



THE BIBLIOMETRIC ANALYSIS OF MACHINE LEARNING USE IN THE DETECTION OF PHYCOCYANIN PIGMENT

Muhammad Haziq Naim Suhaini^a, Nur Afiqah Rosly^{a*}, Sharul Tazrajiman Tajudin^b, Khaleel Ahmad^c

^a Department of Maritime Science, Faculty of Defence Science and Technology, National Defence University of Malaysia, Sg. Besi Camp, 57000 Kuala Lumpur, Malaysia

^b School of Computing and Informatics, Universiti Teknologi Brunei, Jalan Tungku Link Gadong, BE1410 Brunei Darussalam

^c Department of Computer Science & Information Technology, School of Technology, Maulana Azad National Urdu University, Gachibowli, Hyderabad, Telangana-500032, India

ARTICLE INFO

ARTICLE HISTORY

Received : 20-03-2025

Revised : 15-05-2025

Accepted : 01-07-2025

Published : 30-11-2025

KEYWORDS

Bibliometric analysis

Harmful algal blooms

Machine learning

Phycocyanin

ABSTRACT

In freshwater systems, cyanobacteria harmful algal blooms (HABs) have been a major source of worry for environmental and public health agencies around the world. Machine learning can be used to detect Phycocyanin pigment which is an indicator to identify HABs in the water area. However, the use of machine learning is still low compared to remote sensing method. This research was conducted bibliometric analysis by using VOS Viewer software to show the gap and evolution of this research topic through published works. This research used the data of the publications from Scopus database which using machine learning to detect Phycocyanin pigment instead of remote sensing. It has shown that the machine learning method in Phycocyanin detection became more common in scientific community. The total publications that mentioned machine learning in Phycocyanin pigment detection in HABs has been increased gradually started from 2012 and the momentum still going strong. For the conclusion, machine learning has been used more frequently compared to 20 years ago in detection of Phycocyanin pigment in HABs and more researchers became more interested to make research in this specific field.

1.0 INTRODUCTION

Cyanobacteria are phytoplankton which frequently found in freshwater lakes, and when given the right weather conditions, they could grow vast numbers under eutrophic conditions, forming harmful algae blooms (HABs) [6]. Rousso et al. (2022) stated phycocyanin is a cyanobacterial auxiliary photosynthetic pigment that is increasingly being employed for cyanobacteria quantitation [9]. In freshwater systems, cyanobacteria harmful algal blooms have been a major source of worry for environmental and public health agencies around the world. HABs released toxic substances that degraded the water quality and posed health risks to humans and aquatic organisms [7]. For instance, the Taihu Lake ecosystem's balance is destroyed by the HABs epidemic, which also pollutes the lake's water [2]. The increment in algae bloom also reported at the Guarapiranga and Billings reservoirs at Sao Paulo, Brazil [1]. In the case of Guarapiranga Reservoir, the amount of phycocyanin compared to April 2019, climbed by about 500% in April 2020 [1]. The significant increase in algae bloom has crucial effects to the marine environment. HABs have a variety of detrimental effects on ecosystems, including the production of toxic biotoxins, damage to or obstruction of aquatic organisms' gills, and the development of anoxia or hypoxia conditions in the water column that may result in a large-scale event of aquatic ecosystem extinction. Many potentially fatal diseases in humans, such as Ciguatera Fish Poisoning (CFP), Amnesic Shellfish Poisoning (ASP), Neurotoxic Shellfish Poisoning (NSP), and Paralytic Shellfish Poisoning, are later caused by the biotoxins or phytotoxins that accumulate in the tissue or gastrointestinal tracts of the contaminated aquatic organisms

*Corresponding Author | Suhani, M. H. N. | afiqah@upnm.edu.my

© The Authors 2025. Published by Penerbit UPNM. This is open access article under the CC BY license.

(PSP). Biotoxin-containing shellfish and fish have been blamed for fatalities and have been observed in in The United States of America, Mexico, Canada, Chile, and Korea [12]. Furthermore, the ability of HABs to create poisons in dense biomass makes them a severe public health hazard across the globe. HABs can cause dermatitis, gastrointestinal distress, liver failure, or even the demise of domestic and livestock animals when they are exposed to toxins-containing water [1]. They can also cause issues with fisheries, shipping, recreational activities, water quality, and treatment [10].

Southeast Asia has seen HABs since the 1970s, and the situation there is still worrying [17]. In the early to mid-1990s, Paralytic Shellfish Toxin (PST)-producing species blooms in the Philippines did rise in frequency and duration, but they have since stabilised. Yet, the number of locations impacted by these blooms is still increasing, albeit more slowly than it was in the 1990s. In contrast, Malaysia did not experience an increase in the frequency of poisonous blooms during the three decades following the discovery of *Pyrodinium bahamense* in 1976. Ciguatera fish poisoning cases are reported from the Philippines and Malaysia, although the causative organisms are still poorly understood. Amnesic shellfish poisoning (ASP) has not yet been established in either the Philippines or Malaysia. Although there has not been a considerable expansion in the known populations during the past ten years, awareness of the spread of additional PST-producing species in Southeast Asia, such as species of *Alexandrium* and *Gymnodinium catenatum*, has grown since the 1990s and early 2000s. According to Yñiguez et al (2021), the regular fish-killing algal blooms of diverse species, including *Prorocentrum cordatum*, *Margalefidinium polykrikoides*, *Chattonella spp.*, and unarmored dinoflagellates, have become a significant more recent problem in the two countries as well as for Southeast Asia in general (e. g., *Karlodinium australe* and *Takayama sp.*) [17].

Due to the recent increase in the frequency of bloom outbreaks in the Strait of Malacca, research and monitoring into HABs have been heavily pursued in Malaysia. There have been reports of several HABs species from the Strait, including the red tide exterminators *Cochlodinium sp.*, *Ceratium sp.*, and *Noctiluca scintillans*, as well as hazardous *Alexandrium species* and *Pyrodinium bahamense*. Millions of dollars have been lost because of numerous bloom occurrences that resulted in significant fish fatalities at aquaculture locations in the northern half of the Strait [14]. The public health issue on the east coast of Peninsular Malaysia has been brought up by recent reports on occurrences of paralytic shellfish poisoning (PSP) between 2013 and 2014 at Kuantan Port, Pahang, Malaysia. Ten individuals were affected by the occurrences after eating paralytic shellfish toxins-contaminated rock oysters (PSTs). *A. tamiyavanichii* has been identified as the causal organism following the events and a fast reaction monitoring effort. This species has been found in Malaysian seas before; it was initially identified in Sebatu, Malacca, in 1991, where three people had a poisoning after eating mussels that had been contaminated with PSTs. Moreover, reports of the species have come from a few tropical and subtropical coastal nations, including Thailand, the Philippines, Brazil, and Japan [4].

Several studies have documented HABs occurrences that took place in Indonesia's eutrophic and hypertrophic coastal areas and estuaries, including those in Ambon Bay, Jakarta Bay, Pieh Island-West Sumatera and Lampung Bay. Phytotoxin contamination of shellfish or fish has also been linked to poisoning incidents in several Indonesian locales, including Sebatik Island in East Kalimantan, the Makassar waters, Ambon Bay, Cirebon, the Lewotobi and Lewouran regions of East Nusa Tenggara, and Lampung Bay. In Benoa Bay, 29 different phytoplankton species were found. Potentially toxic microalgae included *Coscinodiscus spp.*, *Pseudo-nitzschia spp.*, *Skeletonema spp.*, *Chaetoceros spp.*, *Rhizosolenia spp.*, and *Ceratium spp.* *Coscinodiscus spp.* stood out among them as the most prevalent and widely spread species in Benoa Bay [12]. The results of the HABs study in Ambon Bay showed that *Trichodesmium*, a hazardous algae genus from the Cyanophyta division, was found in Ambon Bay [5]. Chlorophyll-a (Ch-a), which is regarded as a general sign for plankton biomass, and phycocyanin, a distinctive pigment of inland water cyanobacteria, are the main indicators that have been utilised to detect HABs [1]. The most prevalent light-harvesting phycobiliprotein is phycocyanin, which is present in the phycobilisome complex of cyanobacteria and red algae. Phycocyanin serves an essential function in photosynthesis and offers a wide range of potential uses in foods, cosmetics, and medical diagnosis in addition to being fluorescent labelling probes or photosensitizers [11]. The presence of the pigment phycocyanin gives cyanobacteria, other cryptophytes, and red algae their distinctive blue colour [8].

Currently, satellite remote sensing and marine station monitoring (MSM) are utilised as a typical way to identify HABs [16]. However, there are considerable disadvantages to employing satellite remote sensing data for HABs monitoring. Both optical images (passive remote sensing) and radar images (active remote sensing) are delayed in nature as only earlier data for the images can be accessed [2]. This method cannot

be used to anticipate changes in HABs. Therefore, machine learning is a technique that has been developed to track HABs all over the world. Machine learning method is an effective technique that been used especially in improving chlorophyll-a retrievals in the global ocean and inland water environment [3]. Machine learning can be used to detect phycocyanin pigment as an indicator to identify HABs in the water area. This method created a prediction model based on the testing and training variables' data. The analysis of seawater collection data using machine learning techniques will significantly increase the predictive power of HABs and lower the cost of marine environmental monitoring [16]. Machine learning models such as creating an artificial neural network (ANN) and a support vector machine (SVM) were used to analyse cyanobacteria cell density. The importance of HABs forecasting has grown among stakeholders and in environmental sciences. For example, Wang et al (2023) has employed a machine learning method (LightGBM) for predicting the areal coverage of cyanobacterial blooms at an hourly scale [15]. The flowchart for his research has shown on Figure 1.

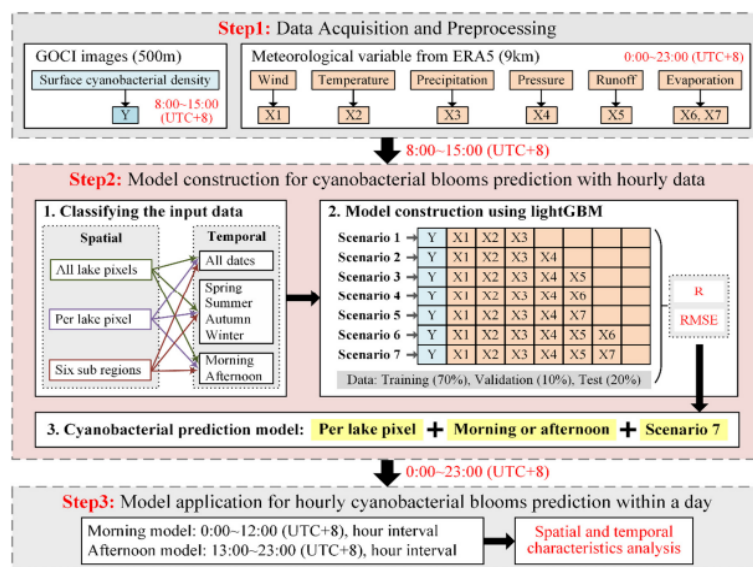


Figure 1. Flowchart for using machine learning to predict area coverage of cyanobacterial blooms on an hourly scale (LightGBM) [15]

Wen et al. (2022) has employed four models of machine learning to make the prediction of HABs location [15]. Figure 2 has shown the four components that make up the structure of the local spatiotemporal HABs forecasted model (STHFM), including the HABs MEFs extraction model (HMEN), the Spatio-temporal feature clustering model (STFCM), the auto-regressive integrated moving average model (ARIMA), and the HABs prediction network model (HPNM).

Another example is the development of an early-warning system for the prediction of two different types of HABs particularly fish kills and toxic bloom occurrences in Bolinao-Anda, Philippines, using data that obtained from in situ sensors. This system was developed using the random forest, which is a machine learning algorithm. The fish kill and toxic bloom models accurately predicted the occurrence of fish kills and shellfish bans by 96.1% and 97.8%, respectively, demonstrating the effectiveness of the random forest method [18]. Even though machine learning has more ability that surpassed the previous technology like remote sensing method, the use of machine learning is still low compared to remote sensing method. This study will attempt to achieve the following objectives, which are to analyse the number of existing publications on machine learning of phycocyanin from the period of 2002-2022, to identify the frequency of using machine learning methods to detect phycocyanin pigment in harmful algae blooms and to investigate the trend of using machine learning to detect phycocyanin in harmful algae blooms. This study highlighted the gap between this method and the improvement that can be made for future studies. This study will provide a current overview of the research and publishing landscape on the topic.

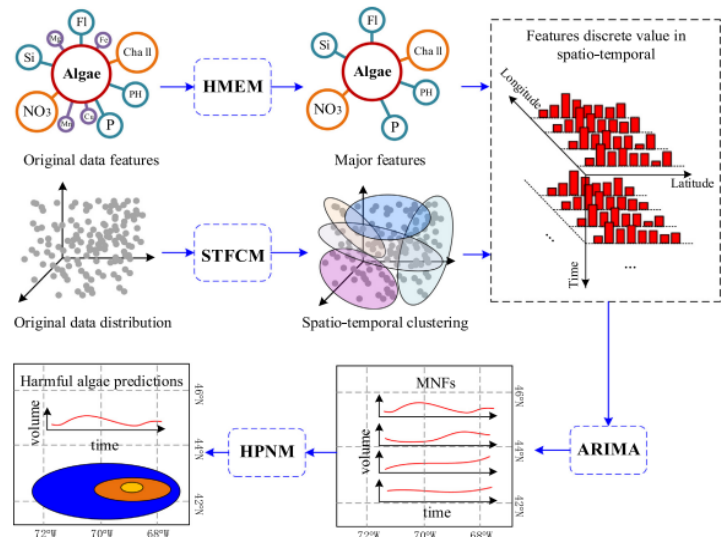


Figure 2. The framework of HABs detection using four models of machine learning [16]

2.0 METHODOLOGY

This research used bibliometric analysis with the application of VOS Viewer tool. Bibliometric techniques were utilised in this work to discover and analysed scientific literature on the application of machine learning in detecting phycocyanin pigment that present in HABs. Bibliometric analysis statistically analyses academic literature and explains publication trends, document types, subject areas, main authors, and publication trends in the HABs study domain. Figure 3 describes the research design of this study.

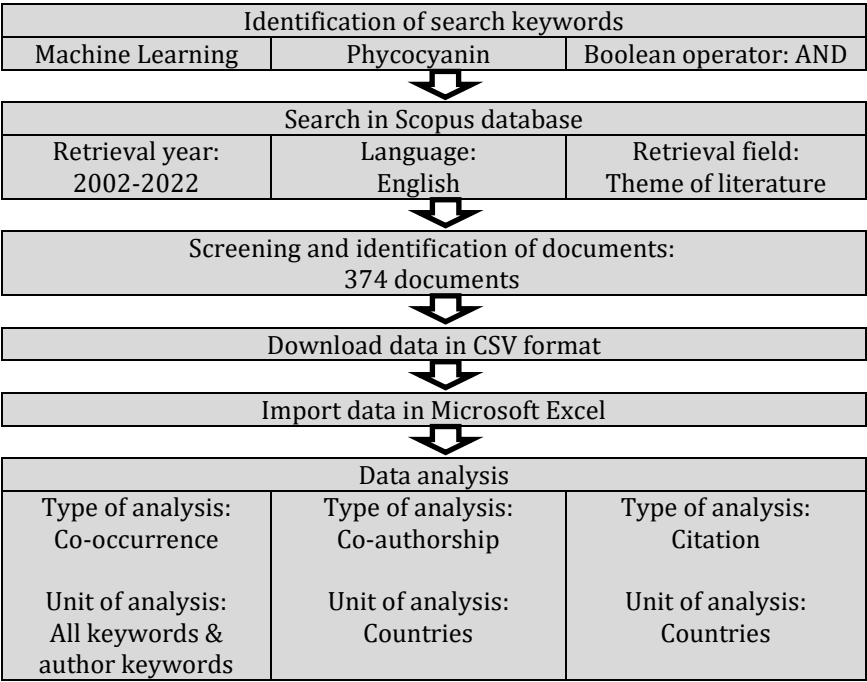


Figure 3. Research design

In this study, Scopus database was chosen to extract the articles in CSV format on the detection of phycocyanin in HABs using machine learning. The data was selected for twenty years range start from 2002 to 2022. Based on the initial results, 386 documents were collected, with the earliest publication was made in 1997, while the most current was made in this year (2022). The following search strategy was used for data collection from the Scopus database: (TITLE (“machine learning” AND “phycocyanin”) AND (EXCLUDE (PUBYEAR,2023) OR EXCLUDE (PUBYEAR,1997))). The number of documents returned drops to 374 if the search phrases are changed to restrict the results to only journal papers published between 2002 and 2022. In a nutshell, the year, source, author, association, nation or territory, subject, and document type were

taken into consideration when choosing these results. Then, the data were analysed according to type of analysis and unit of analysis presented in VOS Viewer.

3.0 ANALYSIS

3.1 Document Type and Publishing Language

In total 374 literature items were retrieved according to the search strategy. Table 1 showed the document types that were lead the publications were research articles (287, 76.8%) followed by reviews (60, 16%), and then conference papers (13, 3.5%). Meanwhile the publication language that was leading by China by 134 publications, followed by United States of America by 96 publications, United Kingdom by 28 publications, South Korea by 25 publications, Australia by 25 publications and India by 20 publications. Figure 4 has shown the trend of the publications from year 2002 to 2022.

Table 1. Publications type and percentage

Document type	Percentage	Number
Articles	76.8%	287
Review	16.0%	60
Conference papers	3.5%	13
Book	1.3%	5
Book chapters	1.3%	5
Note	0.8%	3
Editorial	0.3%	1
TOTAL	100%	374

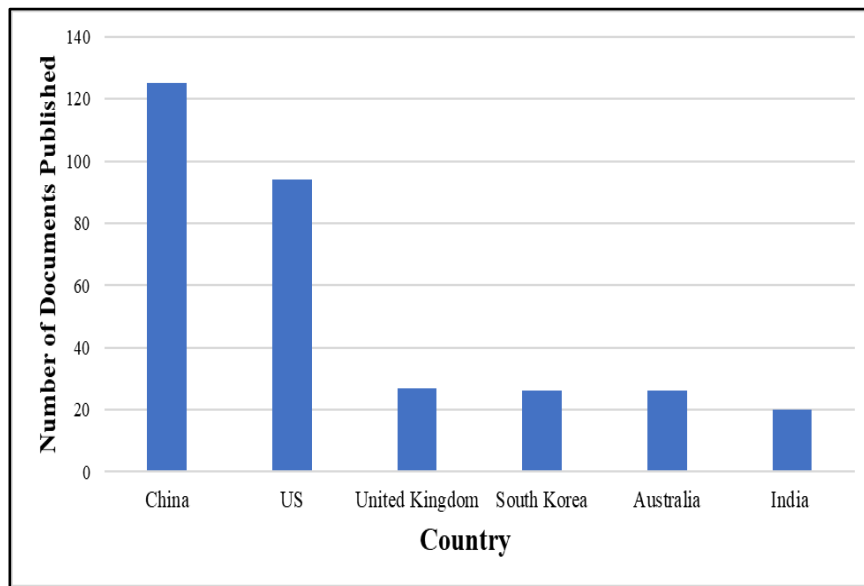


Figure 4. Number of publications based on language

3.2 Most Cited Publications

Most active and cited publications between 2002 and 2022 have been shown on Table 2 which is the article/review named “*An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities*” which published in 2015 is the most cited journal article, which is 265. “*A survey on river water quality modelling using artificial intelligence models: 2000–2020*” is a review that published in 2020, is in the second place which is 230, accounted for the highest number of reviews. The third place is a review named A review of high value-added molecules production by microalgae considering the classification that published in 2020 with the cited number of 156.

Table 2. Most cited publications

No.	Title	Authors	Year	Source	Cited by
1.	An Introduction to the NASA Hyperspectral InfraRed Image (HyspIRI) Mission and Preparatory Activities	Lee et al.	2015	Remote Sensing of Environment	265
2.	A Survey on River Water Quality Modelling Using Artificial Intelligence Models:2000-2020	Tiyasha et al.	2020	Journal of Hydrology	230
3.	Survey Of Hyperspectral Earth Observation Applications from Space in The Sentinel-2 Contest	Transon et al.	2018	Remote Sensing	156
4.	A Review of High Value-Added Molecules Production by Microalgae in Light of The Classification	Levasseur et al.	2020	Biotechnology Advances	154
5.	Non-adiabatic Excited-State Molecular Dynamics: Theory and Applications for Modelling Photophysics in Extended Molecular Materials	Nelson et al.	2020	Chemical Reviews	146

3.3 Co-Occurrence Network Analysis

Based on the Figure 5, frequency statistics of keyword co-occurrences by VOS viewer, and the most frequently used keywords for the area of study between 2002 and 2022 including machine learning, remote sensing, harmful algae blooms, cyanobacteria, neural networks, deep learning, machine learning techniques, satellite imagery, sentinel, lake water, chlorophyll-a, image resolution, alga blooms, lake analysis and spectral analysis. According to Tamala et al. (2022), the VOS Viewer software was used to examine each keyword, and the co-occurrences of the keyword with other keywords as well as the links, total link strengths, and co-occurrences of the keyword were computed [13]. Co-occurrence describes the connections between the keywords. Links refer to the co-occurrence of the keywords with another, and the overall link strength relates to all the references that have been noted between the keyword and the others. Additionally, the occurrences show how many articles contained the keyword. Table 3 showed the list of keywords related to machine learning.

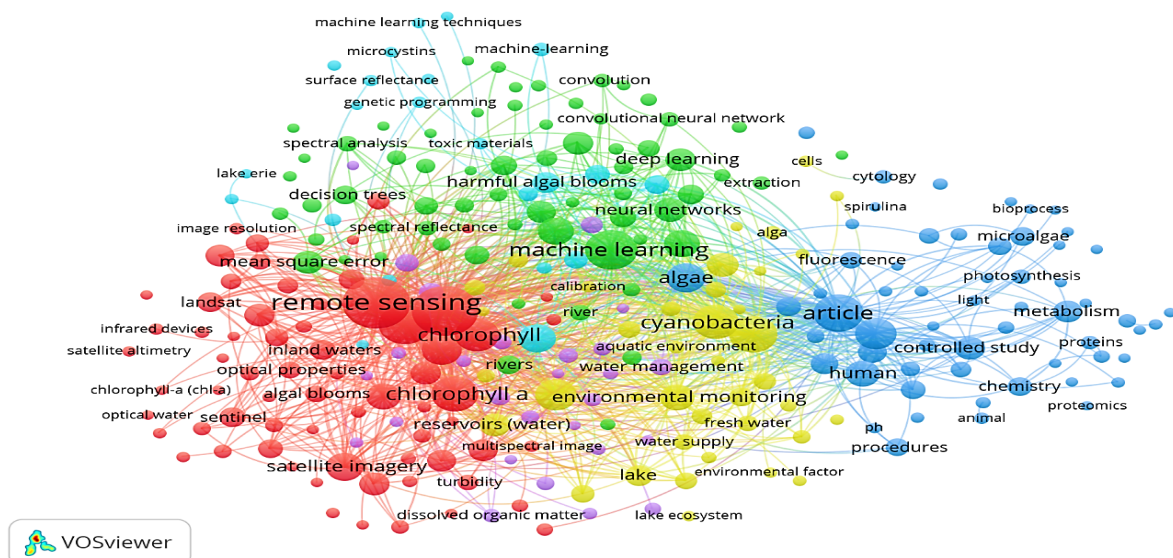


Figure 5. Co-occurrence all keywords analysis

Table 3. Keywords related to machine learning

Keywords	Links	Total Link Strength (TLS)	Occurrences
Machine Learning	242	1026	79
Support Vector Machine (SVM)	96	167	12
Decisions Tree	112	220	18
Random Forests	74	90	9
Artificial Neural Network (ANN)	190	490	37
TOTAL			155

The occurrences for keyword “Machine learning” has shown the number of 79 which means the number of the articles that contained the keyword “Machine learning” is 79 publications. The researcher also takes the keywords that related to “Machine learning” keyword for example support vector machines, random forests, artificial neural network (ANN), and decisions tree keywords. The total co-occurrence of these keywords including the “Machine learning” keyword is 155. It means 155 publications or 41.4% have been used machine learning method to detect phycocyanin pigment in HABs out of 374 publications.

3.4 Trend of Publications

Figure below has shown the graph for publications that mention the use of machine learning to detect the phycocyanin pigment in HABs. The publications have been increased drastically after 2018. The graph showed that there has been a steady growth in research interest in machine learning for phycocyanin identification in HABs, and this trend is expected to continue. The advanced development of the machine learning played the important role in the increase of the number of the publications in this field. Based on the graph on Figure 6, the number of publications increased drastically started from 2019 until 2022. The maximum number of publications were in year 2022 which is 134 publications. It increased almost 33% compared to previous year number of publications in 2020 in which the total number of publications is 90.

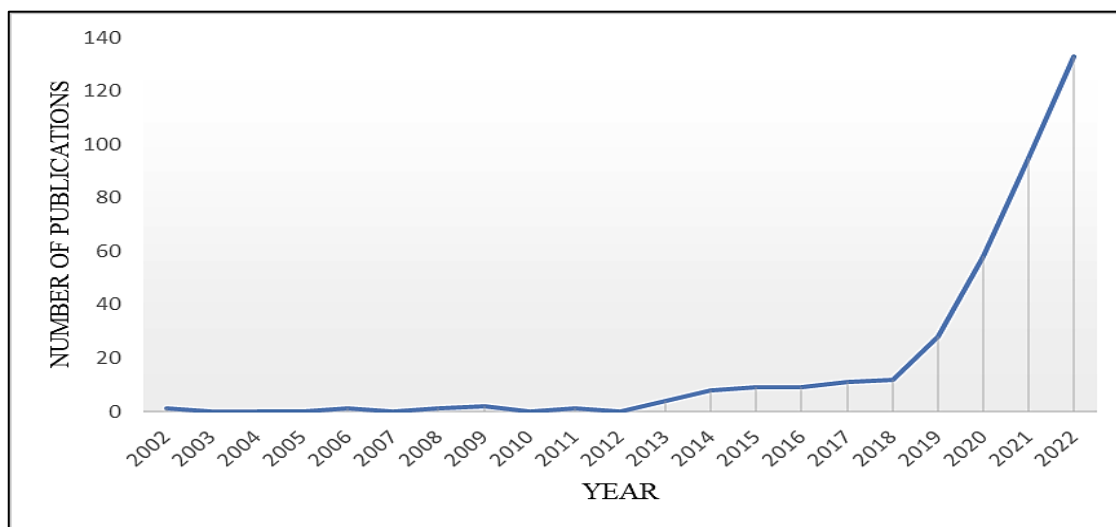


Figure 6. Trend of publications from year 2002 until 2022

4.0 CONCLUSION

The bibliometric analysis has shown that the machine learning method in HABs detection becomes more common especially in the last few years specifically started from 2012. The researchers become more interested to conduct the study in this field due to its new technology and capability that surpassed the previous detection method. This research solely uses one database which is Scopus database and the software used in this research is VOS Viewer. The proposed method enables to analyse the publications and trends in HABs and machine learning study area. Therefore, other scientific database such as web of science (WoS) and different bibliometric analysis tools (CiteSpace or BibExcel) are suggested to use while conducting future studies. It can show the new perspective of the study that can lead to another improvement of using machine learning in Phycocyanin pigment detection in HABs.

5.0 CONFLICT OF INTEREST

The authors declare no conflicts of interest.

6.0 AUTHORS CONTRIBUTION

Suhaini, M. H. N. (Writing - original draft; Writing - review & editing)

Rosly, N. A. (Conceptualisation; Methodology; Project administration; Supervision)

Tajudin, S. T. (Methodology; Validation)

Ahmad, K. (Validation; Formal analysis)

7.0 ACKNOWLEDGEMENT

The authors fully acknowledged Ministry of Higher Education (MOHE) and National Defence University of Malaysia (NDUM) for their support in making this research feasible.

List of Reference

- [1] Alcantara, E., Coimbra, K., Ogashawara, I., Rodrigues, T., Mantovani, J., Rotta, L. H., ... & Cunha, D. G. F. (2021). A satellite-based investigation into the algae bloom variability in large water supply urban reservoirs during COVID-19 lockdown. *Remote Sensing Applications: Society and Environment*, 23, 100555.
- [2] Cao, H., Han, L., & Li, L. (2022). A deep learning method for cyanobacterial harmful algae blooms prediction in Taihu Lake, China. *Harmful Algae*, 113, 102189.
- [3] Chusnah, W. N., & Chu, H. J. (2022). Estimating chlorophyll-a concentrations in tropical reservoirs from band-ratio machine learning models. *Remote Sensing Applications: Society and Environment*, 25, 100678.
- [4] Liow, G. R., Lau, W. L. S., Law, K., Hii, K. S., Noor, N. M., Leaw, C. P., & Lim, P. T. (2019). Phytoplankton community changes in Kuantan Port (Malaysia), with emphasis on the paralytic-shellfish toxin-producing dinoflagellate *Alexandrium tamiyavanichii*. *Regional Studies in Marine Science*, 26, 100504.
- [5] Mahmudi, M., Serihollo, L. G., Herawati, E. Y., Lusiana, E. D., & Buwono, N. R. (2020). A count model approach on the occurrences of harmful algal blooms (HABs) in Ambon Bay. *Egyptian Journal of Aquatic Research*, 46(4), 347-353.
- [6] Ogashawara, I. (2020). Determination of phycocyanin from space—a bibliometric analysis. *Remote Sensing*, 12(3), 567.
- [7] Park, Y., Lee, H. K., Shin, J. K., Chon, K., Kim, S., Cho, K. H., ... & Baek, S. S. (2021). A machine learning approach for early warning of cyanobacterial bloom outbreaks in a freshwater reservoir. *Journal of Environmental Management*, 288, 112415.
- [8] Renugadevi, K., Nachiyar, C. V., Sowmiya, P., & Sunkar, S. (2018). Antioxidant activity of phycocyanin pigment extracted from marine filamentous cyanobacteria *Geitlerinema* sp TRV57. *Biocatalysis and agricultural biotechnology*, 16, 237-242.
- [9] Rousso, B. Z., Bertone, E., Stewart, R. A., Hobson, P., & Hamilton, D. P. (2022). Automation of species-specific cyanobacteria phycocyanin fluorescence compensation using machine learning classification. *Ecological Informatics*, 69, 101669.
- [10] Shan, S., Chen, Z., Koh, K. Y., Cui, F., & Chen, J. P. (2022). Development and application of lanthanum peroxide loaded sepiolite nanocomposites for simultaneous removal of phosphate and inhibition of cyanobacteria growth. *Journal of Colloid and Interface Science*, 624, 691-703.
- [11] Shao, S., Zhao, L., Li, P., Su, H., Chen, X., & Zhang, Y. (2022). A study of the protein-protein interactions in the phycocyanin monomer from *Synechocystis* sp. PCC 6803 using a bacterial two-hybrid system. *Engineering Microbiology*, 2(2), 100019.
- [12] Suteja, Y., Dirgayusa, I. G. N. P., Cordova, M. R., Rachman, A., Rintaka, W. E., Takarina, N. D., ... & Purwiyanto, A. I. S. (2021). Identification of potentially harmful microalgal species and eutrophication status update in Benoa Bay, Bali, Indonesia. *Ocean & Coastal Management*, 210, 105698.
- [13] Tamala, J. K., Maramag, E. I., Simeon, K. A., & Ignacio, J. J. (2022). A bibliometric analysis of sustainable oil and gas production research using VOSviewer. *Cleaner Engineering and Technology*, 7, 100437.

- [14] Teng, S. T., Lim, H. C., Lim, P. T., Dao, V. H., Bates, S. S., & Leaw, C. P. (2014). Pseudo-nitzschia kodamae sp. nov.(Bacillariophyceae), a toxigenic species from the Strait of Malacca, Malaysia. *Harmful Algae*, 34, 17-28.
- [15] Wang, S., Zhang, X., Wang, C., & Chen, N. (2023). Temporal continuous monitoring of cyanobacterial blooms in Lake Taihu at an hourly scale using machine learning. *Science of the Total Environment*, 857, 159480.
- [16] Wen, J., Yang, J., Li, Y., & Gao, L. (2022). Harmful algal bloom warning based on machine learning in maritime site monitoring. *Knowledge-Based Systems*, 245, 108569.
- [17] Yñiguez, A. T., Lim, P. T., Leaw, C. P., Jipanin, S. J., Iwataki, M., Benico, G., & Azanza, R. V. (2021). Over 30 years of HABs in the Philippines and Malaysia: What have we learned?. *Harmful Algae*, 102, 101776.
- [18] Yñiguez, A. T., & Ottong, Z. J. (2020). Predicting fish kills and toxic blooms in an intensive mariculture site in the Philippines using a machine learning model. *Science of The Total Environment*, 707, 136173.