND 'maritime'	56
3	'business intelligence' AND 'maritime'	7
	'data analytics' AND 'maritime'	17
4	('artificial intelligence' OR 'big data' OR 'business intelligence' OR 'data analytics') AND 'maritime'	140
5	(('artificial intelligence' OR 'big data' OR 'business intelligence' OR 'data analytics') AND ('maritime' OR 'shipping'))	197
6	(('artificial intelligence' OR 'big data' OR 'business intelligence' OR 'data analytics') AND ('maritime' OR 'shipping' or 'port'))	308
After initial analysis of 308 articles found that 'machine learning' is one of the top keywords		
7	((('artificial intelligence' OR 'big data' OR 'business intelligence' OR 'data analytics' OR 'machine learning') AND ('maritime' OR 'shipping' or 'port')))	602
8	((('artificial intelligence' OR 'big data' OR 'business intelligence' OR 'data analytics' OR 'machine learning') AND ('maritime' OR 'shipping' or 'port'))) *NOT TOPIC: ('port number*')*	550
9	(('artificial intelligence' OR 'big data' OR 'business intelligence' OR 'data analytics' OR 'machine learning') AND ('maritime' OR 'shipping' or 'port')) NOT TOPIC: ('port number*') *Refined by: LANGUAGES: (ENGLISH)*	541 (A:234, C:309, R:5, EA:3, EM: 1)
10	(('artificial intelligence' OR 'big data' OR 'business intelligence' OR 'data analytics' OR 'machine learning') AND ('maritime' OR 'shipping' or 'port')) NOT TOPIC: ('port number*') Refined by: LANGUAGES: (ENGLISH) *Manually screened by relevance*	279 (A:132, C:141, R:3, EA:2, EM:1)

*A. Article, C. Conference proceedings, R. Review, EA. Early access, EM. Editorial material.

'machine learning') covers articles about big data and AI, while the second part ('maritime' OR 'shipping' or 'port') captures the big data and AI studies in the maritime industry. After refining the search excluding the keyword 'port number' and including studies written only in English, 541 articles were left for consideration. We excluded articles with the keyword 'port number' as they mainly represent the electronics and telecommunications discipline, and not seaport research. Finally, we screened the 541 studies for relevance to the big data and AI in the maritime industry. The relevance for the extracted studies is rooted in the definition of the big data and AI in the maritime industry used in this study and excludes 262 irrelevant articles from the following categories: (1) computer science studies focusing on the information systems and artificial intelligence theory and methods without application in maritime domain; (2) general shipping studies and not maritime shipping; (3) studies where port, shipping or logistics are land-based; and (4) cases of the homonymy problem of keywords (for example, a 'port' is a communication endpoint in computer networking, and a 'vessel' is a part of the circulatory system in biology). The remaining 279 studies were selected for bibliometric analysis. Of these, 132 are journal articles, 141 are conference proceedings, three are review articles, two are early access and one is editorial material. Most review studies only analyse journal articles, but we included other document types given that the topic of interest here is new and trending. Furthermore, the extracted bibliography data contains author, title, abstract, source, cited references, times cited, documents type, keywords and conference information for each of the 279 articles.

2.2. Bibliometric analysis method

Bibliometric analysis is a literature review methodology that refers to statistical and quantitative analysis of published studies (Broadus 1987). Bibliometrics is 'more objective and reliable' than other literature review techniques (Aria and Cuccurullo 2017, 959). When conducted and reported properly, bibliometrics reviews are 'systematic, transparent and reproducible' (Aria and Cuccurullo 2017, 959). Initially, scholars used bibliometrics for analysis of published studies based on the number of publication and citations counts. With recent developments in bibliometric methods and software, and tools such as Vosviewer (Van Eck and Waltman 2009) and Bibliometrix (Aria and Cuccurullo 2017), we can map a scientific field or topic through co-citation analysis, bibliographic coupling, keyword co-occurrence and other techniques (Zupic and Čater 2015).

Having extracted the complete bibliography data of 279 studies from the WoS database, we used bibliometric citation analysis measures such as the number of publications and citations to (1) test Lotka's Law, (2) present an overview of publication trend, and (3) find the most relevant journals, impactful articles, relevant institutions and impactful authors.

Lotka's Law is a bibliometric measure of authorship concentration that describes the frequency of scientific publication by authors. The central assumption is that a few authors are highly productive in any given field, while a relatively substantial number of authors produce only a single article. As per Lotka's law, the number of authors publishing x number of articles is about $1/x^b$ of those publishing only one article. Thus, a higher b value indicates a higher degree of authorship concentration and a low value indicates the absence of a dedicated group of authors in a particular scientific discipline. The general formula of Lotka's Law is:

$$f(y) = \frac{C}{x^b} \quad (1)$$

Here, $f(y)$ is the number of occurrences of studies written by each author of a population, C and are research field specific constants.

We present university collaboration and authors' collaboration using a *Walktrap clustering algorithm* proposed by Pons and Latapy (2005). A clear advantage of this network analysis algorithm is in terms of the quality of the computed partition and the running time for large networks (Pons and Latapy 2005). Furthermore, we use a *three-fields-plot* based on a Sankey diagram that visualises how journals, keywords and countries are interrelated.

The *bibliographic coupling* approach, introduced by Kessler (1958), is used for identifying the cluster of underlying research and mapping contemporary research developments. When two or more published studies share at least one common citation, they are said to be bibliographically coupled (Kessler 1963). Therefore, bibliographically coupled studies are likely to share the same underlying research theme. We can express a bibliographic coupling network of published studies using the following generic formula (Aria and Cuccurullo 2017):

$$B_{coup} = A \times A' \quad (2)$$

Here, A is a $document \times citedreference$ matrix, B_{coup} consist of the matrix b_{ij} which represents the number of common references between article .

For bibliographic coupling analysis, we used the open-source R programming software-based Bibliometrix package (Aria and Cuccurullo 2017). Although Bibliometrix offers a few alternatives, we used the *multidimensional scaling (MDS) method* (Cox and Cox 2000) for information visualisation and the *Louvain clustering algorithm* (Blondel et al. 2008) for detecting communities in networks. Blondel et al. (2008) proved that the Louvain clustering algorithm allows networks of unprecedented size accessible to computational analysis and outperforms all the other methods in terms of computation time, memory-efficient manner and quality of the communities detected. In cluster analysis, normalisation approaches compensate for different occurrence levels among items, and MDS also benefits from normalisation (Zupic and Tomaž 2015). Thus, we normalised bibliographic coupling data using the 'association' measure (Van Eck and Waltman 2009) in Bibliometrix.

The bibliographic coupling analysis identified four underlying research clusters within the big data and AI research in the maritime industry. For research clusters, we considered the 50 most bibliographically coupled studies. For robustness, we considered different numbers of studies (for example, 100, 150, 200 and 279), and obtained the same four clusters but with greater detail. Hence, for better visibility, we proceeded with 50 studies.

3. Bibliometric analysis and findings

This study reviewed big data and AI research in the maritime industry published during 1995–2019. We reviewed a sample of 279 relevant studies published in 214 publication outlets over the last 25 years, written by a total of 842 authors, with an average of 4.78 citations per document. The majority of authors are part of multi-authored studies (824 authors, or 97.9%), while only 2.1% are single-authored studies (18 authors).

3.1. Standalone research domain

In the context of big data and AI in maritime, scientific productivity shows that there are 736 authors in the field that have produced only one article. On the other hand, one author has contributed to a maximum of seven published studies. Pao (1986) found that the value of the *b* in *Lotka's law* typically ranges from 1.78 to 3.78 for most disciplines. The estimated *b* value of 3.446 in this study falls between the range observed by Pao (1986). Thus, big data and AI in maritime is a standalone research domain, albeit with a high degree of authorship concentration.

3.2. Yearly publication and time trend

We present the yearly publication trend in Figure 2. The number of publications on big data and AI in maritime has been growing at a rate of 20.97 articles per year. The output of the first 20 years (1995–2014) was 59 publications, significantly lower than the 220 in the last five years (2015–2019). This indicates a growing interest in big data and AI research in the maritime industry.

Figure 2 also shows the impacts (that is, citations) of studies published in a given year. Studies published in recent years have not received many citations as it takes time for research to make an impact. The total global citation score (TGCS) is a bibliometric measure that indicates the number of citations of studies by all other studies indexed in the WoS database. Another citation metric, the total local citation score (TLCS), refers to the number of citations of studies by the sample of 279 studies in big data and AI in maritime. Both TGCS and TLCS have shown a growing trend since 2008–2009.

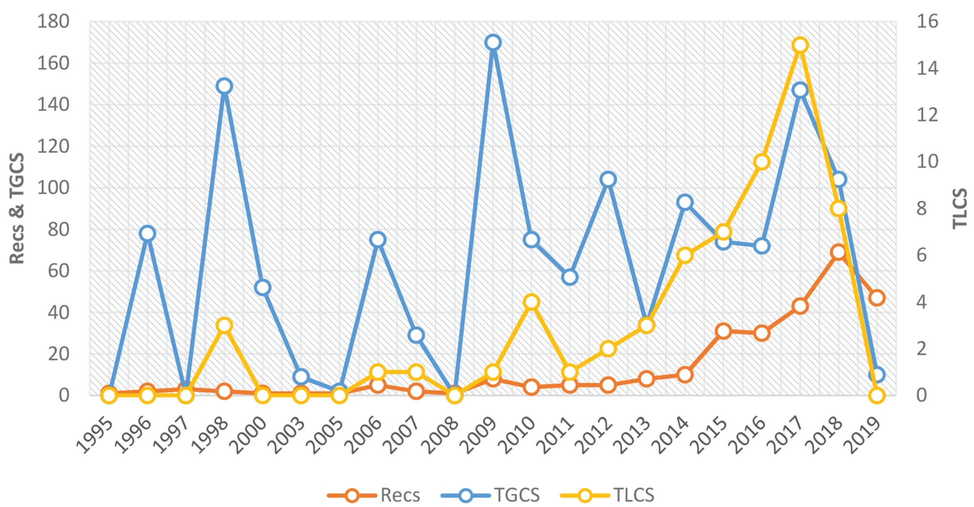

Figure 2. Overview of bibliography data *(Recs. Number of publications, TGCS. Total global citation score, TLCS. Total local citation score).*

3.3. Top journals

The sample of 279 studies in big data and maritime were published in 214 academic outlets. Almost one-fifth of these studies (see Table 2) were published in 14 outlets, including two dedicated conference proceedings that focus explicitly on big data. Seven (2.5%) papers were published in *Ocean Engineering*, and six (2.2%) in the *Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems*. Notably, the latter is the proceedings of only one year, namely 2018. The other outlets that published four or more studies on big data and AI in the maritime industry are *Sensors* (5), *2017 International Conference On Engineering, Technology And Innovation* (4), *Expert Systems With Applications* (4), *IEEE Access* (4), and *Journal of Navigation* (4); these constitute the essential publication outlets for research on big data and AI in maritime.

Journal of Navigation is the frontier journal with the highest level of impact on the body of knowledge on the big data and AI in maritime (with TGCS of 118 and TLCS of 5). Interestingly, it published only four relevant articles, while seven articles appear in *Ocean Engineering*.

3.4. Most impactful articles

Bibliometric measures such as TGCS and TLCS indicate seminal or recent breakthrough studies. Table 3 reports the most cited papers ranked according to TLCS and TGCS. By comparing TLCS

Table 2. Most relevant publication outlets.

Rank	Publication outlet	Recs	TLCS	TLCS/ t	TGCS	TGCS/ t	TLCR
1	Ocean Engineering	7	2	0.29	31	7.58	3
2	Debs’18: Proceedings of the 12Th ACM International Conference on Distributed and Event-Based Systems	6	0	0	0	0	0
3	Sensors	5	0	0	18	5.25	9
4	2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)	4	1	0.33	6	2	0
5	Expert Systems with Applications	4	1	0.5	59	11.17	0
6	IEEE Access	4	0	0	4	3	1
7	Journal of Navigation	4	5	0.96	118	13.05	3
8	Maritime Policy & Management	3	4	1.33	24	8	2
9	Proceedings of The Institution of Mechanical Engineers Part M-Journal of Engineering For The Maritime Environment	3	1	0.25	15	4.5	1
10	Proceedings 2015 IEEE International Conference on Big Data	3	4	0.8	14	2.8	1

* Ranking by Recs. We present journals with a minimum of three relevant publications. In cases of equal Recs, we present journals with higher TGCS. (Recs. Number of publications, TGCS. Total global citation score, TLCS. Total local citation score, TLCR. Total local citated references).

Table 3. Most influential articles.

Rank	Article*	TLCS	TLCS/ t	Article**	TGCS	TGCS/ t
1	Yang et al. (2018a)	4	2	Yang et al. (2018a)	20	10
2	Millefiori et al. (2016)	4	1	Sala et al. (2018)	19	9.50
3	Adland, Jia, and Strandenes (2017)	3	1	Pietrzykowski and Uriasz (2009)	93	8.45
4	Heilig and Voß (2017)	3	1	Arridha (2017)	20	6.67
5	Mascaro, Nicholso, and Korb (2014)	5	0.83	Abdel-Aal, Elhadidy, and Shaahid (2009)	73	6.64
6	Kim, Kim, and Park (2017)	2	0.67	Mascaro, Nicholso, and Korb (2014)	34	5.67
7	Perera and Mo (2016)	2	0.50	Tsou and Hsueh (2010)	56	5.60
8	Lee et al. (2018b)	1	0.50	Yang, Ye (2018)	10	5
9	Fernández et al. (2018)	1	0.50	Gambardella, Rizzoli, and Zaffalon (1998)	97	4.41
10	Lee et al. (2018a) and Zerbino et al. (2018b)	1	0.50	Lalla-Ruiz, MeliáN-Batista, and Moreno-Vega (2012)	32	4

*Ranking by TLCS/t. **Ranking by TGCS/t. In cases of equal TLCS/t or TGCS/t, we present studies with higher TLCS or TGCS. (TGCS. Total global citation score, TLCS. Total local citation score).

and TGCS, we can identify seminal works that received most citations within the sample of 279 studies and beyond it, respectively. Reportedly, Yang et al. (2018a), with a TGCS/t of ten and TLCS/t of two, is the most influential publication within the big data and AI in the maritime domain. Based on TLCS/t, other influential studies are Millefiori et al. (2016), Adland et al. (2017) and Heilig and Voß (2017) with a TLCS/t of one.

3.5. Most relevant institutions

Table 4 shows the most relevant institutions that publish research on big data and AI in the maritime industry. These 15 institutions produce almost one-third of the total output. The top three institutions are in China and account for 32 articles (11.5%). They are the *Wuhan University of Technology* with 15 articles (5.4%), *Shanghai Maritime University* with nine (3.2%) and *Dalian Maritime University* with eight (2.9%). Two of these institutions are specialised maritime universities. The other top relevant institutions comprise universities and research centres from Greece (*University Aegean*), Singapore (*Nanyang Technological University*), Norway (*Norwegian School of Economics*), Italy (*The NATO STO-CMRE* and the *University of Genoa*), the United Kingdom (*University of Plymouth*) and Spain (*Polytechnic University of Valencia*).

The university collaboration network is shown in Figure 3. A bibliometric network consists of nodes and edges. Here, the nodes are 20 universities and the edges indicate relations between pairs of nodes by research collaborations. The presence of several dispersed university clusters indicates that big data and AI in the maritime is an emerging research domain. The blue cluster (six universities) is the largest and most influential. This cluster is also the most diverse, including universities from China *(Wuhan University of Technology)*, the UK *(University of Cambridge)*, Netherlands *(Delft University of Technology)*, the United States *(Northwestern State University)* and South Africa *(University of Pretoria)*. The other two prominent clusters are the green cluster (four universities), followed by red cluster (five universities). Two clusters (the orange and purple clusters) represent research collaboration between only two universities.

3.6. Most impactful authors

Authors shape a research field with their publications. Table 5 presents the most impactful authors in the big data and AI in research the maritime domain. Dimitrios Zissis, Luca Cazzanti and Leonardo M. Millefiori are the top three authors; collectively, they contributed to 17 studies. George Arcieri, Alexander Artikis, Federico Barber, Konstantinos Chatzikokolakis, Christophe Claramunt, Brage Mo and Lokukaluge P. Perera contributed to four studies each. In terms of TLCS, Luca Cazzanti and Leonardo M. Millefiori are the forerunners, while Federico Barber leads in terms of TGCS. Luca Cazzanti is the most influential author, with a TLCS of seven and a TGCS of 18. He also has the highest number of yearly TLCS (1.7) and TGCS (4.05).

Table 4. Most relevant institutions.

Rank	Institution*	Recs	Percent	TLCS	TGCS
1	Wuhan University of Technology	15	5.4	2	25
2	Shanghai Maritime University	9	3.2	1	12
3	Dalian Maritime University	8	2.9	4	38
4	University Aegean	6	2.2	6	15
5	Nanyang Technology University	5	1.8	0	20
6	Norwegian School of Economics	5	1.8	4	26
7	University Genoa	4	1.4	2	66
8	University Politecnica Valencia	4	1.4	3	37
9	University Plymouth	4	1.4	2	26
10	NATO STO Centre Maritime Research & Experimentation CMRE	4	1.4	6	15

* Ranking by Recs. We present institutions with a minimum of four relevant publications. In cases of equal Recs, we present institutions with higher TGCS. (Recs. Number of publications, TGCS. Total global citation score, TLCS. Total local citation score).

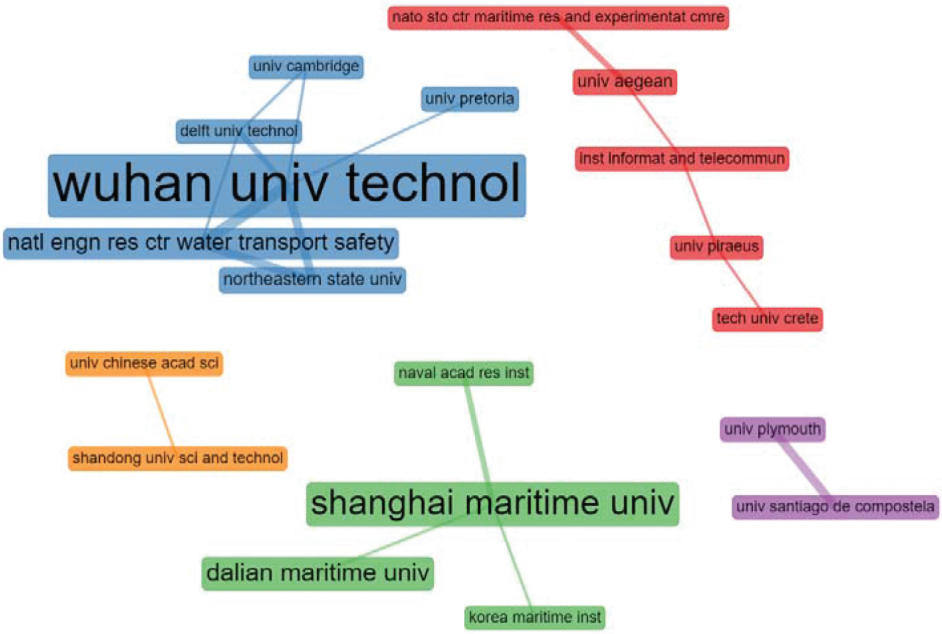

Figure 3. University collaboration network *(20 nodes, 1 minimum edge, walktrap clustering algorithm, association normalization).*

Table 5. Most impactful authors.

Rank	Author	Recs	Percent	TLCS	TLCS/t	TGCS	TGCS/t	TLCR
1	Zissis D	7	2.5	6	1.58	16	5	4
2	Cazzanti L	5	1.8	7	1.7	18	4.05	0
3	Millefiori LM	5	1.8	7	1.7	15	3.58	2
4	Arcieri G	4	1.4	6	1.45	12	2.83	2
5	Artikis A	4	1.4	1	0.2	6	1.73	2
6	Barber F	4	1.4	3	0.36	37	4.42	2
7	Chatzikokolakis K	4	1.4	0	0	6	2.17	4
8	Claramunt C	4	1.4	1	0.17	8	1.5	0
9	Mo B	4	1.4	3	0.83	13	3.75	2
10	Perera LP	4	1.4	3	0.83	13	3.75	2

*Ranking by Recs. We present authors with a minimum of four relevant publications. (Recs. Number of publications, TGCS. Total global citation score, TLCS. Total local citation score, TLCR. Total local cited reference).

We present the authors' collaborative network with 20 nodes in Figure 4. Each node represents an author and edges indicate co-authorship relation among them. With five authors, the purple cluster shows the most influential and significant author collaboration network. The pink, brown and orange clusters consist of three authors' network, separately. Other clusters—the green, blue and red clusters—consist of only two authors, indicating weak collaboration compared to other clusters. While only two of the prominent clusters (purple and brown) are interconnected, there is enough room to improve overall author collaboration within the big data and AI in maritime research.

3.7. *Three-fields plot*

The interconnections among journals, research topics and countries can provide useful insights. Hence, we present an innovative three-fields plot in Figure 5, which shows the interactions among the most relevant publication outlets (left), author keywords (middle) and countries (right) within

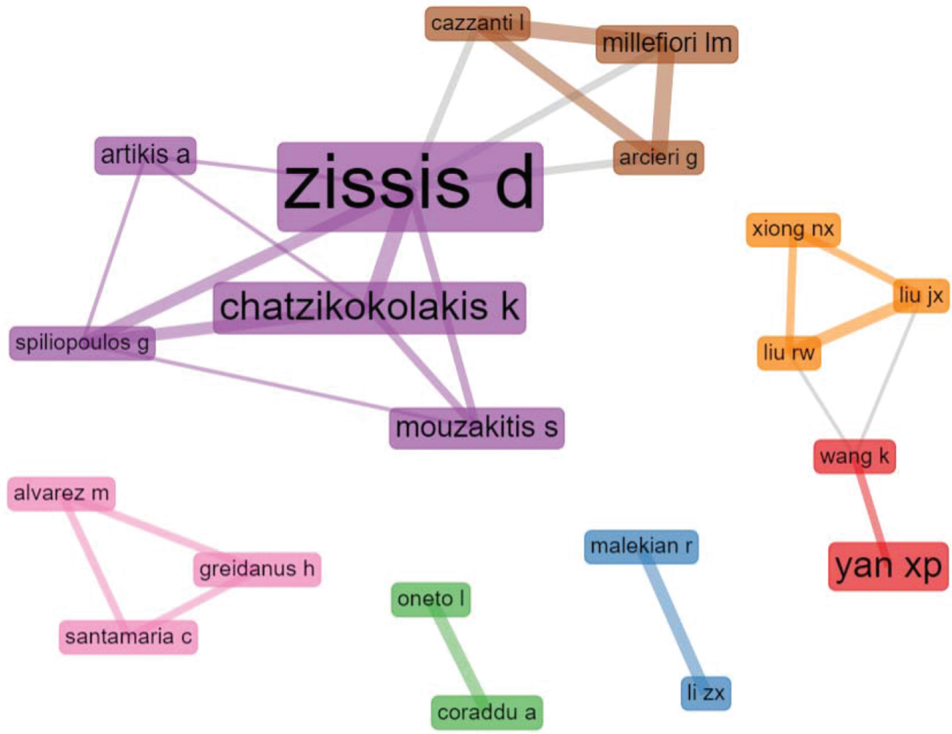

Figure 4. Author collaboration network *(20 nodes, 1 minimum edge, walktrap clustering algorithm, association normalization).*

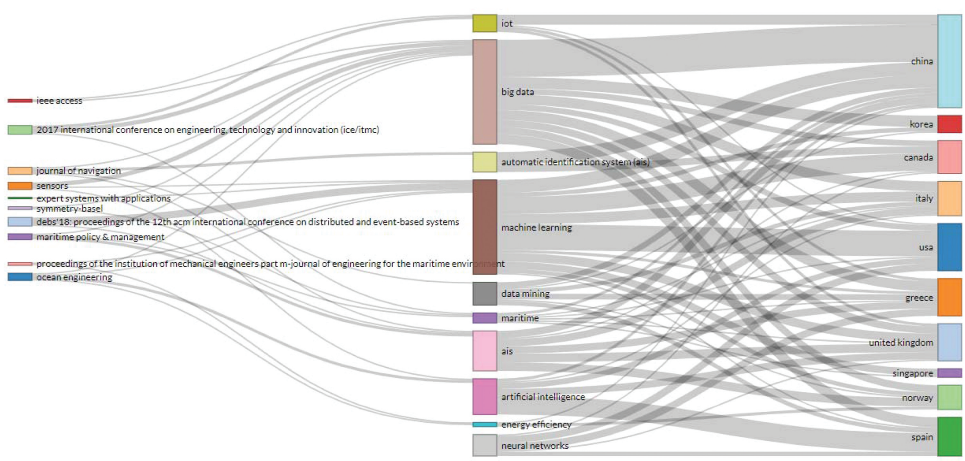

Figure 5. Three-fields-plot.

the big data and AI in maritime research. We find that studies on the Internet of Things (IoT) are mostly published in *IEEE Access*, the majority of which are authored by Chinese scholars. Similarly, the journal *Sensors* published the majority of the big data studies, again mainly authored by Chinese scholars. In general, China excels in big data, the USA in machine learning and Spain in AI. On an aggregate level, European countries have a great interest in AIS, data mining and neural networks. Among the journals, *Ocean Engineering* publishes studies in a comparatively broad range of topics including energy efficiency, artificial intelligence, machine learning and big data.

4. Research clusters and future research directions

Every research domain has several underlying research themes. In big data and AI in maritime research, using the bibliographic coupling analysis, we have revealed four underlying clusters: (1) digital transformation, (2) applications of big data from AIS, (2) energy efficiency and (4) predictive analytics (see Figure 6). The first three clusters are connected and built on each other. Meanwhile, predictive analytics can be considered as a standalone cluster, too. Appendix A presents the prominent studies in each cluster.

We critically reviewed the contents of the bibliographically coupled studies within each cluster and derived sub-clusters that share the same underlying research topic and extract relevant future research directions. For content analysis and sub-clustering, we used a concept matrix (Salipante, Notz, and Bigelow 1982) in Excel that includes columns for the article title, year of publication, author(s), journal, keywords, data, methodology, article cluster, sub-cluster, key finding(s) and abstract. The sub-clusters emerged following an iterative content analysis of all the articles in a cluster. Figure 7 depicts the clusters and their respective sub-clusters.

4.1. Digital transformation in the maritime industry

Broadly, we categorise the first cluster as digital transformation in the maritime industry. After content analysis of the respective studies for this cluster, we found three interconnected sub-clusters and briefly discussed the studies within each of the sub-clusters.

4.1.1. Digitalisation in maritime transport

This sub-cluster focuses on digitalisation in shipping. ‘Digitalization is the use of digital technologies to change a business model and provide new revenue and value-producing opportunities, that is, the process of entering a digital business’ (Gartner Inc. n.d.). This sub-cluster is comprised of three articles—Fruth and Teuteberg (2017), Heilig and Voß (2017) and Sanchez-Gonzalez et al. (2019)—all of which are literature reviews.

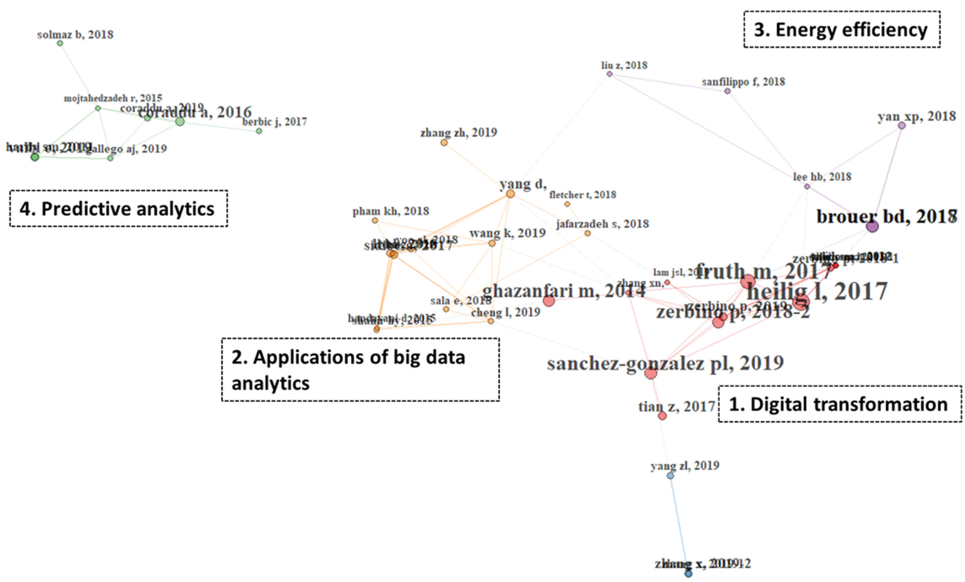

Figure 6. Underlying research clusters. **Bibliometrix R code**: NetMatrix <- biblioNetwork(M, analysis = ‘coupling’, network = ‘references’, sep = ‘;’);Net = networkPlot(NetMatrix, type = ‘mds’,normalize = ‘association’, cluster = ‘louvain’, n = 50, size = 4, size.cex = T, labelsize = 3, label.cex = T, remove.isolates = T)

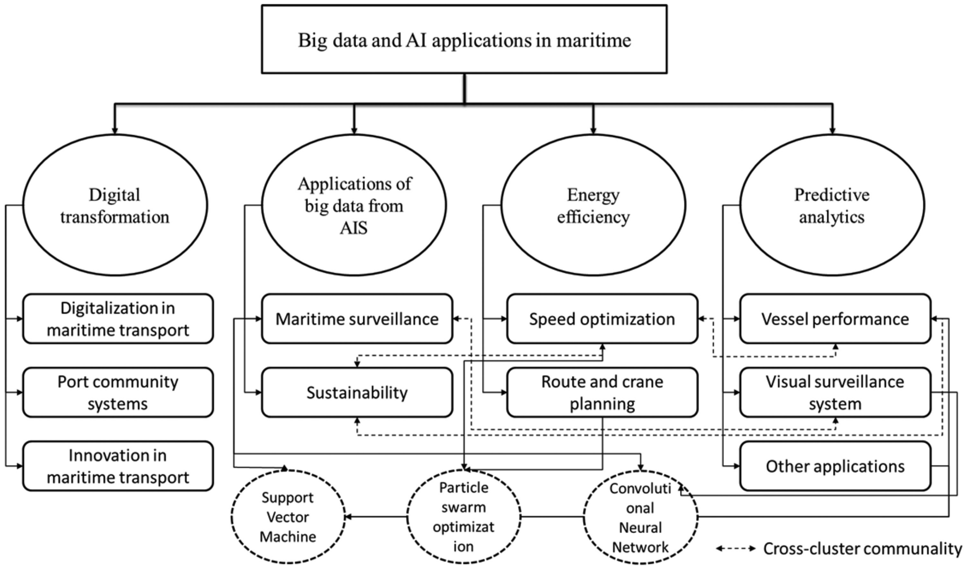

Figure 7. Clusters and their respective sub-clusters. (Support Vector Machine and Convolutional Neural Network are two widely used machine learning methods in the maritime surveillance, vessel performance and visual surveillance system sub-clusters. Particle Swarm Optimization method is common among studies in the sub-clusters of energy efficiency).

Fruth and Teuteberg (2017) presented a systematic literature review (SLR) of the digitalisation in maritime logistics considering 124 studies relevant for academia and practice. Sanchez-Gonzalez et al. (2019) also used the SLR approach to review digitalisation in the maritime transport literature. Compared to Fruth and Teuteberg (2017), the keyword search approach of Sanchez-Gonzalez et al. (2019) is more comprehensive and robust. Overall, the majority of the identified research topics are common in both studies; for example, autonomous ships, big data, artificial intelligence, cybersecurity, IoT and virtual reality.

Meanwhile, Heilig and Voß (2017) presented a state-of-the-art review of the information systems in port operations, although they did not follow a systematic approach to select the studies for review. They classified port information systems into 10 groups: (1) national single window, (2) port community systems, (3) vessel traffic services, (4) terminal operating systems, (5) gate appointment systems, (6) automated gate systems, (7) automated yard systems, (8) port road and traffic control information systems, (9) intelligent transportation systems, and (10) port hinterland intermodal information systems. All of these systems rely to some extent on technologies such as RFID, GPS, AIS and EDI and produce big data that can be analysed using AI techniques for value creation in the maritime business.

Future research: All of the studies in this sub-cluster mention some challenges due to the digitalisation wave in shipping. The most important issue is the risk of cyber-attacks (Fruth and Teuteberg 2017; Sanchez-Gonzalez et al. 2019), followed by the integration of new systems and technologies (Heilig and Voß 2017). Also, Fruth and Teuteberg (2017) argued for more theoretical research that can map the behaviour of different actors within the maritime supply chain. Furthermore, future research should explore applications of cloud computing and 3D printing in maritime transport (Sanchez-Gonzalez et al. 2019).

4.1.2. Port community systems

Among the port information systems classifications of Heilig and Voß (2017), the port community system (PCS) receives more considerable attention in the big data and AI research in the maritime domain. 'A PCS is an inter-organizational system (IOS) that electronically integrates heterogeneous compositions of public and private actors, technologies, systems, processes, and standards within a port community' (Heilig and Voß 2017, 190). Before making any investment, it is essential to evaluate a PCS system. Thus, incorporating 34 business intelligence criteria, Ghazanfari, Rouhani, and Jafari (2014) proposed a fuzzy TOPSIS model for PCS evaluation that can help maritime organisations make better decisions when assessing, selecting or purchasing a PCS.

Later, a group of authors studied PCS from the process mining (Zerbino et al. 2018b; 2019) and knowledge management perspectives (Zerbino et al. 2018a). Similar to Ghazanfari, Rouhani, and Jafari (2014), Zerbino et al. (2018b) applied process mining to evaluate PCS. They argued that existing methodologies are flawed in terms of detecting deviations and frauds in information systems/PCS. Process mining, an emerging type of business analytics, is a set of business-process-related diagnostic and validation tools that can overcome these flaws (Zerbino et al. 2018b; 2019). In the same vein, depending on the knowledge management concept, Zerbino et al. (2018a) examined the barriers of PCS implementation and found that strategic, organisational and technological barriers need more attention.

Future research: Meanwhile, Ghazanfari, Rouhani, and Jafari (2014) recommended the comparative application of other MCDM methods in PCS evaluation and selection of a PCS with appropriate business intelligence level. Zerbino et al. (2018b) suggested further development of the process mining tool; for example, by expanding the methodology into an online PCS evaluation tool. Investigation of the effects of process mining on port performance indicators and related costs could be another future research agenda (Zerbino et al. 2019). Future research should also examine strategies to overcome the barriers of PCS implementation (Zerbino et al. 2018a).

4.1.3. Innovation in maritime transport

Innovation in the maritime industry has traditionally been a slow and incremental process. Two of the studies in this sub-cluster (Tian et al. 2017; Lam and Zhang 2019) introduced customer-centric innovative solutions, while another study (Zhang and Lam 2019) examined the barriers to such solutions. Tian et al. (2017) reviewed the innovative concept of the Internet of Vessels (IoV), which integrated all key technologies, such as sensing, automation, telecommunications, computers, information and smart control, into a platform. Similar to the IoT, the IoV is 'a network of smart interconnected vessels and shore facilities with a series of digital entities' (Tian et al. 2017, 1). Later, Lam and Zhang (2019) introduced seven design requirements and examined how these influence customer values in the context of liner companies. They found that the use of eco ship and container technology, big data solution for ship information management and automation of system are the three most effective solutions. Meanwhile, Zhang and Lam (2019) found that a lack of understanding in improving business using analytics, a lack of executive sponsorships and a lack of skills are the three main barriers to big data analytics adoption among maritime organisations.

Future research: In the context of IoV, future research should conduct a cost-benefit analysis (Tian et al. 2017). Also, cross-regional information integration and management strategies for IoV would be helpful (ibid). Future research should also examine the dynamics of innovative technology adaptation barriers using causal models such as structural equation modelling (Zhang and Lam 2019).

4.2. Applications of big data from AIS

All studies in this cluster use AIS data to investigate a wide range of research topics. Among the studies that comprise this cluster, three are literature review studies. The first one focuses on anomaly detection techniques (Sidibé and Shu 2017), the second on the applications of AIS data (Yang et al. 2019) and the last on the visualisation of maritime traffic data (Wang et al. 2019). While

Yang et al. (2019) provided a detailed review of topics covered in this research cluster, Sidibé and Shu (2017) focused on anomaly detection techniques covering studies published during 2011–2016, and Wang et al. (2019) argued that appropriate visualisation of maritime traffic can the improve safety and reliability of maritime transport. Other studies cover topics ranging from maritime security to search and rescue operations.

4.2.1. Maritime surveillance

Maritime surveillance provides support to widespread maritime policies such as maritime security (Yeo et al. 2019), illegal bunkering (Yeo et al. 2019), tracking of marine oil transportation (Cheng et al. 2019) and search and rescue (Pham, Boy, and Luengo-Oroz 2018).

Some studies in this sub-cluster focused on anomaly detection. There are typically three approaches to anomaly detection: data mining, statistical and machine learning (Sidibé and Shu 2017). Handayani and Sediono (2015) used machine learning algorithms called Bayesian networks (BN) for anomaly detection in vessel tracking. Shahir et al. (2015) presented a novel approach to anomaly detection by representing patterns (a family of multi-vessel scenarios with common kinematic characteristics) using the left-to-right Hidden Markov Model (HMM) and classifying them using the Support Vector Machine (SVM) model.

The Singapore Port and IBM recently collaborated on a project called SAFER, which uses a machine learning approach to predict vessel arrival time, potential traffic hot spots, unusual behaviour of vessels and illegal bunkering (Yeo et al. 2019). Cheng et al. (2019) proposed a framework combining data mining and statistical modelling to predict tanker routes along the 21st-century Maritime Silk Road. Meanwhile, Pham, Boy, and Luengo-Oroz (2018) focused on social good and proposed an automated identification of the rescue model using machine learning algorithms.

Future research: Although anomaly detection literature is rather mature, future research should consider real-time anomaly detection of vessels and the use of advanced big data technologies such as the Hadoop MapReduce framework (Sidibé and Shu 2017). Cheng et al. (2019) argued that future research could use their proposed shipping route extraction method for collision avoidance and environmental impacts of maritime oil transportation. Furthermore, applications of deep learning algorithms such as Convolutional Neural Networks (CNNs) for pattern recognition with different location and scale, and Recurrent Neural Networks (RNNs) for the better sequential structure of data can improve the effectiveness of the automated identification of a rescue model (Pham, Boy, and Luengo-Oroz 2018).

4.2.2. Environmental and economic sustainability

Sustainability refers to the conservation of resources for future generations while fulfilling present needs. Studies in this sub-cluster evaluate ship performance from sustainability perspectives, particularly environmental and economic sustainability. Fletcher et al. (2018) used several machine learning algorithms to estimate shipping emission. They found that Supervised Mixture Probabilistic Latent Factor Regression (SMPPLS) and Gaussian Process Regression (GPR) out-performed other algorithms. Using big data from AIS, Jafarzadeh and Schjølberg (2018) showed that offshore and passenger ships benefit most from hybrid (that is, diesel-electric) and electric propulsion in terms of emissions reduction, followed by container and Ro-Ro ships. Meanwhile, we found an interesting application of big data from AIS in the context of high-seas fishing. Sala et al. (2018) analysed AIS data from 3620 fishing vessels using machine learning, finding that the government subsidises more than half of high-seas fishing, without which the majority of the high-seas fishing vessels would be out of operation.

Future research: From an environmental sustainability perspective, using machine learning, future research can attempt to predict particulate matter emissions from shipping (Fletcher et al.

2018). In the same vein, based on their original finding, Jafarzadeh and Ingrid (2018) suggested that future studies should examine the impact of hybrid and electric propulsion on the energy efficiency of offshore and passenger ships.

4.3. Energy efficiency

Scholars have used big data and AI to achieve better energy efficiency in maritime transport from different perspectives. While the majority of the studies focus on vessel speed optimisation (Lee et al. 2018a; Yan et al. 2018), others focus on route planning (Liu et al. 2018) and vessel crane control (Sanfilippo et al. 2017). Brouer, Karsten, and Pisinger (2017) considered a broader scenario and optimised a liner shipping network focusing on container routing and speed optimisation.

4.3.1. Speed optimisation

For better energy efficiency of vessels, liner companies usually consider slow steaming as the best practice (Cariou 2011), although it can contradict service level agreements (SLAs). Also, the majority of the existing studies do not consider weather conditions while optimising vessel speed. Thus, using big data, Lee et al. (2018a) minimised fuel consumption while maximising SLA using the particle swarm optimisation method. Yan et al. (2018) found the optimal speed of inland ships using the distributed parallel k-means clustering algorithm. Their proposed approach is effective in reducing vessel energy consumption and CO_2 emissions.

Future research: In future research, port-related uncertainties (congestion, equipment breakdown, etc.) should be included in the optimisation model for vessel speed optimisation (Lee et al. 2018a). Also, future studies should attempt theoretical development for ship energy efficiency optimisation (Yan et al. 2018).

4.3.2. Route and crane planning

Apart from speed optimisation, energy efficiency in maritime transport can be achieved from intelligent shipping route planning and advanced controlling of vessel cranes. Sanfilippo et al. (2017) found the particle swarm optimisation method slightly better than genetic algorithms for vessel crane controlling. Liu et al. (2018) acknowledged the use of the same methods as Sanfilippo et al. (2017) for route planning. However, they proposed an improved hybrid model for intelligent shipping route planning that integrates the genetic algorithms and particle swarm optimisation methods. Both studies demonstrate the possibilities for energy consumption optimisation.

Future research: Sanfilippo et al. (2017) proposed some avenues for future research; for instance, a comparison of genetic algorithms and particle swarm optimisation methods with other methods such as proportional–integral–derivative (PID) and nonlinear controlling methods (ibid). Sanfilippo et al. (2017) also argued for open-source development of control methods for rapid testing of new and different methods. In the context of intelligent shipping route planning, Liu et al. (2018) proposed considering dynamic constraints (such as water velocity and wind speed in real-time) rather than static ones. Also, 3D information instead of 2D (for example, water depth ignored) should be considered in future research for shipping route planning (ibid).

4.4. Predictive analytics

Although dominated by machine learning algorithms, predictive analytics in the maritime research covers a wide range of applications, ranging from predicting vessel propulsion failure to predicting toxic blooms in coastal waters.

4.4.1. Vessel performance

Modern ships have many sensors that collect data on a wide range of factors, including temperature, flow rates and pressure. The vast majority of those data are yet to be utilised in decision making. Utilising such data, Coraddu et al. (2016) forecast potential future failures of a vessel propulsion system and found that SVM is better than regularised least square (RLS). Coraddu et al. (2019) predicted vessel hull condition using two unsupervised machine learning models and found that both the One-Class SVM (OCSVM) and Global KNN (GKNN) perform well and are identical in terms of prediction accuracy.

Future research: For future research, both Coraddu et al. (2016) and Coraddu et al. (2019) discussed applications of their approach in a real-world scenario, and both argued for the use of unsupervised machine learning methods in vessel propulsion performance prediction.

4.4.2. Visual surveillance system

Another important application is in the context of the visual surveillance system. Solmaz et al. (2018) found that deep learning-based CNN models improve the accuracy of visual recognition and verification of maritime vessels and land vehicles. Taking the CNN one step further, Gallego et al. (2019) proposed the Convolutional Long Short Term Memory Selectional AutoEncoders (CMSAE) neural architecture to predict maritime oil spills using Side-Looking Airborne Radar (SLAR) images.

Future research: In the context of visual recognition, future research should utilise metadata from the sensors and the aircraft to improve prediction performance (Gallego et al. 2019). Also, the use of auxiliary data such as flight speed, altitude, and wind speed can improve prediction accuracy by discriminating between true targets and sensor noise (ibid).

4.4.3. Other applications

In a real-world configuration, Mojtahedzadeh et al. (2015) examined a robotic system that can safely remove goods from containers that arrive in random order. When information about the goods is incomplete, a machine learning-based probabilistic model performs well (ibid). In predicting real-time wave heights, Berbić et al. (2017) found that SVM is more accurate than the artificial neural network (ANN). Using a random forest model, Valbi et al. (2019) predicted the occurrence of PSP toxic dinoflagellate Alexandrium minutum in the coastal waters of the northwest Adriatic Sea with more than 80% accuracy.

Future research: Overall, future research should consider a combination of different machine learning models such as SVM and ANN (Berbić et al. 2017). In the context of developing robotic systems, future studies should consider the influence of uncertainty in the attributes of objects/goods on their detection performance (Mojtahedzadeh et al. 2015).

5. Conclusions

This study provides a comprehensive review of the big data and AI studies in the maritime domain, combining bibliometric analysis with systematic content analysis. We collected bibliography data of 279 studies from the WoS database, authored by 842 scholars published in 214 academic outlets. In the following sub-section, we summarise the key findings to the four objectives stated in the introduction section.

5.1. 'Big data and AI in the maritime industry' is a standalone and growing research domain

This is based on the application of Lotka's law, which uses authorship concentration as a metric to identify research domains. An estimation of Lotka's law ($1.78 < b = 3.45 < 3.78$) indicates the existence of big data and AI in maritime as a standalone research domain with a high degree of authorship concentration. Furthermore, the exponential growth of publications on the topic in recent years (see Figure 2) confirms growing interest.

5.2. Key journals, articles, universities and authors

The most central scientific journals are *Ocean Engineering, Sensors*, and *Expert Systems with Applications*. The most influential articles based on yearly TLCS are Yang et al. (2018a), Millefiori et al. (2016) and Adland, Jia, and Strandenes (2017). In terms of the number of publications, Wuhan University of Technology, Shanghai Maritime University and Dalian Maritime University are the most active institutions. The central authors, ranked by the number of publications (in parentheses), are Zissis D (7), Cazzanti L (5) and Millefiori LM (5). There are two distinct university collaboration networks, with the biggest centred at the Wuhan University of Technology, the other around Shanghai Maritime University. In terms of author collaboration, the most active network is centred on Zissiz, D. and Chatzikokolakis, K.

5.3. Emerging research clusters within the maritime application of big data and AI

Using bibliographic coupling analysis, we reveal four underlying research clusters within big data and AI in maritime research. The clusters are (1) digital transformation, exploring the impact of digital technologies on business models and operations; (2) applications of big data from AIS, addressing how data analysis, particularly AIS data, can be applied to improve safety, security, and both environmental/commercial efficiency; (3) energy efficiency, covering topics such as speed optimisation and route/crane planning; and (4) predictive analytics, focussing on ship systems maintenance, traffic and accident scenario analysis and other decision making and forecasting challenges.

5.4. Avenues for future research

Overall, the main body of existing research seems to focus on developing new digital technology and data analytics. While technological development is needed, the mere existence of technology does not affect until it has become more widely adopted into the maritime industry. This process of diffusion brings implications not only on each stakeholder in isolation, but for the entire maritime industry as a system, including legislation, culture, commercial structure, trust and collaboration, and all issues related to the 'softer' parts of institutional change. Thus, we recommend that technology development is balanced with a more robust research emphasis on these issues. One particular area that deserves more attention from socio-cultural and commercial research perspectives is identifying and addressing the drivers and barriers on increasing transparency and trust in the maritime industry, including legal issues such as cybersecurity and data ownership within a digital network.

Acknowledgments

The authors thank the anonymous reviewers for constructive suggestions. The authors acknowledge support from the MARKOM 2020 project funded by the Norwegian Government. The language proofreading has been funded by SESAME II project.

Author credit statement

Ziaul Haque Munim: Conceptualization, data curtain, bibliometric analysis, Writing- research method, clusters & revised draft. **Mariia Dushenko**: Writing- bibliometric results, research method, revised draft. **Veronica Jaramillo Jimenez**: Writing- introduction & revised draft. **Mohammad Hassan Shakil**: Writing-bibliometric analysis & revised draft. **Marius Imset**: Writing-conclusion & revised draft.

Disclosure statement

The authors declare no conflict of interest.

ORCID

Ziaul Haque Munim http://orcid.org/0000-0002-5942-708X

References

Abdel-Aal, R. E., M. A. Elhadidy, and S. M. Shaahid. 2009. "Modeling and Forecasting the Mean Hourly Wind Speed Time Series Using GMDH-based Abductive Networks." *Renewable Energy* 34 (7): 1686–1699. doi:10.1016/j.renene.2009.01.001.

Adland, R., H. Jia, and S. P. Strandenes. 2017. "Are AIS-based Trade Volume Estimates Reliable? The Case of Crude Oil Exports." *Maritime Policy and Management* 44 (5): 657–665. doi:10.1080/03088839.2017.1309470.

Aria, M., and C. Cuccurullo. 2017. "Bibliometrix: An R-tool for Comprehensive Science Mapping Analysis." *Journal of Informetrics* 11 (4): 959–975. doi:10.1016/j.joi.2017.08.007.

Arridha, R. 2017. "Classification Extension Based on IoT-big Data Analytic for Smart Environment Monitoring and Analytic in Real-time System." *International Journal of Space-Based and Situated Computing* 7 (2): 82–93. doi:10.1504/IJSSC.2017.086821.

Berbić, J., E. Ocvirk, D. Carević, and L. Goran. 2017. "Application of Neural Networks and Support Vector Machine for Significant Wave Height Prediction." *Oceanologia* 59 (3): 331–349. doi:10.1016/j.oceano.2017.03.007.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10): P10008. doi:10.1088/1742-5468/2008/10/P10008.

Broadus, R. 1987. "Toward a Definition of "Bibliometrics"." *Scientometrics* 12 (5–6): 373–379. doi:10.1007/BF02016680.

Brouer, B. D., C. V. Karsten, and D. Pisinger. 2016. "Big Data Optimization in Maritime Logistics." In *Big Data Optimization: Recent Developments and Challenges*, edited by A. Emrouznejad, 319–344. Cham: Springer International Publishing.

Brouer, B. D., C. V. Karsten, and D. Pisinger. 2017. "Optimization in Liner Shipping." *4OR* 15 (1): 1–35. doi:10.1007/s10288-017-0342-6.

Cariou, P. 2011. "Is Slow Steaming a Sustainable Means of Reducing CO2 Emissions from Container Shipping?" *Transportation Research Part D: Transport and Environment* 16 (3): 260–264. doi:10.1016/j.trd.2010.12.005.

Cheng, L., Z. Yan, Y. Xiao, Y. Chen, F. Zhang, and L. ManChun. 2019. "Using Big Data to Track Marine Oil Transportation along the 21st-century Maritime Silk Road." *Science China Technological Sciences* 62 (4): 677–686. doi:10.1007/s11431-018-9335-1.

Coraddu, A., L. Oneto, A. Ghio, S. Savio, D. Anguita, and M. Figari. 2016. "Machine Learning Approaches for Improving Condition-based Maintenance of Naval Propulsion Plants." *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment* 230 (1): 136–153.

Coraddu, A., S. Lim, L. Oneto, K. Pazouki, R. Norman, and A. J. Murphy. 2019. "A Novelty Detection Approach to Diagnosing Hull and Propeller Fouling." *Ocean Engineering* 176: 65–73.

Cox, T. F., and M. A. A. Cox. 2000. *Multidimensional Scaling*. Boca Raton, Florida: Chapman and hall/CRC.

d'Amore, M., R. Baggio, and E. Valdani. 2015. "A Practical Approach to Big Data in Tourism: A Low Cost Raspberry Pi Cluster." In *Information and Communication Technologies in Tourism 2015*, 169–181. Cham: Springer.

Fernández, P., J. P. Suárez, A. Trujillo, C. Domínguez, and J. M. Santana. 2018. "3D-Monitoring Big Geo Data on a Seaport Infrastructure Based on FIWARE." *Journal of Geographical Systems* 20 (2): 139–157. doi:10.1007/s10109-018-0269-2.

Fletcher, T., V. Garaniya, S. Chai, R. Abbassi, H. Yu, T. C. Van, R. J. Brown, and F. Khan. 2018. "An Application of Machine Learning to Shipping Emission Inventory." *Royal Institution of Naval Architects. Transactions. Part A. International Journal of Maritime Engineering* 160 (4): A381–A96.

Franks, B. 2012. *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*. Vol. 49. Hoboken, New Jersy: John Wiley & Sons.

Fruth, M., and F. Teuteberg. 2017. "Digitization in Maritime logistics—What Is There and What Is Missing?" *Cogent Business & Management* 4 (1): 1411066. doi:10.1080/23311975.2017.1411066.

Gallego, A.-J., P. Gil, A. Pertusa, and R. B. Fisher. 2019. "Semantic Segmentation of SLAR Imagery with Convolutional LSTM Selectional AutoEncoders." *Remote Sensing* 11 (12): 1402. doi:10.3390/rs11121402.

Gambardella, L. M., A. E. Rizzoli, and M. Zaffalon. 1998. "Simulation and Planning of an Intermodal Container Terminal." *Simulation* 71 (2): 107–116. doi:10.1177/003754979807100205.

Gartner Inc. n.d. "Information Technology Glossary." Accessed on October 20, 2019. https://www.gartner.com/en/information-technology/glossary/digitalization
Ghazanfari, M., S. Rouhani, and M. Jafari. 2014. "A Fuzzy TOPSIS Model to Evaluate the Business Intelligence Competencies of Port Community Systems." *Polish Maritime Research* 21 (2): 86–96. doi:10.2478/pomr-2014-0023.
Handayani, D., and W. Sediono. 2015. "Anomaly Detection in Vessel Tracking: A Bayesian Networks (Bns) Approach." *International Journal of Maritime Engineering (RINA Transactions Part A)* 157 (A3): 145–152.
Heilig, L., E. Lalla-Ruiz, and V. Stefan. 2017. "Digital Transformation in Maritime Ports: Analysis and a Game Theoretic Framework." *NETNOMICS: Economic Research and Electronic Networking* 18 (2–3): 227–254. doi:10.1007/s11066-017-9122-x.
Heilig, L., and V. Stefan. 2017. "Information Systems in Seaports: A Categorization and Overview." *Information Technology and Management* 18 (3): 179–201. doi:10.1007/s10799-016-0269-1.
Jafarzadeh, S., and S. Ingrid. 2018. "Operational Profiles of Ships in Norwegian Waters: An Activity-based Approach to Assess the Benefits of Hybrid and Electric Propulsion." *Transportation Research Part D: Transport and Environment* 65: 500–523. doi:10.1016/j.trd.2018.09.021.
Kessler, M. M. 1958. *Concerning Some Problems of Intrascience Communication*. Massachusetts, USA: Massachusetts Institute of Technology, Lincoln Laboratory.
Kessler, M. M. 1963. "Bibliographic Coupling between Scientific Papers." *Journal of the Association for Information Science and Technology* 14 (1): 10–25.
Kim, S., H. Kim, and Y. Park. 2017. "Early Detection of Vessel Delays Using Combined Historical and Real-time Information." *Journal of the Operational Research Society* 68 (2): 182–191. doi:10.1057/s41274-016-0104-4.
Lalla-Ruiz, E., B. MeliáN-Batista, and J. Marcos Moreno-Vega. 2012. "Artificial Intelligence Hybrid Heuristic Based on Tabu Search for the Dynamic Berth Allocation Problem." *Engineering Applications of Artificial Intelligence* 25 (6): 1132–1141. doi:10.1016/j.engappai.2012.06.001.
Lam, J., S. Lee, and X. Zhang. 2019. "Innovative Solutions for Enhancing Customer Value in Liner Shipping." *Transport Policy* 82: 88–95. doi:10.1016/j.tranpol.2018.09.001.
Lee, H., N. Aydin, Y. Choi, S. Lekhavat, and Z. Irani. 2018a. "A Decision Support System for Vessel Speed Decision in Maritime Logistics Using Weather Archive Big Data." *Computers & Operations Research* 98: 330–342. doi:10.1016/j.cor.2017.06.005.
Lee, P., O. Tae-Woo, K. Kwon, and X. Ruan. 2019. "Sustainability Challenges in Maritime Transport and Logistics Industry and Its Way Ahead." *Sustainability* 11 (5): 1331. doi:10.3390/su11051331.
Lee, P. T.-W., S.-W. Lee, Z.-H. Hu, K.-S. Choi, N. Y. H. Choi, and S.-H. Shin. 2018b. "Promoting Korean International Trade in the East Sea Economic Rim in the Context of the Belt and Road Initiative." *Journal of Korea Trade* 22 (3): 212–227. doi:10.1108/JKT-03-2018-0015.
Liang, T. P., and Y. -H. Liu. 2018. "Research Landscape of Business Intelligence and Big Data analytics: A Bibliometrics Study." *Expert Systems with Applications* 111: 2–10. doi: 10.1016/j.eswa.2018.05.018.
Liu, Z., J. Liu, F. Zhou, R. W. Liu, and N. Xiong. 2018. "A Robust GA/PSO-Hybrid Algorithm in Intelligent Shipping Route Planning Systems for Maritime Traffic Networks." *Journal of Internet Technology* 19 (6): 1635–1644.
Mascaro, S., A. E. Nicholso, and K. B. Korb. 2014. "Anomaly Detection in Vessel Tracks Using Bayesian Networks." *International Journal of Approximate Reasoning* 55 (1): 84–98. doi:10.1016/j.ijar.2013.03.012.
Millefiori, L. M., D. Zissis, L. Cazzanti, and G. Arcieri. 2016. "A Distributed Approach to Estimating Sea Port Operational Regions from Lots of AIS Data." Paper presented at the 2016 IEEE International Conference on Big Data (Big Data). Washington DC, USA.
Mirović, M., M. Miličević, and O. Ines. 2018. "Big Data in the Maritime Industry." *NAŠE MORE: Znanstveno-strucni Casopis Za More I Pomorstvo* 65 (1): 56–62. doi:10.17818/NM/2018/1.8.
Mojtahedzadeh, R., A. Bouguerra, E. Schaffernicht, and A. J. Lilienthal. 2015. "Support Relation Analysis and Decision Making for Safe Robotic Manipulation Tasks." *Robotics and Autonomous Systems* 71: 99–117. doi:10.1016/j.robot.2014.12.014.
Munim, Z. H. 2019. "Autonomous Ships: A Review, Innovative Applications and Future Maritime Business Models." *Supply Chain Forum: An International Journal* 20 (4): 266–279. doi:10.1080/16258312.2019.1631714.
Pao, M. L. 1986. "An Empirical Examination of Lotka's Law." *Journal of the American Society for Information Science* 37 (1): 26–33. doi:10.1002/asi.4630370105.
Perera, L. P., and M. Brage. 2016. "Marine Engine Operating Regions under Principal Component Analysis to Evaluate Ship Performance and Navigation Behavior." *IFAC-PapersOnLine* 49 (23): 512–517. doi:10.1016/j.ifacol.2016.10.487.
Pham, K. H., J. Boy, and M. Luengo-Oroz. 2018. "Data Fusion to Describe and Quantify Search and Rescue Operations in the Mediterranean Sea." Paper presented at the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA). Turin, Italy.
Pietrzykowski, Z., and J. Uriasz. 2009. "The Ship Domain–a Criterion of Navigational Safety Assessment in an Open Sea Area." *The Journal of Navigation* 62 (1): 93–108. doi:10.1017/S0373463308005018.
Pons, P., and M. Latapy. 2005. "Computing Communities in Large Networks Using Random Walks." Paper presented at the International symposium on computer and information sciences. Istanbul, Turkey.

Sala, E., J. Mayorga, C. Costello, D. Kroodsma, M. L. D. Palomares, D. Pauly, U. Rashid Sumaila, and D. Zeller. 2018. "The Economics of Fishing the High Seas." *Science Advances* 4 (6): eaat2504. doi:10.1126/sciadv.aat2504.

Salipante, P., W. Notz, and J. Bigelow. 1982. "A Matrix Approach to Literature Reviews." *Research in Organizational Behavior* 4: 321–348.

Sanchez-Gonzalez, P.-L., D. Díaz-Gutiérrez, T. J. Leo, and L. R. Núñez-Rivas. 2019. "Toward Digitalization of Maritime Transport?" *Sensors* 19 (4): 926. doi:10.3390/s19040926.

Sanfilippo, F., L. I. Hatledal, K. Y. Pettersen, and H. Zhang. 2017. "A Benchmarking Framework for Control Methods of Maritime Cranes Based on the Functional Mockup Interface." *IEEE Journal of Oceanic Engineering* 43 (2): 468–483. doi:10.1109/JOE.2017.2691920.

Shahir, H. Y., U. Glasser, A. Y. Shahir, and H. Wehn. 2015. "Maritime Situation Analysis Framework: Vessel Interaction Classification and Anomaly Detection." Paper presented at the 2015 IEEE International Conference on Big Data (Big Data). Santa Clara, USA.

Sidibé, A., and G. Shu. 2017. "Study of Automatic Anomalous Behaviour Detection Techniques for Maritime Vessels." *The Journal of Navigation* 70 (4): 847–858. doi:10.1017/S0373463317000066.

Solmaz, B., E. Gundogdu, V. Yucesoy, A. Koç, and A. A. Alatan. 2018. "Fine-grained Recognition of Maritime Vessels and Land Vehicles by Deep Feature Embedding." *IET Computer Vision* 12 (8): 1121–1132. doi:10.1049/iet-cvi.2018.5187.

Tian, Z., F. Liu, L. Zhixiong, R. Malekian, and Y. Xie. 2017. "The Development of Key Technologies in Applications of Vessels Connected to the Internet." *Symmetry* 9 (10): 211. doi:10.3390/sym9100211.

Tsou, M.-C., and C.-K. Hsueh. 2010. "The Study of Ship Collision Avoidance Route Planning by Ant Colony Algorithm." *Journal of Marine Science and Technology* 18 (5): 746–756.

UNCTAD. 2018. "Review of Maritime Transoprt." In edited by R. Asariotis, M. Assaf, H. Benamara, J. Hoffmann, A. Premti, L. Rodríguez, M. Weller, and F. Youssef. Geneva, Switzerland: United Nations Conference on Trade and Development.

Valbi, E., F. Ricci, S. Capellacci, S. Casabianca, M. Scardi, and A. Penna. 2019. "A Model Predicting the PSP Toxic Dinoflagellate Alexandrium Minutum Occurrence in the Coastal Waters of the NW Adriatic Sea." *Scientific Reports* 9 (1): 4166. doi:10.1038/s41598-019-40664-w.

Van Eck, N., and L. Waltman. 2009. "Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping." *Scientometrics* 84 (2): 523–538. doi:10.1007/s11192-009-0146-3.

Wang, K., M. Liang, L. Yan, J. Liu, and R. W. Liu. 2019. "Maritime Traffic Data Visualization: A Brief Review." Paper presented at the 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA). Suzhou, China.

Yan, X., K. Wang, Y. Yuan, X. Jiang, and R. R. Negenborn. 2018. "Energy-efficient Shipping: An Application of Big Data Analysis for Optimizing Engine Speed of Inland Ships considering Multiple Environmental Factors." *Ocean Engineering* 169: 457–468. doi:10.1016/j.oceaneng.2018.08.050.

Yang, D., W. Lingxiao, S. Wang, H. Jia, and K. X. Li. 2019. "How Big Data Enriches Maritime Research–a Critical Review of Automatic Identification System (AIS) Data Applications." *Transport Reviews* 39 (6): 755–773. doi:10.1080/01441647.2019.1649315.

Yang, T., H. Feng, C. Yang, R. Deng, G. Guo, and L. Tieshan. 2018a. "Resource Allocation in Cooperative Cognitive Radio Networks Towards Secure Communications for Maritime Big Data Systems." *Peer-to-Peer Networking and Applications* 11 (2): 265–276. doi:10.1007/s12083-016-0482-z.

Yang, W., Y. Tian, L. Tie-shan, P. Dongcheng, and Z. Yihua. 2018b. "Visualization Analysis of Shipping Recruitment Information Based on R." Paper presented at the 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI). Xiamen, China.

Yeo, G., S. H. Lim, L. Wynter, and H. Hassan. 2019. "MPA-IBM Project SAFER: Sense-Making Analytics for Maritime Event Recognition." *Interfaces* 49 (4): 269–280. doi:10.1287/inte.2019.0997.

Zerbino, P., D. Aloini, R. Dulmin, and V. Mininno. 2018a. "Knowledge Management in PCS-enabled Ports: An Assessment of the Barriers." *Knowledge Management Research & Practice* 16 (4): 435–450. doi:10.1080/14778238.2018.1473830.

Zerbino, P., D. Aloini, R. Dulmin, and V. Mininno. 2018b. "Process-mining-enabled Audit of Information Systems: Methodology and an Application." *Expert Systems with Applications* 110: 80–92. doi:10.1016/j.eswa.2018.05.030.

Zerbino, P., D. Aloini, R. Dulmin, and V. Mininno. 2019. "Towards Analytics-Enabled Efficiency Improvements in Maritime Transportation: A Case Study in A Mediterranean Port." *Sustainability* 11 (16): 4473. doi:10.3390/su11164473.

Zhang, X., and J. S. L. Lam. 2019. "A Fuzzy Delphi-AHP-TOPSIS Framework to Identify Barriers in Big Data Analytics Adoption: Case of Maritime Organizations." *Maritime Policy and Management* 46 (7): 781–801. doi:10.1080/03088839.2019.1628318.

Zupic, I., and Č. Tomaž. 2015. "Bibliometric Methods in Management and Organization." *Organizational Research Methods* 18 (3): 429–472. doi:10.1177/1094428114562629.

Appendix

Appendix A. Cluster-wise studies.

Cluster 1: Digital transformation	Cluster 2: Applications of big data from AIS	Cluster 3: Energy efficiency	Cluster 4: Predictive analytics
Fruth and Teuteberg (2017)	Yang et al. (2019)	Brouer, Karsten, and Pisinger (2017)	Solmaz et al. (2018)
Heilig and Voß (2017)	Wang et al. (2019)	Lee et al. (2018a)	Coraddu et al. (2019)
Sanchez-Gonzalez et al. (2019)	Cheng et al. (2019)	Yan et al. (2018)	Coraddu et al. (2016)
Ghazanfari, Rouhani, and Jafari (2014)	Sala et al. (2018)	Liu et al. (2018)	Mojtahedzadeh et al. (2015)
Zhang and Lam (2019)	Pham, Boy, and Luengo-Oroz (2018)	Sanfilippo et al. (2017)	Gallego et al. (2019)
Lam, Lee, and Zhang (2019)	Jafarzadeh and Ingrid (2018)		Berbić et al. (2017)
Zerbino et al. (2019)	Shahir et al. (2015)		Valbi et al. (2019)
Zerbino et al. (2018a)	Handayani and Sediono (2015)		
Tian et al. (2017)	Sidibé and Shu (2017)		
Zerbino et al. (2018b)	Fletcher et al. (2018)		
	Yeo et al. (2019)		