 doi:10.20944/preprints202304.0052.v1

A Bibliometric Study of Machine Learning in Biofilm

Yuanzhao Ding [1,*]

[1] School of Geography and the Environment, University of Oxford, South Parks Road, Oxford OX1 3QY, UK; ORCID: 0000-0003-0116-3648
* Correspondence: armstrongding@163.com

Abstract: Biofilm is a complex community of microorganisms that are attached to surfaces and encased in a self-produced extracellular matrix. Machine learning (ML) techniques have been applied to various aspects of biofilm research, such as predicting biofilm formation, identifying key genes, and designing new therapeutic strategies. In this study, we conducted a bibliometric analysis of machine learning in biofilm research to provide a comprehensive overview of the current state of the field. We searched the Web of Science database for articles published included "machine learning biofilm". A total of 126 articles were identified and analysed. Our results showed that the number of publications on machine learning in biofilm has been increasing rapidly over the past decade, indicating a growing interest in the application of ML techniques to biofilm research. The analysis also revealed that the most common research topics in this area were related to biofilm formation, prediction, and control. Furthermore, the most frequently used ML techniques in biofilm research were artificial neural networks and support vector machines. Overall, our study provides valuable insights into the current trends and future directions of machine learning in biofilm research. It also highlights the importance of interdisciplinary collaboration between biofilm researchers and ML experts to drive innovation in this field.

Keywords: Bibliometric analysis; Biofilm; Big data; Machine learning; Artificial intelligence

1. Introduction

Biofilm is a fascinating and complex community of microorganisms that adhere to surfaces and secrete a self-produced extracellular matrix, known as a biofilm matrix [1]. This matrix acts as a protective barrier, making the biofilm resilient to antibiotics, immune system responses, and other environmental factors [2]. Biofilms are widely distributed in nature and can colonize various surfaces, including medical implants, water distribution systems, and food processing equipment [3]. Biofilms are associated with a range of problems, including infections, biofouling, and corrosion, making them a significant area of interest in scientific research [4].

Over the past decade, there has been a growing interest in the application of machine learning (ML) techniques in biofilm research [5-8]. ML is a subfield of artificial intelligence that enables computer systems to learn and improve from experience without being explicitly programmed [9]. By using ML algorithms, large and complex datasets generated from biofilm research, such as genomics, proteomics, and metabolomics data, can be analyzed to identify key genes, proteins, and metabolites that are involved in biofilm formation and function [10].

The objective of this study is to conduct a comprehensive bibliometric analysis of the use of ML in biofilm research to gain insights into the current state of the field [11]. Bibliometrics is a quantitative analysis of scientific publications that provides valuable insights into the structure, dynamics, and trends of a particular research area [12,13]. A bibliometric analysis of ML in biofilm research can identify the most influential publications, authors, and institutions in the field, as well as the most common research topics and trends [14].

This analysis can also reveal how researchers have used ML to advance our understanding of biofilm formation, growth, and function. The integration of ML and biofilm research has enabled researchers to analyse large and complex datasets to identify key factors that influence biofilm formation, growth, and function, such as environmental factors, genetic makeup, and metabolic processes. Furthermore, ML can also be used to develop new approaches to biofilm detection, prevention, and control, making it a valuable tool in the fight against biofilm-associated problems.

In this study, we will conduct a thorough search of relevant databases for articles published with "machine learning biofilm". By focusing on this relatively recent time, we can gain insights into the most current research trends in this area. We will then use bibliometric analysis tools, such as VOSviewer, to analyze the identified articles and generate bibliometric networks that can reveal important information about the structure and dynamics of the field. Our analysis will include co-authorship analysis, keyword co-occurrence analysis, and citation analysis, among others, to identify the most influential authors, institutions, and publications, as well as the most common research topics and trends.

Overall, this study will provide a comprehensive overview of the current state of the field of ML in biofilm research. By identifying key research themes, trends, and influential institutions in the field, we can better understand the challenges and opportunities associated with the integration of ML and biofilm research. Moreover, this analysis can guide future research efforts, helping to advance our understanding of the complex biological processes involved in biofilm formation, growth, and function.

2. Material and methods

To gain a better understanding of the current state of research on the application of ML in the context of biofilm, we conducted a thorough search of the Web of Science, ensuring that our analysis is up-to-date and relevant. Searching "machine learning biofilm," we identified a total of 126 articles that met our inclusion criteria [15-17]. To gain insights into the key themes and trends in this body of literature, we employed VOSviewer, a powerful software tool for constructing and visualizing bibliometric networks [18-20]. Specifically, we performed a range of bibliometric analyses, including co-occurrence analysis, country/region analysis, and institution analysis [21].

In our co-authorship analysis, we examined the most influential authors in the field of ML and biofilm [22]. This analysis revealed several highly productive authors who have made significant contributions to the field. In our co-occurrence analysis, we identified the most common research topics and themes within the body of literature [23]. Our analysis revealed that several key themes emerged, including biofilm formation, biofilm detection, and the use of ML to identify and predict bacterial growth patterns [24]. Additionally, we identified several subtopics within these broader themes, such as the role of different environmental factors in biofilm formation, and the development of ML algorithms to accurately predict bacterial growth [25].

Finally, our citation analysis revealed the most influential publications in this field, as well as the most highly cited articles [26]. We found that several articles have received many citations, indicating that they have had a significant impact on the field. Notably, many of these highly cited articles focused on the use of ML to predict biofilm formation, and the development of new algorithms to better understand bacterial growth patterns.

Our bibliometric analysis provides valuable insights into the current state of research on the application of ML in the context of biofilm [11]. By identifying the most influential authors, institutions, and publications, as well as the most common research themes and trends, we are better equipped to understand the challenges and opportunities associated with this exciting field of study [27,28]. Furthermore, our findings may inform future research efforts, helping to advance our understanding of the complex biological processes involved in biofilm formation and growth [29,30].

3. Results

Our analysis showed that the number of publications on ML in biofilm has been increasing rapidly over the past decade, indicating a growing interest in the application of ML techniques to biofilm research. The analysis also revealed that the most common

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2023 doi:10.20944/preprints202304.0052.v1

research topics in this area were related to biofilm formation, prediction, and control. Furthermore, the most frequently used ML techniques in biofilm research were artificial neural networks [31-33] and support vector machines [34-36].

In terms of authorship, the most influential authors were from a diverse range of institutions, including academic and research institutions, as well as industrial and commercial organizations. The most influential institutions were also diverse and included universities, research centres, and hospitals. The most highly cited publications were focused on the development of ML algorithms for predicting biofilm formation and identifying key genes involved in biofilm formation.

The central portion of Figure 1 depicts the two most significant words in the field, namely "machine learning" and "biofilm formation". Additionally, the figure highlights several other important words related to biofilm formation, such as "*Pseudomonas aeruginosa*", "expression", "adaptation", "evolution", "virulence", "growth", "dynamics", "mechanism", and "resistance". These words provide insights into the most critical bacteria (*P. aeruginosa*), processes (adaptation, expression, and growth), and study targets (dynamics, mechanism, and resistance) in biofilm formation research. The words related to machine learning, such as "prediction", "performance", and "system", are linked to the most significant target for ML in biofilm research. This analysis emphasizes the importance of machine learning, particularly in predicting the biofilm formation. Overall, Figure 1 provides a clear overview of the most important words in the field and their interrelationships.

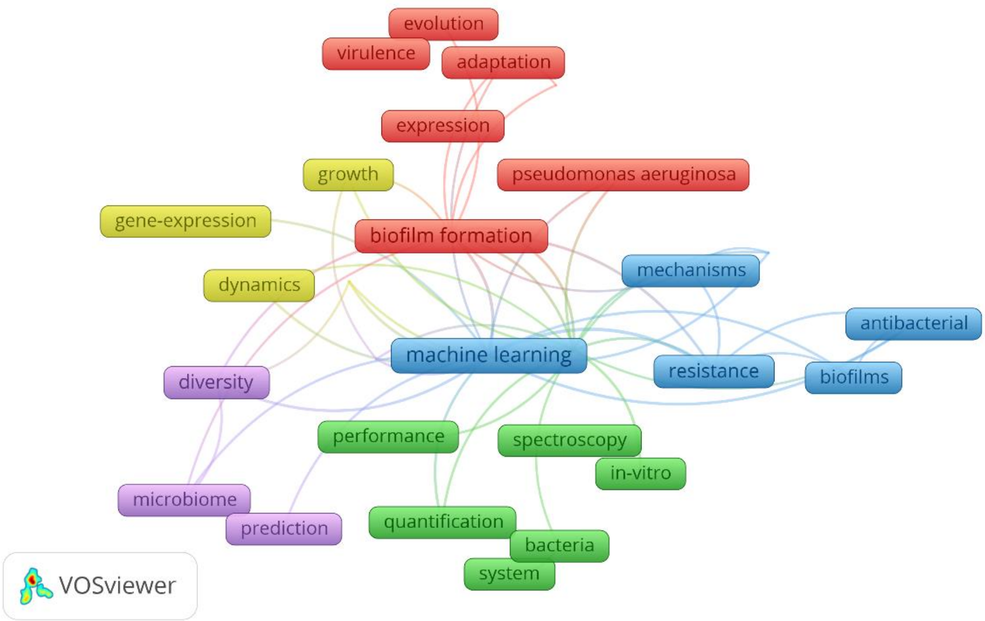

Figure 1. Most important words in this research field and their connection by VOSviewer.

As depicted in Figure 2, there is a global interest in research activities pertaining to the field under investigation, and this has resulted in numerous collaborations between different countries and regions. The visualization highlights the prominent roles of the United States and China in this area, with both countries demonstrating a high level of engagement and contributing significantly to the body of knowledge. However, other countries such as Canada, the United Kingdom, Japan, India, Spain, Germany, Switzerland, Italy, Sweden, Denmark, and South Korea have also been actively involved in this field and have made noteworthy contributions to the advancement of research. The contributions of these countries have enhanced the diversity of research perspectives and approaches, making the field richer and more comprehensive. It is evident that global collaborations and exchange of ideas have played a crucial role in driving progress in this area, and it is expected that this trend will continue in the future.

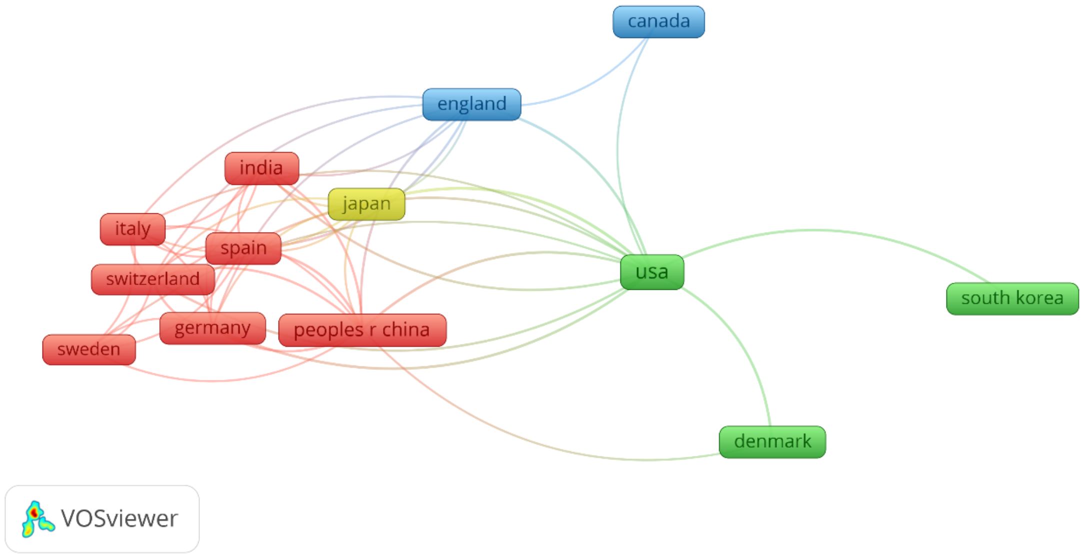

Figure 2. Scientific collaboration network from different countries by VOSviewer.

According to Figure 3, the most important institutions in this research area are Stanford University, the University of Illinois, the University of California at Irvine and San Diego, Sapienza University, the Agricultural University of Athens, Oregon State University, the University of Toronto, the Chinese Academy of Sciences, the Memorial Sloan Kettering Cancer Center, the Jožef Stefan Institute, and the Max Planck Society.

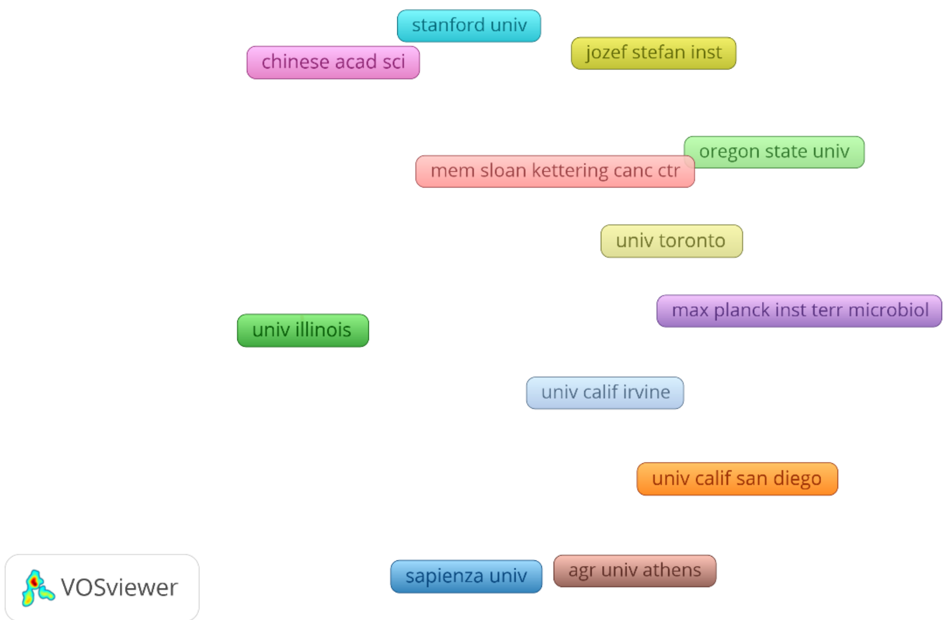

Figure 3. Most important institutions in the field by VOSviewer.

4. Discussion

4.1. Recent and significant papers

In the pursuit of developing effective strategies for managing bacterial biofilms, scientists have continued to explore the potential applications of ML techniques. Upon examining Table 1, which outlines recent and significant papers in the field, it is apparent that ML techniques have been widely employed in the study of biofilms. In fact, one

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2023 doi:10.20944/preprints202304.0052.v1

notable study conducted by a group of researchers utilized an ML model to predict the presence of biofilm inhibitory molecules [37]. This investigation incorporated a combination of descriptor, fingerprint, and hybrid models, and the accuracy of these models was found to be impressive, achieving 93%, 88%, and 90%, respectively. Additionally, the software created by this study, Molib, has become a widely utilized tool for predicting small molecules with biofilm inhibitory properties. The implications of this software's success are particularly promising, as it presents an opportunity for therapeutic intervention against bacteria that can form biofilms.

One such investigation involved the use of *P. aeruginosa*, a commonly studied model organism, to examine the chemical components of essential oils and their potential impact on biofilm formation [5,6]. In this study, the researchers utilized eleven different classification models (F1-F11) to analyse the data and assess the accuracy of the ML predictions. The results showed that the models achieved prediction accuracies ranging from 69% to 98%, demonstrating the effectiveness of ML in identifying essential oil chemical components that may impact biofilm formation. Through their analysis, the authors were able to identify specific essential oil chemical components that were likely responsible for modulating bacterial biofilm formation in both positive and negative ways. This information is invaluable for scientists who are working to develop interventions that can effectively manage biofilms and prevent their potentially harmful effects.

Another article discussed the challenges in treating biofilm-associated infections caused by *Staphylococcus aureus* and *Staphylococcus epidermidis* [7]. It explored the potential of essential oils (EOs) as a treatment option and analyzed the ability of 89 EOs to modulate biofilm production in different strains of the bacteria. ML algorithms were applied to the chemical compositions of the EOs to determine their anti-biofilm potencies and identify the chemical components responsible for biofilm production, inhibition, or stimulation for each strain.

In another study, EOs are investigated as natural alternatives to chemotherapeutic drugs for inhibiting biofilm in chronic *S. aureus* infections [8]. 61 EOs were tested for biofilm modulation and antibacterial activity. Chemical composition was analysed by GC/MS and ML algorithms correlated potency with active components. Select EOs inhibited biofilm growth at 1.00% concentration and were characterized for their ability to alter biofilm organization through scanning electron microscope (SEM) studies.

Another paper presented a novel computational methodology that combines meta-analysis and ML to identify important genes and pathways in biofilm-forming bacteria [38]. This approach was used to analyse gene expression profiles in different strains of *S. aureus* and identify a set of 36 candidate genes, 11 of which are reported for the first time. These genes are predicted to be important in biofilm formation and can be considered as a signature target list to develop anti-biofilm therapeutics. The study highlights the potential of combining meta-analysis and ML to gain deeper insights into biofilm mechanisms and develop effective therapeutic strategies.

Another study developed a machine-learning-aided cocktail assay for prompt and reliable biofilm detection [39]. Lanthanide nanoparticles with different properties were formulated into the cocktail kits, and the physicochemical heterogeneities of biofilms were transformed into luminescence intensity. The random forest algorithm was used to identify unknown biofilms with an overall accuracy of over 80%. Antibiotic-loaded cocktail nanoprobes efficiently eradicated biofilms, and the technique can serve as a reliable diagnostic tool for biofilm infections. It can also provide instructions for the design of assays for detecting biochemical compounds beyond biofilms.

Table 1. Summary for recent and important biofilm machine learning studies.

Model organism	Target/Biofilm process	ML models	ML accuracy	Main contributions	Year	Reference
NA	Biofilm inhibitory molecules	Classification	88% - 93%	ML to predict biofilm inhibitory molecules	2020	[37]
Pseudomonas aeruginosa	Essential oil chemical components	Binary Classification	69% - 98%	ML to identify chemical components responsible for bacterial biofilm formation	2018-2022	[5,6]
Staphylococcus aureus and Staphylococcus epidermidis	Essential oil chemical components	Binary Classification	68.7% - 90.6%	ML to identify chemical component that modulate biofilm production		[7]
S. aureus	Essential oil chemical components	Binary Classification	NA	ML to predict essential oils modulate biofilm production and inhibit biofilm	2019	[8]
S. aureus	acyl-CoA thioesterase	Classification	59.46 - 94.59%	Identification of 36 candidate genes including an acyl-CoA thioesterase enzyme and ten hypothetical proteins	2021	[38]
S. aureus, P. aeruginosa, Acinetobacter baumannii, Stenotrophomonas maltophilia, Escherichia coli	Biofilm infection	Random forest	95.0% - 100%	Using lanthanide nanoparticles detects pathogenic biofilms based on random forest	2022	[39]

4.2. Bibliometric analysis on the machine learning in biofilm research.

Our bibliometric analysis provides a comprehensive overview of the current state of ML in biofilm research. The analysis highlights the growing interest in the application of ML techniques to biofilm research and the importance of interdisciplinary collaboration between biofilm researchers and ML experts to drive innovation in this field [25].

The most common research topics in this area were related to biofilm formation, prediction, and control, which reflects the urgent need for new strategies for controlling biofilm-related problems [40]. The most frequently used ML techniques in biofilm research were artificial neural networks [31-33] and support vector machines [34-36], which suggests that these techniques are well-suited for analysing complex biofilm-related data.

In terms of authorship and institutions, our analysis showed that the most influential authors and institutions were from a diverse range of fields, which highlights the interdisciplinary nature of biofilm research. This diversity is important for driving innovation.

4.3. Future recommendation of using ML in bacterial and biofilm studies.

The application of big data and ML techniques has become increasingly prevalent in various fields, such as species distribution [41,42], education [43,44], and cancer prediction [45,46]. Based on machine learning, the policymakers can adjust the policy and help the people [41,43,47]. However, despite the numerous studies on bacteria and biofilm, the use of ML in this area remains limited, as highlighted in previous paragraphs.

Bacteria and biofilm are subjects of extensive research in different environmental and industrial settings, including pollutant removal [48-50], electricity generation [51-53], concrete enhancement [54-56], and multi-functional building [57-59]. To enhance or weaken biological processes, scientists have devised methods to genetically modify bacterial genes [48], resulting in stronger or weaker biofilms. Despite these advances, the underlying mechanisms of these biological processes remain uncertain [60].

Given the wide-ranging applications and importance of bacteria and biofilm research, it is imperative to explore the potential benefits of integrating ML techniques in this area [61]. By leveraging the vast amounts of data generated through research, ML can help unravel the complex interactions and mechanisms involved in bacterial processes and provide novel insights and predictions [62]. Furthermore, the development of new ML algorithms specifically tailored to the unique challenges posed by bacterial data can potentially yield more accurate and efficient results [63,64].

Overall, the integration of ML in bacteria and biofilm research has the potential to advance our understanding of these important biological processes and may lead to new discoveries and applications in various fields.

Reference

1. Flemming, H.-C.; Wingender, J. The biofilm matrix. *Nature reviews microbiology* **2010**, *8*, 623-633.
2. Dufour, D.; Leung, V.; Lévesque, C.M. Bacterial biofilm: structure, function, and antimicrobial resistance. *Endodontic Topics* **2010**, *22*, 2-16.
3. Galie, S.; García-Gutiérrez, C.; Miguélez, E.M.; Villar, C.J.; Lombó, F. Biofilms in the food industry: health aspects and control methods. *Frontiers in microbiology* **2018**, *9*, 898.
4. Wang, H.; Christiansen, D.E.; Mehraeen, S.; Cheng, G. Winning the fight against biofilms: the first six-month study showing no biofilm formation on zwitterionic polyurethanes. *Chemical Science* **2020**, *11*, 4709-4721.
5. Artini, M.; Papa, R.; Sapienza, F.; Božović, M.; Vrenna, G.; Tuccio Guarna Assanti, V.; Sabatino, M.; Garzoli, S.; Fiscarelli, E.V.; Ragno, R. Essential Oils Biofilm Modulation Activity and Machine Learning Analysis on Pseudomonas aeruginosa Isolates from Cystic Fibrosis Patients. *Microorganisms* **2022**, *10*, 887.
6. Artini, M.; Patsilinakos, A.; Papa, R.; Božović, M.; Sabatino, M.; Garzoli, S.; Vrenna, G.; Tilotta, M.; Pepi, F.; Ragno, R. Antimicrobial and antibiofilm activity and machine learning classification analysis of essential oils from different Mediterranean plants against Pseudomonas aeruginosa. *Molecules* **2018**, *23*, 482.
7. Patsilinakos, A.; Artini, M.; Papa, R.; Sabatino, M.; Božović, M.; Garzoli, S.; Vrenna, G.; Buzzi, R.; Manfredini, S.; Selan, L. Machine learning analyses on data including essential oil chemical composition and in vitro experimental antibiofilm activities against Staphylococcus species. *Molecules* **2019**, *24*, 890.
8. Papa, R.; Garzoli, S.; Vrenna, G.; Sabatino, M.; Sapienza, F.; Relucenti, M.; Donfrancesco, O.; Fiscarelli, E.V.; Artini, M.; Selan, L. Essential oils biofilm modulation activity, chemical and machine learning analysis—Application on Staphylococcus aureus isolates from cystic fibrosis patients. *International Journal of Molecular Sciences* **2020**, *21*, 9258.
9. Lavallin, A.; Downs, J.A. Machine learning in geography–Past, present, and future. *Geography Compass* **2021**, *15*, e12563.
10. Johnson, C.H.; Ivanisevic, J.; Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews Molecular cell biology* **2016**, *17*, 451-459.
11. Li, P.; Tong, X.; Wang, T.; Wang, X.; Zhang, W.; Qian, L.; Liao, J.; Diao, W.; Zhou, J.; Wu, W. Biofilms in wound healing: A bibliometric and visualised study. *International Wound Journal* **2023**, *20*, 313-327.
12. Talafidaryani, M.; Jalali, S.M.J.; Moro, S. Tracing the evolution of digitalisation research in business and management fields: Bibliometric analysis, topic modelling and deep learning trend forecasting. *Journal of Information Science* **2023**, 01655515221148365.
13. Anwar, M.A.; Zhang, Q.; Asmi, F.; Hussain, N.; Plantinga, A.; Zafar, M.W.; Sinha, A. Global perspectives on environmental kuznets curve: A bibliometric review. *Gondwana Research* **2022**, *103*, 135-145.
14. Zhang, S.; Mao, G.; Crittenden, J.; Liu, X.; Du, H. Groundwater remediation from the past to the future: A bibliometric analysis. *Water Research* **2017**, *119*, 114-125.
15. AlRyalat, S.A.S.; Malkawi, L.W.; Momani, S.M. Comparing bibliometric analysis using PubMed, Scopus, and Web of Science databases. *JoVE (Journal of Visualized Experiments)* **2019**, e58494.
16. Archambault, É.; Campbell, D.; Gingras, Y.; Larivière, V. Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American society for information science and technology* **2009**, *60*, 1320-1326.
17. Gorraiz, J.; Schloegl, C. A bibliometric analysis of pharmacology and pharmacy journals: Scopus versus Web of Science. *Journal of information science* **2008**, *34*, 715-725.

18. Shah, S.H.H.; Lei, S.; Ali, M.; Doronin, D.; Hussain, S.T. Prosumption: bibliometric analysis using HistCite and VOSviewer. *Kybernetes* **2020**, *49*, 1020-1045.
19. Ji, B.; Zhao, Y.; Vymazal, J.; Mander, Ü.; Lust, R.; Tang, C. Mapping the field of constructed wetland-microbial fuel cell: A review and bibliometric analysis. *Chemosphere* **2021**, *262*, 128366.
20. Ramírez-Malule, H.; Quinones-Murillo, D.H.; Manotas-Duque, D. Emerging contaminants as global environmental hazards. A bibliometric analysis. *Emerging contaminants* **2020**, *6*, 179-193.
21. Zhu, Y.; Li, J.J.; Reng, J.; Wang, S.; Zhang, R.; Wang, B. Global trends of Pseudomonas aeruginosa biofilm research in the past two decades: A bibliometric study. *MicrobiologyOpen* **2020**, *9*, 1102-1112.
22. Liu, P.; Xia, H. Structure and evolution of co-authorship network in an interdisciplinary research field. *Scientometrics* **2015**, *103*, 101-134.
23. Xue, J.; Reniers, G.; Li, J.; Yang, M.; Wu, C.; van Gelder, P. A bibliometric and visualized overview for the evolution of process safety and environmental protection. *International Journal of Environmental Research and Public Health* **2021**, *18*, 5985.
24. Ju, Y.; Zhang, F.; Yu, P.; Zhang, Y.; Zhao, P.; Xu, P.; Sun, L.; Bao, Y.; Long, H. A Bibliometric Analysis of Research on Bacterial Persisters. *BioMed Research International* **2023**, *2023*.
25. Hashemi, S.J.; Bak, N.; Khan, F.; Hawboldt, K.; Lefsrud, L.; Wolodko, J. Bibliometric analysis of microbiologically influenced corrosion (MIC) of oil and gas engineering systems. *Corrosion* **2018**, *74*, 468-486.
26. Rickert, C.A.; Hayta, E.N.; Selle, D.M.; Kouroudis, I.; Harth, M.; Gagliardi, A.; Lieleg, O. Machine learning approach to analyze the surface properties of biological materials. *ACS Biomaterials Science & Engineering* **2021**, *7*, 4614-4625.
27. Qi, Y.; Chen, X.; Hu, Z.; Song, C.; Cui, Y. Bibliometric analysis of algal-bacterial symbiosis in wastewater treatment. *International Journal of Environmental Research and Public Health* **2019**, *16*, 1077.
28. Zhang, T.; Yin, X.; Yang, X.; Man, J.; He, Q.; Wu, Q.; Lu, M. Research trends on the relationship between microbiota and gastric cancer: a bibliometric analysis from 2000 to 2019. *Journal of Cancer* **2020**, *11*, 4823.
29. Moura, L.K.B.; Tapety, F.I.; Mobim, M.; Lago, E.C.; de LobÃ, E.S.; Leal, C.M.d.C.L.; Santos, T.C.; Monte, T.L. Bacterial association and oral biofilm formation: A bibliometric analysis. *African Journal of Microbiology Research* **2016**, *10*, 1654-1661.
30. Colares, G.S.; Dell'Osbel, N.; Wiesel, P.G.; Oliveira, G.A.; Lemos, P.H.Z.; da Silva, F.P.; Lutterbeck, C.A.; Kist, L.T.; Machado, Ê.L. Floating treatment wetlands: A review and bibliometric analysis. *Science of the Total Environment* **2020**, *714*, 136776.
31. Lesnik, K.L.; Liu, H. Predicting microbial fuel cell biofilm communities and bioreactor performance using artificial neural networks. *Environmental science & technology* **2017**, *51*, 10881-10892.
32. Lahiri, D.; Nag, M.; Sarkar, T.; Dutta, B.; Ray, R.R. Antibiofilm activity of α-amylase from Bacillus subtilis and prediction of the optimized conditions for biofilm removal by response surface methodology (RSM) and artificial neural network (ANN). *Applied Biochemistry and biotechnology* **2021**, *193*, 1853-1872.
33. de Ramón-Fernández, A.; Salar-García, M.J.; Fernández, D.R.; Greenman, J.; Ieropoulos, I.A. Evaluation of artificial neural network algorithms for predicting the effect of the urine flow rate on the power performance of microbial fuel cells. *Energy* **2020**, *213*, 118806.
34. Shengxian, C.; Yanhui, Z.; Jing, Z.; Dayu, Y. Experimental study on dynamic simulation for biofouling resistance prediction by least squares support vector machine. *Energy Procedia* **2012**, *17*, 74-78.
35. Modak, S.; Lahorkar, A.; Valadi, J. Recent Advances in Applications of Support Vector Machines in Fungal Biology. *Laboratory Protocols in Fungal Biology: Current Methods in Fungal Biology* **2022**, 117-136.
36. Li, X.; Chen, S.; Zhang, J.; Yu, L.; Chen, W.; Zhang, Y. Optimization of Ultrasonic-Assisted Extraction of Active Components and Antioxidant Activity from Polygala tenuifolia: A Comparative Study of the Response Surface Methodology and Least Squares Support Vector Machine. *Molecules* **2022**, *27*, 3069.
37. Srivastava, G.N.; Malwe, A.S.; Sharma, A.K.; Shastri, V.; Hibare, K.; Sharma, V.K. Molib: A machine learning based classification tool for the prediction of biofilm inhibitory molecules. *Genomics* **2020**, *112*, 2823-2832.
38. Subramanian, D.; Natarajan, J. Integrated meta-analysis and machine learning approach identifies acyl-CoA thioesterase with other novel genes responsible for biofilm development in Staphylococcus aureus. *Infection, Genetics and Evolution* **2021**, *88*, 104702.
39. Wang, J.; Jiang, Z.; Wei, Y.; Wang, W.; Wang, F.; Yang, Y.; Song, H.; Yuan, Q. Multiplexed identification of bacterial biofilm infections based on machine-learning-aided lanthanide encoding. *ACS nano* **2022**, *16*, 3300-3310.

40. Alaoui Mdarhri, H.; Benmessaoud, R.; Yacoubi, H.; Seffar, L.; Guennouni Assimi, H.; Hamam, M.; Boussettine, R.; Filali-Ansari, N.; Lahlou, F.A.; Diawara, I. Alternatives Therapeutic Approaches to Conventional Antibiotics: Advantages, Limitations and Potential Application in Medicine. *Antibiotics* **2022**, *11*, 1826.
41. Chen, S.; Ding, Y. Machine Learning and Its Applications in Studying the Geographical Distribution of Ants. *Diversity* **2022**, *14*, 706.
42. Gobeyn, S.; Mouton, A.M.; Cord, A.F.; Kaim, A.; Volk, M.; Goethals, P.L.M. Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecological Modelling* **2019**, *392*, 179-195.
43. Chen, S.; Ding, Y. A Machine Learning Approach to Predicting Academic Performance in Pennsylvania's Schools. *Social Sciences* **2023**, *12*, 118.
44. Chen, S.; Ding, Y.; Liu, X. Development of the growth mindset scale: Evidence of structural validity, measurement model, direct and indirect effects in Chinese samples. *Current Psychology* **2021**, 1-15.
45. Cammarota, G.; Ianiro, G.; Ahern, A.; Carbone, C.; Temko, A.; Claesson, M.J.; Gasbarrini, A.; Tortora, G. Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nature reviews gastroenterology & hepatology* **2020**, *17*, 635-648.
46. Chen, S.; Ding, Y. A Feasibility Study of Machine Learning Models for Cancer Rate Prediction. *Preprints.org* **2023**, 2023030491.
47. Ding, Y. Heavy metal pollution and transboundary issues in ASEAN countries. *Water Policy* **2019**, *21*, 1096-1106.
48. Ding, Y.; Peng, N.; Du, Y.; Ji, L.; Cao, B. Disruption of putrescine biosynthesis in Shewanella oneidensis enhances biofilm cohesiveness and performance in Cr(VI) immobilization. *Appl Environ Microbiol* **2014**, *80*, 1498-1506, doi:10.1128/AEM.03461-13.
49. Ding, Y.; Zhou, Y.; Yao, J.; Szymanski, C.; Fredrickson, J.K.; Shi, L.; Cao, B.; Zhu, Z.; Yu, X.Y. In situ molecular imaging of the biofilm and its matrix. *Anal Chem* **2016**, *88* 11244-11252, doi:10.1021/acs.analchem.6b03909.
50. Ding, Y.; Zhou, Y.; Yao, J.; Xiong, Y.; Zhu, Z.; Yu, X.Y. Molecular evidence of a toxic effect on a biofilm and its matrix. *Analyst* **2019**, *144*, 2498-2503 doi:10.1039/c8an02512f.
51. Zhao, C.e.; Wu, J.; Ding, Y.; Wang, V.B.; Zhang, Y.; Kjelleberg, S.; Loo, J.S.C.; Cao, B.; Zhang, Q. Hybrid conducting biofilm with built-in bacteria for high-performance microbial fuel cells. *ChemElectroChem* **2015**, *2*, 654-658.
52. Yang, Y.; Ding, Y.; Hu, Y.; Cao, B.; Rice, S.A.; Kjelleberg, S.; Song, H. Enhancing bidirectional electron transfer of Shewanella oneidensis by a synthetic flavin pathway. *ACS synthetic biology* **2015**, *4*, 815-823.
53. Zhao, C.-e.; Chen, J.; Ding, Y.; Wang, V.B.; Bao, B.; Kjelleberg, S.; Cao, B.; Loo, S.C.J.; Wang, L.; Huang, W. Chemically functionalized conjugated oligoelectrolyte nanoparticles for enhancement of current generation in microbial fuel cells. *ACS Applied Materials & Interfaces* **2015**, *7*, 14501-14505.
54. Zhang, Z.; Liu, D.; Ding, Y.; Wang, S. Mechanical performance of strain-hardening cementitious composites (SHCC) with bacterial addition. *Journal of Infrastructure Preservation and Resilience* **2022**, *3*, 1-11.
55. Zhang, Z.; Ding, Y.; Qian, S. Influence of bacterial incorporation on mechanical properties of engineered cementitious composites (ECC). *Construction and Building Materials* **2019**, *196*, 195-203.
56. Zhang, Z.; Weng, Y.; Ding, Y.; Qian, S. Use of Genetically Modified Bacteria to Repair Cracks in Concrete. *Materials* **2019**, *12*, 3912.
57. Hamdany, A.H.; Ding, Y.; Qian, S. Visible light antibacterial potential of graphene-TiO2 cementitious composites for self-sterilization surface. *Journal of Sustainable Cement-Based Materials* **2022**, 1-11.
58. Hamdany, A.H.; Ding, Y.; Qian, S. Cementitious Composite Materials for Self-Sterilization Surfaces. *ACI Materials Journal* **2022**, *119*.
59. Hamdany, A.H.; Ding, Y.; Qian, S. Mechanical and antibacterial behavior of photocatalytic lightweight engineered cementitious composites. *Journal of Materials in Civil Engineering* **2021**, *33*, 04021262.
60. Tribedi, P.; Gupta, A.D.; Sil, A.K. Adaptation of Pseudomonas sp. AKS2 in biofilm on low-density polyethylene surface: an effective strategy for efficient survival and polymer degradation. *Bioresources and Bioprocessing* **2015**, *2*, 1-10.
61. Sadeghi, M.; Panahi, B.; Mazlumi, A.; Hejazi, M.A.; Komi, D.E.A.; Nami, Y. Screening of potential probiotic lactic acid bacteria with antimicrobial properties and selection of superior bacteria for application as biocontrol using machine learning models. *LWT* **2022**, *162*, 113471.
62. Long, F.; Fan, J.; Xu, W.; Liu, H. Predicting the performance of medium-chain carboxylic acid (MCCA) production using machine learning algorithms and microbial community data. *Journal of Cleaner Production* **2022**, *377*, 134223.

63. Cordier, T.; Lanzén, A.; Apothéloz-Perret-Gentil, L.; Stoeck, T.; Pawlowski, J. Embracing environmental genomics and machine learning for routine biomonitoring. *Trends in microbiology* **2019**, *27*, 387-397.
64. Long, F.; Wang, L.; Cai, W.; Lesnik, K.; Liu, H. Predicting the performance of anaerobic digestion using machine learning algorithms and genomic data. *Water Research* **2021**, *199*, 117182.