Proceedings of the Sixth International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 7, Paper 7.3
Civil-Comp Press, Edinburgh, United Kingdom, 2024
ISSN: 2753-3239, doi: 10.4203/ccc.7.7.3

Knowledge Structure of Structural Health Monitoring Methods Applied to Railways: A Review Using CiteSpace From 2015-2023

A. C. Pires[1], A. Antunes Dos Santos[1] and G. F. M. D. Santos[2]

[1]School of Mechanical Engineering, State University of Campinas, Brazil
[2]Department of Mechanical Engineering, Federal University of Espirito Santo, Brazil

Abstract

Structural health monitoring has gained popularity in recent years with the technological advancement of sensor technology and data transmission via cloud computing. In the field of railway systems, structural health monitoring is becoming increasingly dynamic and interdisciplinary. This complexity makes it challenging for researchers to determine the current trends, identify research gaps, and understand key concepts. This paper presents a fast and systematic approach to conducting a bibliometric analysis of structural health monitoring methods applied to railways, aiming to give readers an overall understanding of the field. Utilizing data spanning from 2015 to 2023 from the Web of Science, this study identifies key publications, researchers, and institutions on the subject. Moreover, CiteSpace was used to provide intuitive visuals that reveal partnerships between institutions and emerging areas of research through clustering algorithms. The analysis indicates that structural health monitoring in railway applications will increasingly embrace interdisciplinarity, with an emphasis on data-driven methods such as deep learning and big data analytics. Although this application is specific, the step-by-step process aims to assist researchers in identifying promising areas and facilitating the literature review process.

Keywords: structural health monitoring, Web of Science, CiteSpace, railway, bibliometrics, review

1 Introduction

This study evaluates the state of research related to structural health monitoring (SHM) applied to railways. Bibliometric data was collected from Web of Science (WoS) from the beginning of 2015 to the end of 2023 and visualized using CiteSpace.

Learning about a research topic can be challenging, particularly without background knowledge, initial intuition, or a methodological approach. Simply "winging it" often proves inefficient, wasting valuable time that could be better spent on more important tasks. Scientometrics provides a quantitative tool to assess the importance of scientific publications, the interrelationships between authors or institutions, and the evolution of the research area over time.

The main objective of this paper is to offer a comprehensive guide for researchers interested in exploring structural health modeling for railway systems. Additionally, this paper seeks to assist emerging researchers in identifying novel topics and broadening existing ones. To accomplish this, the paper addresses the following questions:

1. Who are the main authors making significant contributions to the field?
2. Which papers are the most important?
3. Within structural health monitoring, which subjects or research areas are essential to understand?
4. Which institutions are leading in the realm of structural health monitoring for railway systems?
5. What are the current emerging trends and research hotspots?

2 Methods

This section will introduce the softwares used for the bibliometric study.

2.1 Web of Science

Web of Science (WoS) is one of the primary multidisciplinary research databases used for academic and research purposes. It enables the user to trace the citation history of articles and provides insights and visual illustrations that show the impact and progression of research over time. Researchers frequently use it for literature reviews, citation analysis, and monitoring scientific trends in research areas.

Using WoS, a series of keywords is chosen to filter the database and generate the dataset of works related to structural health monitoring applied to railways. These keywords are as follows: Topic: ("condition monitoring" OR "structural health monitoring" OR "drive-by monitoring" OR "asset monitoring" OR "track monitoring" OR "road monitoring" OR "bridge monitoring" OR "condition-based maintenance" OR "remaining useful life" OR "RUL" OR "fault detection" OR "anomaly detection" OR "damage detection" OR "prognostic" OR "track defect" OR "data-driven" OR "on-board monitoring" OR "onboard monitoring" OR "vibration monitoring"). From this list, only papers containing the word "railway" were selected. After this step, additional filters were applied to include only works in the English language and publications between 2015 and 2023. This process yielded a total of 1,628 papers at the time of this research. This bibliometric dataset, complete with full records and cited references, was analysed inside WoS before being exported as plain text files for use in CiteSpace.

Figure *1* illustrates the changing popularity of research related to structural health monitoring applied to railways over time. There has been a steady increase in both citations and the number of publications throughout the years.

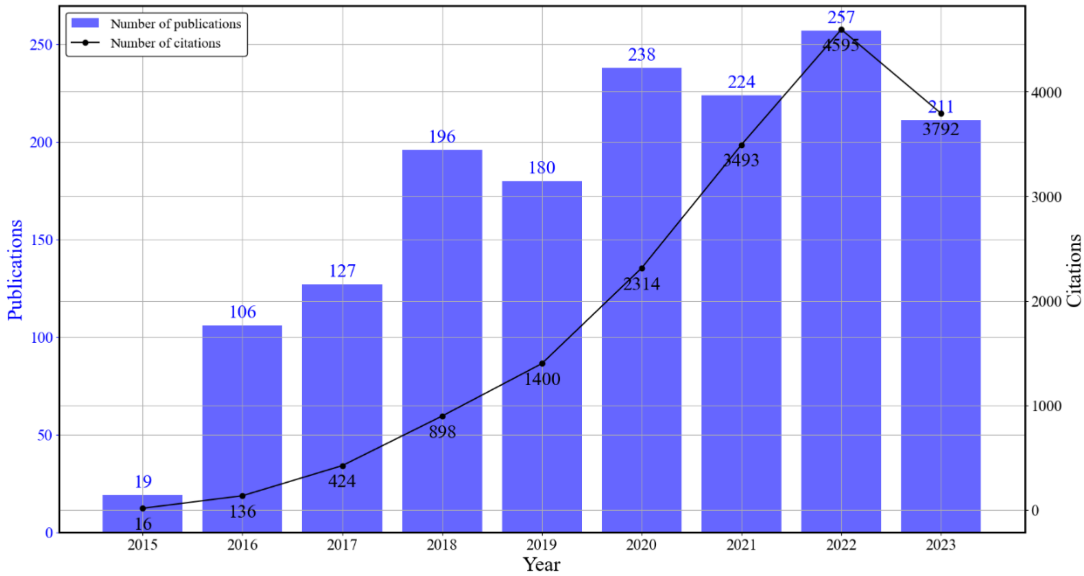


Figure 1: Research trend of structural health monitoring applied to railways between 2015 and 2023.

2.2 CiteSpace

Citespace is a software developed in Java by Dr. Chen Chaomei [1]. It offers insights into specific research areas and helps address questions that the user might have about their area of interest. Examples of such questions are provided at the end of Section 1.

CiteSpace (version 6.2. R4) was employed to visualize certain facets of the data that WoS struggles to display, such as research hotspots, collaborations between institutions, and co-citations, among others. It achieves this by utilizing clustering algorithms to produce visual maps and tables that emphasize the desired characteristics. The two main components of these maps are nodes and links [2]:

1. **Nodes:** These represent the objects the user wishes to analyze, such as cited authors, keywords, cited references, and the researcher's institution. They are depicted as concentric circles with varying colors. The node's size indicates its importance and nodes situated closer together might signify a high co-citation or close areas of research. Red rings highlight nodes that have experienced a sudden burst of popularity over time, suggesting they might be more important than others;
2. **Links:** They describe the relationships between nodes. The link's size indicates the strength of the relationship.

CiteSpace's features can be summarized into five main functions [1,2]:

1. **Co-citation analysis:** Papers frequently cited together in another paper are likely related in content, scope, or theme. Analyzing co-citation can pinpoint specific topics or research frontiers;
2. **Temporal analysis:** The number of citations of an article is influenced by the time since its publication and the context of its release, such as its novelty and the size of its field's community. Temporal analysis enables the user to observe how the citations or other research elements evolved, signaling potential emerging trends;
3. **Burst analysis:** This function is used to detect sudden changes over time. For instance, a sudden and substantial increase in citations might indicate a work or topic that has garnered significant attention in little time, suggesting the importance of the paper. For a node, this is denoted as represented by a red circle;
4. **Identifying research frontiers:** Analyzing recent citations can uncover emerging research areas or hotspots. For a node, this is indicated by a purple ring;
5. **Collaboration analysis:** This function identifies researchers or institutions that collaborate frequently, aiding in pinpointing influential groups or institutions within a field.

2.2 Performance metrics

Within CiteSpace, different metrics are employed to quantify the importance of a node and the quality of the clusters formed. Depending on the objective, a particular metric might be more suited for addressing the intended question. The main metrics used to measure network performance include [1,2]:

1. **Mean silhouette score**: This clustering metric gauges how similar a node is to its own cluster in comparison with other clusters. Its values range from -1 to +1, with larger values indicating greater siomilarity among cluster members;

2. **Modularity Q**: This metric evaluates the quality of the network's division into multiple clusters or modules. Its range is between 0 and 1. Large values suggest that the network has clear, distinct groups where nodes of the same cluster are more densely interconnected than nodes in different clusters. Values approaching 0 indicate that the division is as good as dividing randomly.

It is worth noting that the silhouette score can be misleading when applied to a cluster with very few members. In the visualization maps, a filter was applied to exclude nodes with no links to other nodes and with negligible paper counts. The main importance metrics used to quantify the importance of a node are [1]:

1. **Centrality:** This measures a node's influence and how a node disseminates its information. An analogy for the centrality metric is comparing a tollbooth on a heavily trafficked highway to one with low traffic like a tollbooth on a highway with large traffic vs light traffic. In heavy traffic, the tollbooth serves as a more significant bottleneck than in light traffic, making it more crucial for traffic flow. A higher centrality might represent a key point of a research field or indicate a close relationship with other research areas (heavy traffic);
2. **Citation burst:** This measured the sudden rise in a node's popularity over a specific timeframe. Larger values signify sudden attention or importance in the research community;
3. **Sigma score:** This metric merges both centrality and citation burst, considering the spatial aspect of centrality and the temporal aspect of the citation burst.

With these metrics in mind, Table 1 shows the parameters used inside CiteSpace to produce the visualization maps and tables.

No.	Parameter name	Value used
1	Time slicing	Year span from January 2015 to December 2023, 1 slice per year
2	Term source	Tile, abstract, author, keywords and keywords plus
3	Node type	Author, institution, cited reference, cited author, and keywords
4	Selection criteria	Top 10%
5	Pruning	Pathfinder and pruning sliced networks
6	Links	Default
7	Visualization	Cluster view-static and show merged network

Table 1: Parameters used for bibliometric analysis inside CiteSpace.

3 Results

The results section is divided into subsections aimed at answering each specific question made in the introduction. These questions are:

1. Who are the main authors making significant contributions to the field?
2. Which papers are the most important?
3. Within structural health monitoring, which subjects, or research areas, are essential to understand?
4. Which institutions are leading in the realm of structural health monitoring for railway systems?
5. What are the current emerging trends and research hotspots?

3.1 Who are the main authors making significant contributions?

The top 20 most influential authors in terms of the number of publications are shown in Table 2. The list shows that these authors have a relatively narrow range of contributions, ranging from 11 to 21. Clive Roberts leads with 21 publications, followed by You-Liang Ding with 19 publications. Around one-third of the total research output in this domain is given by the top 20 authors, which is significant. The railway sector inside structural health monitoring is a niche sector, which is a possible reason for the lower number of publications compared to other similar sectors such as structural health monitoring of wind turbines.

No.	Author	Count	Percentage (%)	No.	Author	Count	Percentage (%)
1	Roberts C	21	1.289	14	Ni YQ	13	0.798
2	Ding YL	19	1.166	15	Wang J	13	0.798
3	Kaewunruen S	18	1.105	16	Zhang Y	13	0.798
4	Núñez A	18	1.105	17	Calçada R	12	0.737
5	Jiang B	17	1.044	18	Li Q	12	0.737
6	Stewart E	17	1.044	19	Liu J	12	0.737
7	Wang P	17	1.044	20	Wei XK	12	0.737
8	Dollevoet R	15	0.921	21	Zhang WH	12	0.737
9	Li ZL	15	0.921	22	Bruni S	11	0.675
10	Liu ZG	15	0.921	23	Entezami M	11	0.675
11	Ribeiro D	14	0.859	24	Jia LM	11	0.675
12	Zhao HW	14	0.859	25	Obrien EJ	11	0.675
13	He Q	13	0.798				
				Sum			**35.91**

Table 2: Top 20 active authors of structural health monitoring for railway systems.

3.2 Who are the main authors making significant contributions?

Table 3 shows the top 12 most important articles based on the citation number. The paper with the most citations was from Hodge et al. [3] with 294 citations, followed by de Bruin et al. with 193 citations. Chen et al. [4] has the highest number of citations per year, with 91.5, indicating its importance in the field despite being the most recent paper of the table. The interdisciplinary nature of structural health monitoring in the railway sector was highlighted by the papers, where topics such as wireless sensor networks, deep learning, and big data analytics are all present.

No.	Author	Title	Count	Year	DOI	Citations per year
1	Hodge et al. [3]	Wireless Sensor Networks for Condition Monitoring in the Railway Industry: A Survey	294	2015	10.1109/TITS.2014.2366512	32.67
2	de Bruin et al. [5]	Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks	193	2017	10.1109/TNNLS.2016.2551940	27.57
3	Chen et al. [4]	Data-Driven Fault Diagnosis for Traction Systems in High-Speed Trains: A Survey, Challenges, and Perspectives	183	2022	10.1109/TITS.2020.3029946	91.5
4	Wang et al. [6]	Sparsity guided empirical wavelet transform for fault diagnosis of rolling element bearings	165	2018	10.1016/j.ymssp.2017.08.038	27.5
5	Ghofrani et al. [7]	Recent applications of big data analytics in railway transportation systems: A survey	157	2018	10.1016/j.trc.2018.03.010	26.17
6	Wu et al. [8]	Incipient winding fault detection and diagnosis for squirrel-cage induction motors equipped on CRH trains	156	2020	10.1016/j.isatra.2019.09.020	39
7	Sahal et al. [9]	Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case	132	2020	10.1016/j.jmsy.2019.11.004	33
8	Weston et al. [10]	Perspectives on railway track geometry condition monitoring from in-service railway vehicles	129	2015	10.1080/00423114.2015.1034730	14.33
9	Gou et al. [11]	An Open-Switch Fault Diagnosis Method for Single-Phase PWM Rectifier Using a Model-Based Approach in High-Speed Railway Electrical Traction Drive System	116	2016	10.1109/TPEL.2015.2465299	14.5
10	Neves et al. [12]	Structural health monitoring of bridges: a model-free ANN-based approach to damage detection	114	2017	10.1007/s13349-017-0252-5	16.29
11	Bešinović et al. [13]	Resilience in railway transport systems: a literature review and research agenda	101	2020	10.1080/01441647.2020.1728419	25.25
12	Feng et al. [14]	Model Updating of Railway Bridge Using In Situ Dynamic Displacement Measurement under Trainloads	101	2015	10.1061/(ASCE)BE.1943-5592.0000765	11.22

Table 3: Top 12 references for structural health monitoring applied to railways.

3.3 Which institutions are leading in the realm of structural health monitoring for railway systems?

To identify the most influential institutions, an institution co-authorship network was created. This network boasts a modularity Q of 0.6442 and a mean silhouette score of 0.7, suggesting good clustering performance.

Table 4 enumerates the top 20 most prominent institutions. It can be observed that Southwest Jiaotong University holds the top position with 140 publications, followed closely by Beijing Jiaotong University in the 2nd spot 125 publications. Notably, both are world-renowned Chinese universities.

Figure 2 shows the visualization network of the institution co-authorships, highlighting the top 4 members of each cluster. Over the years, a discernible trend has emerged: there's a rise in co-citations among universities specializing in different fields. In Figure 2a, the co-citations between institutions in 2016 appear sparse compared to the denser network depicted in Figure 2b. Beyond this trend, the field of structural health monitoring seems to be evolving towards greater interdisciplinarity. This is evident from the growing links between institutions from different clusters. While all nodes are associated with structural health monitoring in railways in some capacity, each cluster has its unique characteristics. The fusion of different specialties seems to be an emerging pattern.

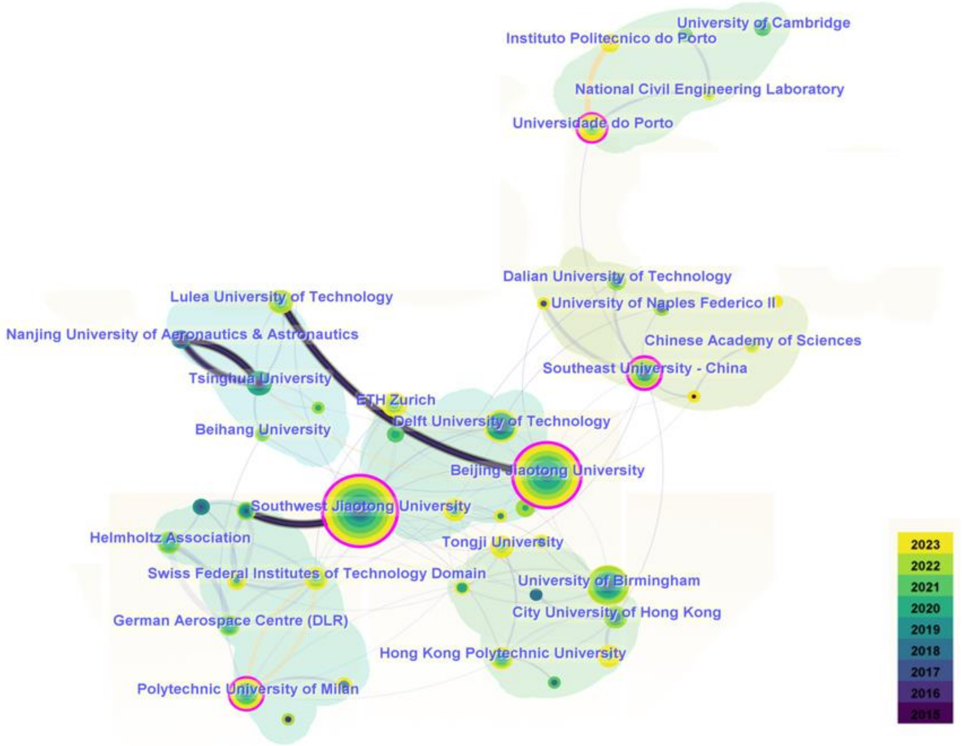
3.4 What are the research hotspots?

Using the same keyword network shown in Section 3.3., Table 5 was generated to show the top 25 keywords with the most citation bursts. Comparing the early years (2015-2017) to the later years (2020-2023), there has been a shift from modelling and prognostics, which are foundational concepts, to more advanced and specific computational techniques such as neural networks, unsupervised learning, and concepts such as digital twins. This can be visualized by the horizontal bar contained in the table, where red indicates a burst of citations, blue indicates a normal number of citations while the faded-out blue shows a lower number of citations. Note that digital twins only started to gain substantial traction in the railway sector from 2021, while deep learning started in 2019 and became a hotspot from 2021 onwards.

No.	Institution	Average publication year	Count	Percentage (%)	No.	Institution	Average publication year	Count	Percentage (%)
1	Southwest Jiaotong University	2016	140	8.60	11	Helmholtz Association	2017	20	1.23
2	Beijing Jiaotong University	2015	125	7.68	12	City University of Hong Kong	2017	19	1.17
3	University of Birmingham	2015	60	3.69	13	Tsinghua University	2015	19	1.17
4	Delft University of Technology	2015	46	2.83	14	ETH Zurich	2020	19	1.17
5	Polytechnic University of Milan	2015	43	2.64	15	Tongji University	2021	17	1.04
6	Southeast University - China	2015	33	2.03	16	German Aerospace Centre (DLR)	2017	15	0.92
7	Universidade do Porto	2015	27	1.66	17	Central South University	2019	14	0.86
8	Lulea University of Technology	2016	27	1.66	18	Norwegian University of Science & Technology (NTNU)	2016	13	0.86
9	Swiss Federal Institutes of Technology Domain	2015	24	1.48	19	Harbin Institute of Technology	2021	13	0.86
10	Hong Kong Polytechnic University	2016	23	1.41	20	University College Dublin	2017	12	0.80
					Sum				**44.56**

Table 4: The top 20 universities in terms of number of publications

a)



b)

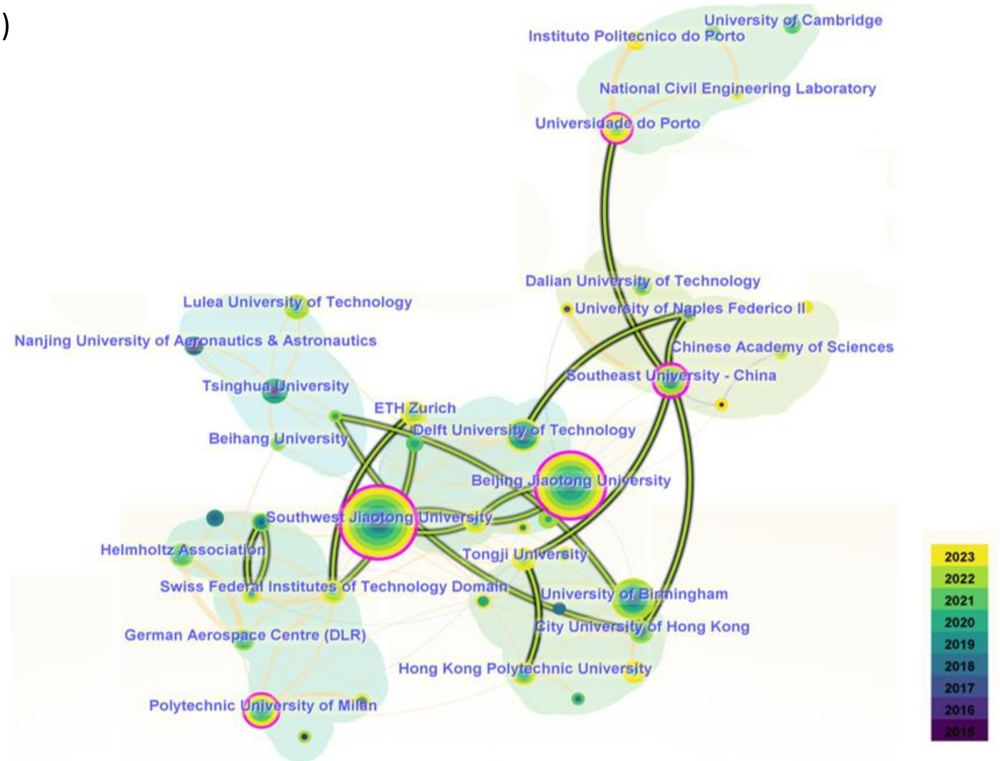


Figure 2: Visualization network map of institution co-authorship; (a) Highlight of the most frequent co-authorships of 2016. (b) Highlight of the most frequent co-authorships of 2022.

No.	Keyword	Strength	Start	End	Year (2015-2023)
1	models	3.59	2015	2017	
2	prognostics	3.42	2015	2020	
3	fault detection and isolation	3.07	2015	2017	
4	wireless sensor networks	3.07	2015	2017	
5	parameters	4.11	2017	2018	
6	speed	3.54	2017	2018	
7	energy harvesting	3.08	2017	2018	
8	empirical mode decomposition	3.53	2018	2020	
9	decomposition	2.99	2018	2018	
10	model	2.97	2018	2018	
11	defect detection	4.08	2019	2020	
12	damage	3.61	2019	2019	
13	neural networks	5.49	2020	2021	
14	rail transportation	3.78	2020	2021	
15	railway track	3.44	2020	2020	
16	feature extraction	5.76	2021	2023	
17	temperature	4.00	2021	2023	
18	vibration	2.89	2021	2021	
19	convolutional neural network	2.79	2021	2021	
20	data driven	2.79	2021	2021	
21	neural network	4.28	2022	2023	
22	deep learning	4.08	2022	2023	
23	unsupervised learning	3.92	2022	2023	
24	digital twin	3.22	2022	2023	
25	anomaly detection	3.18	2022	2023	

Table 5: Top 25 keywords with the most citation bursts from 2015 to 2023.

The highest strengths can be seen with feature extraction with 5.76 and neural networks with 5.49, both concepts intimately related to deep learning. Feature extraction, specifically, is about transforming measured data into features that can clearly distinguish between normal and abnormal conditions. This is done automatically with convolutional neural networks for example, which is item 19 of Table 7

Figure 3 shows a timeline view of the clusters identified by the keyword network (Table 5). Each horizontal line shown is a timeline for a specific cluster. From left to right, the year increases from 2015 to 2023. The size of the circle represents the nodes importance. If the color red is present, then it shows that this specific node had a citation burst. It can be seen that the largest number of citation bursts occurred in the deep learning cluster, as shown by the larger number of red circles in Figure 3. Therefore, it is likely that the future of SHM for railway systems will involve using deep learning and machine learning models even more to solve SHM and predictive maintenance problems.

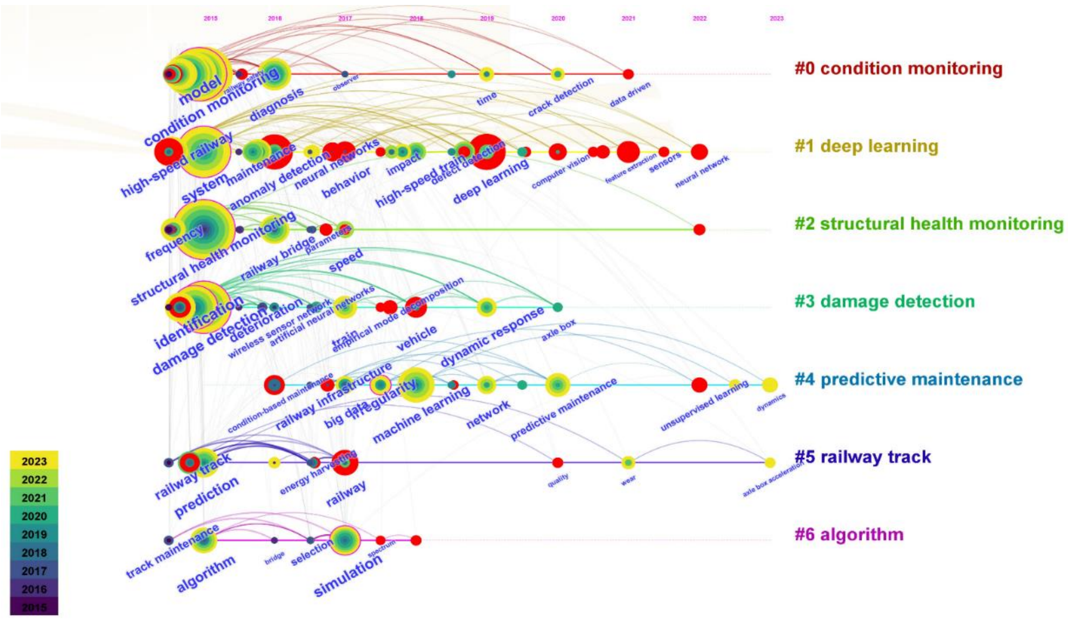


Figure 3: Timeline view of the keyword citation bursts that occurred between 2015 and 2023.

4 Conclusions and Contributions

Structural health monitoring has continued to gain popularity over the years and has become a dynamic and evolving field of research. This paper provides a systematic approach to conducting a bibliometric analysis of SHM methods applied to railways, aiming to give the reader a general view of the field. Furthermore, this study helps identify research gaps, research hotspots and future research directions.

Bibliometric data were collected from Web of Science (WoS) database, that spanning from 2015 to 2023 using filters and keywords. After analyzing the data, the following key conclusions were drawn:

1. The leading universities in the field are Southwest Jiaotong University with 140 publications, and Beijing Jiaotong with 125 publications, ocuppying the first and second spots, respectively. In total, they correspond to 16.28% of all publications in the field;
2. The knowledge clusters of SHM applied to railways are divided into 6 categories: condition monitoring, deep learning, structural health monitoring, damage detection, predictive maintenance, railway track and algorithms;
3. Collaboration between institutions has increased from 2015 to 2023 and is becoming more interdisciplinary, meshing traditional engineering concepts with deep learning and data analytics.
4. Data-driven methods and concepts (deep learning, neural networks, feature extraction, unsupervised learning), anomaly detection and digital twins are research hotspots at the end of 2023, with a trend of it becoming more popular over time.

Acknowledgements

The authors wish to express their acknowledgment to Vale S.A. for funding this study and technical support. Also, to CNPq (grant number 315304/2018-9), which partially funded this project.

References

[1] C. Chen, *CiteSpace: A Practical Guide for Mapping Scientific Literature CiteSpace: Text ming and visualization in scientific literatures View project Sandbox: Graphic Design View project*, no. December 2016. 2016. [Online]. Available: https://www.researchgate.net/publication/308204148
[2] H. Yang, X. Shao, and M. Wu, "A review on ecosystem health research: A visualization based on CiteSpace," *Sustain.*, vol. 11, no. 18, 2019, doi: 10.3390/su11184908.
[3] V. J. Hodge, S. O'Keefe, M. Weeks, and A. Moulds, "Wireless sensor networks for condition monitoring in the railway industry: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1088–1106, 2015, doi: 10.1109/TITS.2014.2366512.
[4] H. Chen, B. Jiang, S. X. Ding, and B. Huang, "Data-Driven Fault Diagnosis for Traction Systems in High-Speed Trains: A Survey, Challenges, and Perspectives,"

IEEE Trans. Intell. Transp. Syst., vol. 23, no. 3, pp. 1700–1716, 2022, doi: 10.1109/TITS.2020.3029946.
T. De Bruin, K. Verbert, and R. Babuska, "Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 3, pp. 523–533, 2017, doi: 10.1109/TNNLS.2016.2551940.
[6] D. Wang, Y. Zhao, C. Yi, K. L. Tsui, and J. Lin, "Sparsity guided empirical wavelet transform for fault diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 101, pp. 292–308, 2018, doi: 10.1016/j.ymssp.2017.08.038.
[7] F. Ghofrani, Q. He, R. M. P. Goverde, and X. Liu, "Recent applications of big data analytics in railway transportation systems: A survey," *Transp. Res. Part C Emerg. Technol.*, vol. 90, no. September 2017, pp. 226–246, 2018, doi: 10.1016/j.trc.2018.03.010.
[8] Y. Wu, B. Jiang, and Y. Wang, "Incipient winding fault detection and diagnosis for squirrel-cage induction motors equipped on CRH trains," *ISA Trans.*, vol. 99, pp. 488–495, 2020, doi: 10.1016/j.isatra.2019.09.020.
[9] R. Sahal, J. G. Breslin, and M. I. Ali, "Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case," *J. Manuf. Syst.*, vol. 54, no. December 2019, pp. 138–151, 2020, doi: 10.1016/j.jmsy.2019.11.004.
[10] P. Weston, C. Roberts, G. Yeo, and E. Stewart, "Perspectives on railway track geometry condition monitoring from in-service railway vehicles," *Veh. Syst. Dyn.*, vol. 53, no. 7, pp. 1063–1091, 2015, doi: 10.1080/00423114.2015.1034730.
[11] B. Gou, X. Ge, S. Wang, X. Feng, J. B. Kuo, and T. G. Habetler, "An Open-Switch Fault Diagnosis Method for Single-Phase PWM Rectifier Using a Model-Based Approach in High-Speed Railway Electrical Traction Drive System," *IEEE Trans. Power Electron.*, vol. 31, no. 5, pp. 3816–3826, 2016, doi: 10.1109/TPEL.2015.2465299.
[12] A. C. Neves, I. González, J. Leander, and R. Karoumi, "Structural health monitoring of bridges: a model-free ANN-based approach to damage detection," *J. Civ. Struct. Heal. Monit.*, vol. 7, no. 5, pp. 689–702, 2017, doi: 10.1007/s13349-017-0252-5.
[13] N. Bešinović, "Resilience in railway transport systems: a literature review and research agenda," *Transp. Rev.*, vol. 40, no. 4, pp. 457–478, 2020, doi: 10.1080/01441647.2020.1728419.
[14] D. Feng and M. Q. Feng, "Model Updating of Railway Bridge Using In Situ Dynamic Displacement Measurement under Trainloads," *J. Bridg. Eng.*, vol. 20, no. 12, pp. 1–12, 2015, doi: 10.1061/(asce)be.1943-5592.0000765.