Health Biotechnology and Biopharma (2023), 7(3): 30-44

Original Research Article

Bibliometric analysis: The future of thalassemia disease by utilizing artificial intelligence technology

Erliyan Redy Susanto [1, 2] Admi Syarif [3,*] Warsito Warsito [4], Khairun Nisa Berawi [5], Hidayat Hidayat [6]

[1]Doctoral Program, Faculty of Mathematics and Natural Sciences, University of Lampung, Indonesia; [2]Information System Department, Engineering and Computer Science Faculty, Universitas Teknokat Indonesia;[3]Computer Science Department, Mathematics and Natural Sciences Faculty, Lampung University, Indonesia; [4]Physics Department, Mathematics and Natural Sciences Faculty, Lampung University; [5]Medical Faculty, Lampung University, Indonesia; [6]Clinical Pathology Laboratory of Abdul Moeloek Hospital, Bandar Lampung, Indonesia

***Corresponding author:** Admi Syarif, Computer Science Department, Mathematics and Natural Sciences Faculty, Lampung University, Indonesia. E- mail: admi.syarif@fmipa.unila.ac.id*

DOI: 10.22034/HBB.2023.22

Received: July 23, 2023; **Accepted:** October 29, 2023

ABSTRACT

The use of artificial intelligence to solve surfaces in thalassemia disease has been the focus of popular research in the last decade. However, there are no scientific reports that provide a systematic overview of this scientific field. We used bibliometric approaches to identify and analyze academic literature on the use of artificial intelligence in thalassemia disease issues and explore emerging research trends, joint writer networks, institutions, countries, and journals. The results of this analysis also obtained information that the issues of sensitivity and specificity, diagnostic accuracy in computer science have the potential to acquire novelty.

***Keywords*:** Artificial intelligence, bibliometric analysis, scopus, thalassemia

INTRODUCTION

Genetic problems can cause blood disorders such as Leukemia and Thalassemia [1]. Thalassemia is a group of inherited blood disorders that affect the production of hemoglobin, the protein in red blood cells that carries oxygen [2]. Thalassemia can be mild, moderate, or severe, and can cause anemia, heart problems, and other health problems [3]. Iron deficiency and other severe anemias can be brought on by Thalassemia. This

may result in a reduction in the volume of red blood cells [4]. This study focuses on exploring research trends related to Artificial Intelligence (AI) in the treatment of Thalassemia. AI is a fast-evolving field that has the potential to revolutionize diagnosis [5] and prevention of Thalassemia [6]. Systems that leverage AI can be used to analyze large amounts of data [7], identify patterns [8], and make predictions. This can help doctors diagnose Thalassemia earlier and more accurately and develop more effective treatments.

Bibliometric analysis has become an important tool in understanding research trends and scientific developments in various fields [9,10]. In this study, bibliometric analysis was used to explore and analyze research related to a particular topic, namely AI in the context of Thalassemia. Bibliometric analysis was performed using the Scopus database. The keywords "Artificial Intelligence" and "Thalassemia" are used to find relevant articles. Search results are limited to articles published in English until 2023. PubMed is the best database in medical science [11]. However, the analytical viewpoint of the PubMed database is more difficult than that of the Scopus database [9]. Artificial intelligence is a branch of computer science and this field is quantitatively more numerous in the Scopus database even though compared to the Web of Science database [10]. Therefore, we use the Scopus database in this study. Bibliometric analysis reveals that the number of published articles on AI in the diagnosis and prevention of Thalassemia has steadily increased over the last decade. Most of these articles are published in high-impact journals. The most active research countries in this field are Thailand, China and Italy. The bibliometric analysis also identified a number of research gaps in this area [11,12]. For example, therefore it is very necessary to have a Decision Support System for the detection of Thalassemia at this time [13]. Further research is also needed to use AI to develop new treatments for Thalassemia such as gene therapy [14,16].

The research gap that can be identified through this bibliometric analysis is the need to increase diagnostic accuracy in the treatment of Thalassemia using AI. In this regard, research can be focused on developing more sophisticated algorithms to diagnose diseases with higher accuracy. In addition, attention should also be paid to increasing the sensitivity and specificity of

using AI to support the diagnosis and treatment of Thalassemia. In addition, this bibliometric analysis also reveals that machine learning and deep learning are the methods most widely discussed in research related to AI in Thalassemia. Therefore, the novelty of this study lies in the development of more advanced and complex learning techniques in the application of AI in the treatment of Thalassemia.

By understanding the gaps and opportunities for updating this research, it is hoped that further research can fill existing knowledge gaps and achieve significant progress in the use of AI for the treatment of Thalassemia. Thus, this research is expected to provide a valuable contribution to the development of science and progress in efforts to provide better care for patients suffering from this disease. The contribution made by this study is to find the potential need to improve diagnostic accuracy in the treatment of Thalassemia using AI so that high sensitivity and specificity is needed in the use of AI to support the diagnosis and treatment of Thalassemia.

MATERIALS AND METHODS

This study followed the stages according to the bibliometric analysis. The right keywords are the key to success in analysis according to bibliometric studies [17]. The article data analyzed in this study were obtained from the Scopus database with the keywords "Thalassemia" and "Artificial Intelligence". This research was carried out in accordance with predetermined systematic steps. The stages of bibliometric analysis are presented in Figure 1.

Data Extraction

Data extraction refers to the process of systematically collecting relevant information from various sources, typically academic publications, and converting it into a structured format that can be analyzed [18]. The purpose of data extraction is to gather the raw information needed for analysis. This stage requires careful attention to detail to ensure accurate and consistent capture of data elements. Data Extraction in bibliometric analysis is the process of gathering and extracting relevant information from the raw data sources used in the analysis [19,20]. Data extraction is the first step in the process of bibliometric analysis and is important to obtain datasets that are suitable for research purposes. By performing good data

extraction, researchers can obtain structured, complete, and relevant datasets for bibliometric analysis [21,22]. Efficient data extraction allows researchers to proceed to the analysis phase with information that has been collected in a systematic manner. Combined with good data preparation, proper data extraction can provide a solid foundation for gaining valuable insights and findings through bibliometric analysis.

Data preparation

Data preparation involves processing and transforming the raw data obtained from the extraction stage into a format suitable for analysis [23]. This includes cleaning the data to correct errors, standardizing inconsistent entries, removing duplicates, and structuring the data in a way that enables meaningful analysis. Data preparation aims to ensure that the data is reliable, consistent, and ready for statistical or computational analysis. Data preparation in bibliometric analysis is an important step to prepare raw data before analysis. The purpose of compiling data is to simplify and organize data so that it can be processed and analyzed more effectively and efficiently [24]. Data preparation was carried out by collecting bibliographic data to be used in the analysis in the BibTeX (.bib) file format [25]. The need to ensure bibliographic data includes information such as article title, author's name, year of publication, journal of publication, and keywords and other important information. If the bibliographic data does not contain complete information, it is necessary to enrich the data.

Analysis configuration

Through the proper configuration of bibliometric analysis, researchers can obtain more focused and relevant analysis results. The selection of methods, units of analysis, calculation techniques, and other parameters must be adjusted to the research objectives and the characteristics of the data used [26,27]. The proper configuration will help reveal trends, patterns and valuable information in the scientific literature related to the research topic. In this study, the type of analysis used was co-occurrence analysis, with the unit of analysis involving all relevant keywords, using the "Full count" calculation method with a threshold of 4. The number of keywords selected was 27.

Visualization

Using graphical visualizations makes it easy to see and understand complex collaboration networks. Visualization can

make it easier to identify related clusters and see research themes emerging from bibliometric data [17,28]. Visualization can also make it easier to identify related clusters and see research themes emerging from bibliometric data.

Interpretation

It is important to interpret the visualization results displayed by VOSviewer by identifying the clusters that are formed and observing the pattern of relationships between documents or keywords in these clusters [29]. Next, pay attention to the keywords that appear most often or are most typical in the analysis results.

RESULTS

Articles that have been obtained from the Scopus database are then identified based on the field of study. This identification is very important because it makes it easier to find out the many fields of study that discuss research topics with the keywords "Artificial Intelligence" and "Thalassemia". Research trends regarding Thalassemia can be seen based on the number of articles published each year. The number of papers we have collected in the last two decades is presented in Figure 2.

In general, the number of papers related to these keywords continues to increase every year. Even though it experienced a decline, in the following years the number continued to increase. It is possible that in the next few years, this issue will still be a hot topic to discuss.

The number of papers by country successively, namely China (4), Italy (4), Thailand (4), undefined (3), Iran (3), United States (2), Singapore (2), Libyan Arab Jamahiriya (2), Indonesia (2), Viet Nam (1), United Kingdom (1), Sri Lanka (1), Spain (1), Qatar (1), Portugal (1), Lebanese (1), Jordan (1), India (1), greece (1), Germany (1), france (1), Cote d'Ivoire (1), Canada (1), and Brazil (1). China, Italy and Thailand are the countries with the highest number of studies related to these keywords. This shows that these countries have a high interest in developing AI technology for the diagnosis and treatment of Thalassemia. The results of this analysis can also be used to identify countries that can become partners in research on artificial intelligence for Thalassemia.

We also get four authors who write the most according to these keywords. Agwil, RO from the Institut National de la Recherche Scientifique, Quebec, Canada, together with Shrivastava, DP from Higher

Colleges of Technology, Abu Dhabi, United Arab Emirates, have published 2 papers in 2010. Paokanta, P. and Srichairatanakool, S. from Chiang Mai University, Chaing Mai, Thailand, has published 1 paper in 2013 and 1 paper in 2015. This shows that there are not many paper authors who consistently discuss this topic which are published in Scopus indexed publishers.

Based on the field of study of the article with these keywords, respectively Computer Science (17), Medicine (14), Mathematics (9), Biochemistry, Genetics and Molecular Biology (7), Engineering (6), Physics and Astronomy (3), Pharmacology, Toxicology and Pharmaceuticals (1), Nursing (1), Multidisciplinary (1), Materials Science (1), Health Professions (1), Decision Sciences (1), and Chemistry (1). The field of study of Computer Science is the most based on these keywords. Then successively followed by Medicine, Mathematics and so on. This shows that there are many authors with a background in Computer Science who are interested in discussing research topics with these keywords.

VosViewer is a tool that can be used to discover research novelties using bibliometric analysis techniques. We use co-occurrence in this bibliometric analysis. Co-occurrence refers to the simultaneous occurrence of two or more keywords or terms in a particular document or collection of documents. This concept is used to identify relationships and patterns of presence of keywords or terms in scientific literature. In this bibliometric analysis, co-occurrence is applied to various elements, such as keywords, words in the title of the article, or terms in the abstract. In this context, co-occurrence can provide insight into topics or subtopics that frequently appear together in the scientific literature. Visualization of the relationship between keywords processed using the VosViewer application is presented in Figure 3.

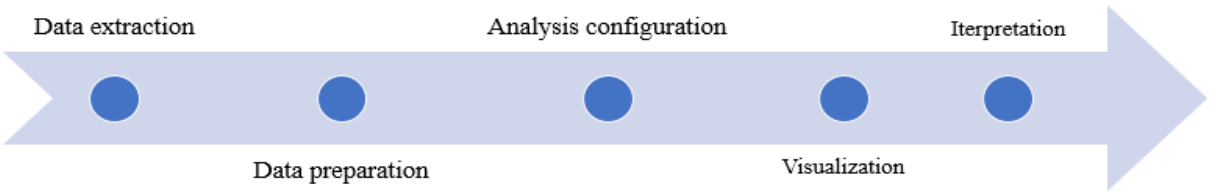

Figure 1. Stage of bibliometric analysis

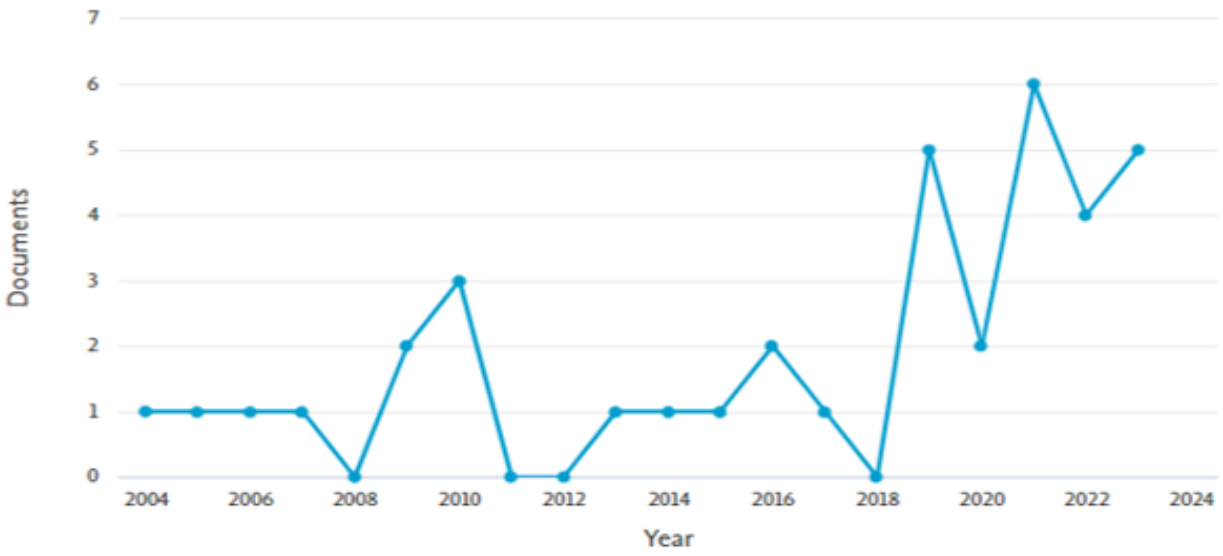

Figure 2. Trend of scopus publication

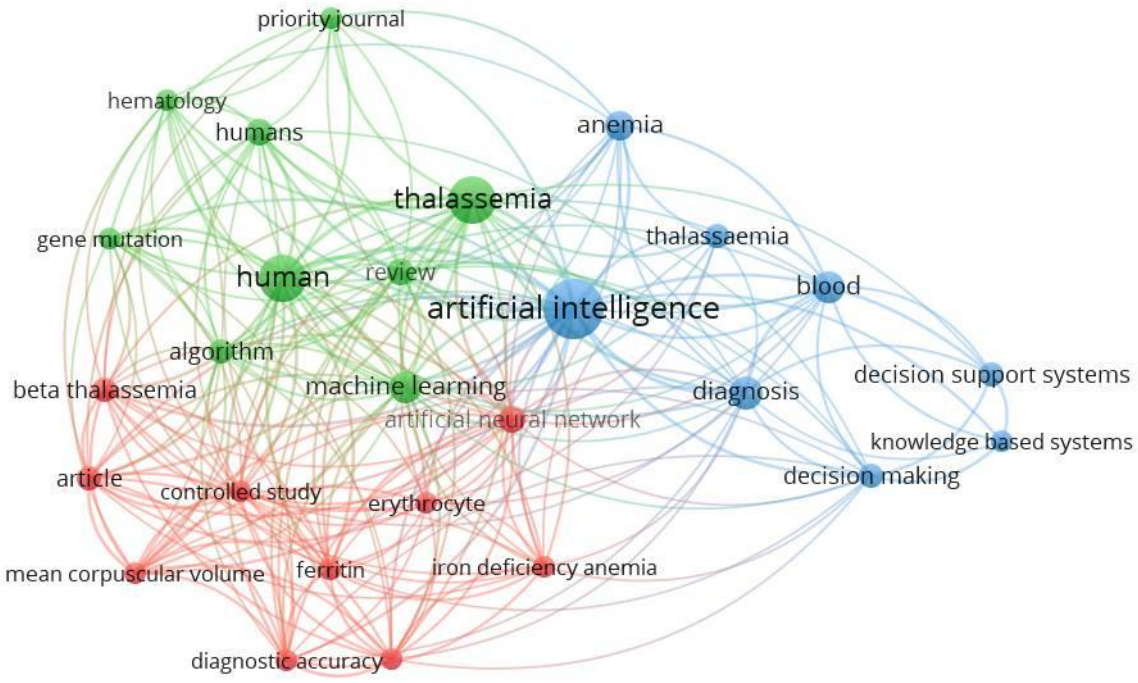

Figure 3. Network visualization

Based on the network visualization shown in Figure 3 with the keywords "Thalassemia" and "Artificial Intelligence", here we can see a topic that we can use as a reference in obtaining new opportunities, namely "diagnostic accuracy". This topic is worthy of consideration as a research topic because it is the furthest node from the two main keywords [30,31]. Meanwhile, to see trends in related research topics using the overlay visualization presented in the Figure 4.

The goal of overlay visualization is to help researchers visually understand the complex relationships between different elements in the scientific literature. By overlaying bibliometric data visualization, we can identify patterns, trends or relationships that might not be obvious. Meanwhile, based on the overlay visualization, the two topics were hot issues discussed recently. Meanwhile, based on the density visualization presented in Figure 5.

The density visualization shows the number of article nodes for each topic. The more articles there are, the thicker and clearer the knot will be. On the other hand, nodes will fade as the number of articles decreases.

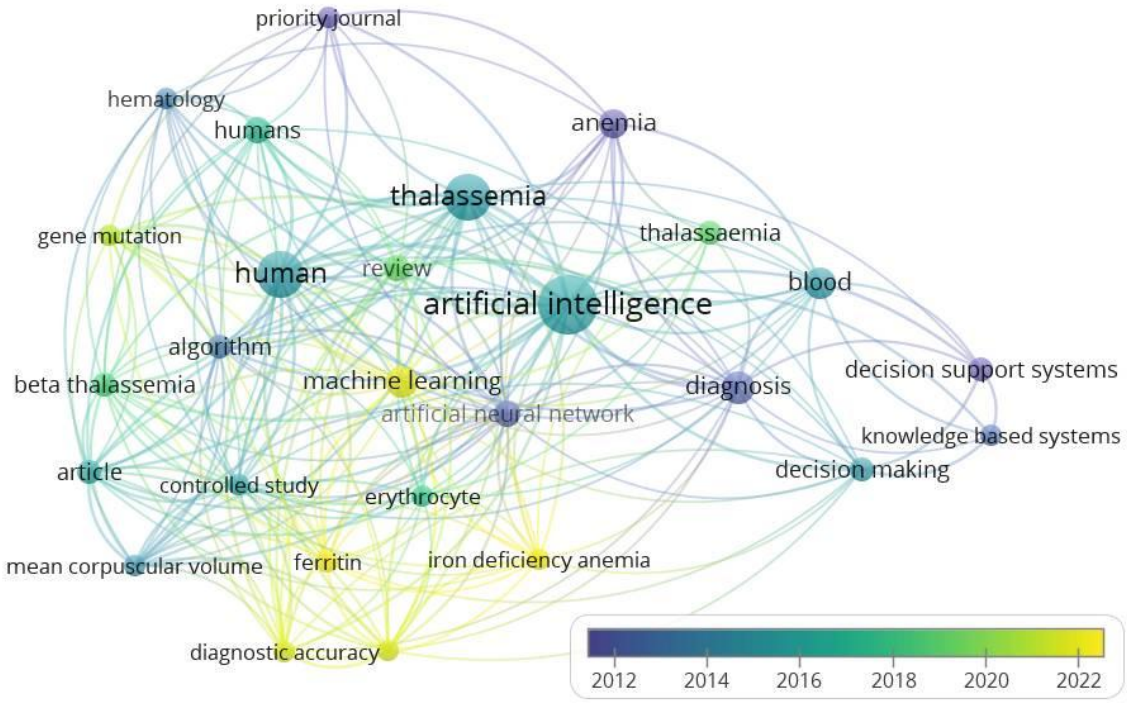

Figure 4. Overlay visualization

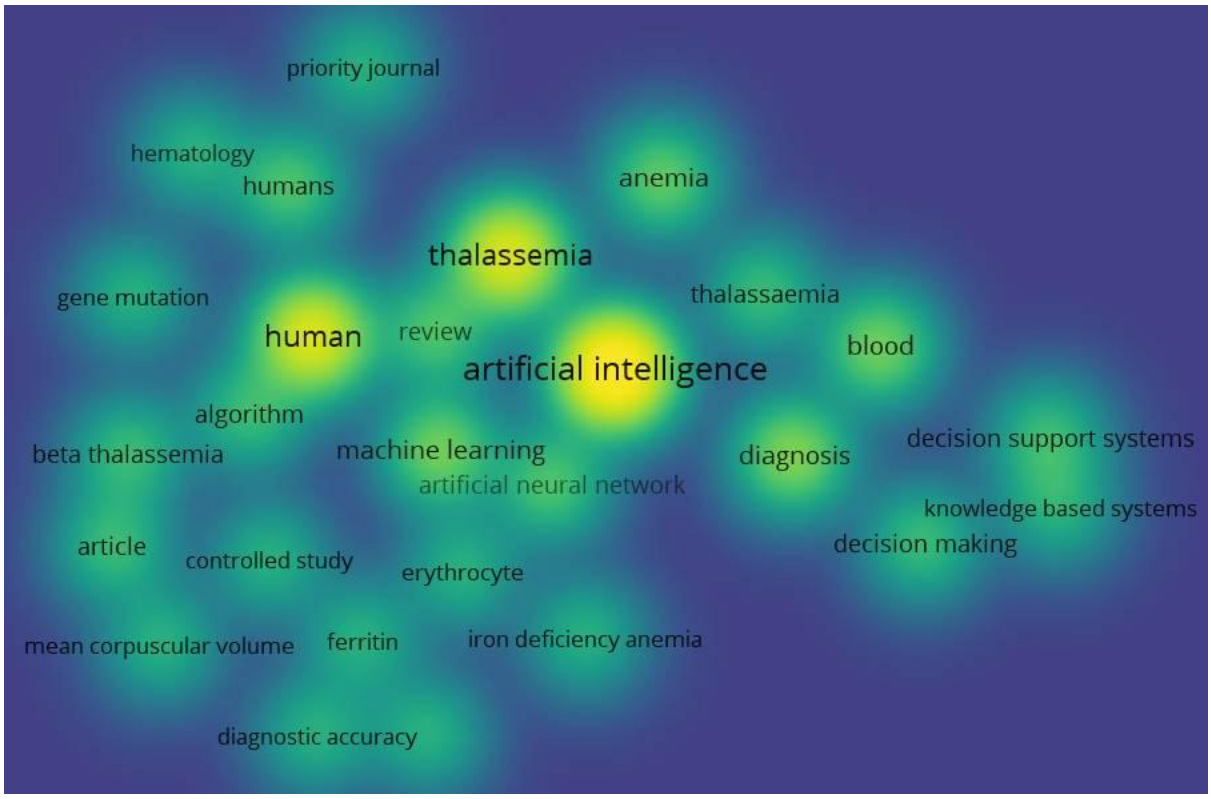

Figure 5. Density visualization

DISCUSSION

Figures 3, 4 and 5 show that the research topic of diagnostic accuracy is the furthest node, is a topic that has begun to be discussed recently, and is still rarely discussed. These findings indicate that there is an unexplored space in the use of AI to increase the diagnosis of Thalassemia. This suggests that deeper and broader research on this topic can make a significant contribution to the fields of Thalassemia and AI. With this new potential, further research on diagnostic accuracy can provide new insights and innovative methods to improve the accuracy and effectiveness of Thalassemia diagnosis. This kind of research can help in improving early detection, more accurate classification, as well as the development of a personalized approach to the treatment

of Thalassemia. However, it should be noted that this novelty still has potential and requires further research and development to fully explore its possibilities. Research involving practical implementation and validation of a larger and diverse population of Thalassemia patient data will be an important step to test the effectiveness and novelty of the AI approach in increasing diagnostic accuracy of Thalassemia problems.

The results of this bibliometric study also open up opportunities for collaboration with international authors. International authors may have more in-depth knowledge of the latest developments in Thalassemia research and may have access to different resources. Collaboration with them could help expand understanding of this disease. Thalassemia research often requires access to patient data and blood or tissue samples. Collaboration with international authors can help gain access to larger data collections and samples, which may be difficult to achieve when working independently. Collaboration with international authors who have different backgrounds or expertise in related fields can open up opportunities for more comprehensive multidisciplinary research. This can help explore new aspects of Thalassemia and create a more holistic solution.

CONCLUSION

This research summarizes the latest advances in Artificial Intelligence technology in Thalassemia disease research. In this study, we found that the emerging trend was “diagnostic accuracy”. Future research on the problem of Thalassemia based on Artificial Intelligence can consider the results of this bibliometric analysis to obtain more possible novelties. The most productive countries in this study were Italy, China and Thailand. There are four authors of the most papers in Scopus-indexed publishers, namely Agwil, RO from the Institut National de la Recherche Scientifique, Quebec, Canada. Shrivastava, DP from Higher Colleges of Technology, Abu Dhabi, United Arab Emirates. Paokanta, P. and Srichairatanakool, S. from Chiang Mai University, Chaing Mai, Thailand. Our findings will provide valuable clues relevant to future research directions on Thalassemia using Artificial Intelligence technologies.

REFERENCES

[1]. Alhoda B, Dehghi K, Jalili A, Zahra Y, Dehghi K, Sadeghian MH, et al. Assessing the frequency of t (12 ; 21) (p13 ; q22) in patients with acute lymphoblastic leukemia in northeastern of Iran. *Health Biotech Biopharm* 2018; 2: 48–62.

[2]. Sa'Id AA, Rustam Z, Novkaniza F, Setiawan QS, Maulidina F, Wibowo VVP. Twin support vector machines for thalassemia classification. *ICT* 2021: 160–64.

[3]. Suganya Devi K, Arutperumjothi G, Srinivasan P. Diagnosis evaluation and interpretation of qualitative abnormalities in peripheral blood smear images—A Review. *Heal. Informatics A Comput. Perspect. Health*; 2021, 341–65.

[4]. Ayatollahi J, Desin F, Sharifyazdi M, Mousavi SA, Akhondimeybodi Z, Heidari F, *et al*. Investigation of factors affecting adult erythrocyt sedimentation rate. *Health Biotech Biopharm*, 2022; 5: 45–54.

[5]. Singh K, Singh P, Angurala M. Future directions in healthcare research. *Comput. Heal. Informatics Biomed. Appl;* 2023, 287–313.

[6]. Mielberg EL. Medllecta. *HPM.* 2021.

[7]. Ali AA, Hassan HS, Anwar EM, Khanna A. Hybrid technique for heart diseases diagnosis based on convolution neural network and long short-term memory. *Appl. Big Data Healthc.*; 2021, 261–80.

[8]. Kumar K, Thakur GSM. Advanced applications of neural networks and artificial intelligence: A Review. *Int J Inf Technol Comput Sci* 2012; 4: 57–68.

[9]. Zyoud SH, Fuchs-Hanusch D. A bibliometric-based survey on AHP and TOPSIS techniques. *ESA.* 2017; 78: 158–81.

[10]. Jovicic SM. Global trend of clinical biomarkers of health and disease during the period (1913–2021): systematic review and bibliometric analysis. *African J Urol* 2021; 27: 141.

[11]. Kokol P, Vošner HB. Discrepancies among Scopus, Web of Science, and PubMed coverage of funding information in medical journal articles. *J Med Libr Assoc* 2018;

106: 81–86.

[12]. Alviz-Meza A, Orozco-Agamez J, Quinayá DCP, Alviz-Amador A. Bibliometric analysis of fourth industrial revolution applied to material sciences based on Web of Science and Scopus databases from 2017 to 2021. *Chem Engineering* 2023; 7: 2.

[13]. Saputra DCE, Sunat K, Ratnaningsih T. A new artificial intelligence approach using extreme learning machine as the potentially effective model to predict and analyze the diagnosis of anemia. *Healthcare* 2023; 11: 697.

[14]. Ferih K, Elsayed B, Elshoeibi AM, Elsabagh AA, Elhadary M, Soliman A, *et al.* Applications of artificial intelligence in thalassemia: A comprehensive review. *Diagnostics*, 2023; 13: 1551.

[15]. Tran DC, Dang AL, Hoang TNL, Nguyen CT, Le TMP, Dinh TNM, *et al*. Prevalence of thalassemia in the Vietnamese population and building a clinical decision support system for prenatal screening for thalassemia. *Electron J Gen Med* 2023; 20: 501.

[16]. Day WJ, Chiriboga CA, Crawford TO, Darras BT, Benguerba K, Mendell JR. Thérapie de remplacement d'un gène, AVXS-101 dans l'amyotrophie spinale de type 1 (ASI-1) : mise à jour de l'étude pivot (STR1VE). *Rev Neurol*, 2019; 175: 97–98.

[17]. Huang Y-J, Cheng S, Yang F-Q, Chen C. Analysis and visualization of research on resilient cities and communities based on VOSviewer. *Int J Environ Res Public Health* 2022; 19: 7068.

[18]. He T, Wang D, Wu Z, Huang C, Xu X, Xu X, *et al.* A bibliometric analysis of research on (R)-ketamine from 2002 to 2021. *Neuropharmacol*, 2022; 218; 109207.

[19]. Tamala JK, Maramag EI, Simeon KA, Ignacio JJ. A bibliometric analysis of sustainable oil and gas production research using VOSviewer. *Clean Eng Technol* 2022; 7: 100437.

[20]. Li WH, Hadizadeh M, Yusof A, Naharudin MN. Analysis of research trends on elbow pain in overhead sports: A bibliometric study based on Web of Science database and

VOSviewer. *Healthcare* 2022; 10: 2242.

[21]. Cascella M, Perri F, Ottaiano A, Cuomo A, Wirz S, Coluccia S. Trends in research on artificial intelligence in anesthesia: A VOSviewer -based bibliometric analysis. *Intel Artif* 2022; 25: 126–37.

[22]. Nandiyanto ABD, Al Husaeni DF. A bibliometric analysis of materials research in Indonesian journal using VOSviewer. *J Eng Res* 2021; 9: 1–16.

[23]. Züllig T, Trötzmüller M, Köfeler HC. Lipidomics from sample preparation to data analysis: a primer. *Anal Bioanal Chem* 2020; 412: 2191.

[24]. Nandiyanto ABD, Al Husaeni DN, Al Husaeni DF. A bibliometric analysis of chemical engineering research using vosviewer and its correlation with Covid-19 pandemic condition. *J Eng Sci Technol* 2021; 16: 4414–22.

[25]. Farooq R, Rehman S, Ashiq M, Siddique N, Ahmad S. Bibliometric analysis of coronavirus disease (COVID-19) literature published in Web of Science 2019–2020. *J Fam Community Med* 2021; 28: 1.

[26]. Meng L, Wen K-H, Brewin R, Wu Q. Knowledge atlas on the relationship between urban street space and residents' health—A bibliometric analysis based on VOSviewer and CiteSpace. *Sustainability* 2020; 12: 2384.

[27]. Vittori A, Cascella M, Leonardi M, Monaco F, Nocerino D, Cuomo A, *et al.* VOSviewer-based bibliometric network analysis for evaluating research on Juvenile Primary Fibromyalgia Syndrome (JPFS). *Children* 2022; 9: 637.

[28]. Martins J, Gonçalves R, Branco F. A bibliometric analysis and visualization of e-learning adoption using VOSviewer. *Univers Access Inf Soc*, 2022.

[29]. Van Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* 2017; 111: 1053–70.

[30]. Mihalic T, Mohamadi S, Abbasi A, Dávid LD. Mapping a sustainable and responsible tourism paradigm: A

bibliometric and citation network analysis. *Sustainability* 2021; 13: 853.

[31]. Vasconcelos RN, Lima ATC, Lentini CAD, Miranda G V., Mendonça LF, Silva MA, et al. Oil Spill Detection and Mapping: A 50-Year bibliometric analysis. *Remote Sens* 2020; 12: 3647.