Bibliometric Analysis Using Publish or Perish and VOS Viewer: Machine Learning in Relation to Cardiovascular Disease

[1*]Akbar Sidqi, [2]Suhartono
[1,2]UIN Maulana Malik Ibrahim, Malang, Indonesia
[*]*akbarsidqi@gmail.com*

ABSTRACT
Cardiovascular disease, commonly known as heart disease, is a condition where blood vessels narrow or become blocked, potentially leading to heart attacks, angina, or strokes. Disorders affecting the heart muscle, valves, or rhythm are also considered forms of heart disease. According to the American Heart Association (2017), cardiovascular disease causes approximately 17.3 million deaths worldwide, with around 3 million of these occurring before the age of 60. Global statistics indicate that there are 9.4 million deaths annually due to cardiovascular disease, of which 45% are attributed to coronary heart disease. It is estimated that this number will increase to 23.3. million in 2030. This research is to identify patterns and trends and comprehensively evaluate data on cardiovascular disease. The method used is bibliometric analysis and VOS Viewer with a minimum number of term occurrences of 3 clusters with 96 relevant terms to visualize entity relationships, patterns and trends. In the analysis process, 50 articles were used to obtain scientific publications related to the topic of cardiovascular disease from the 2020-2024 period. with VOS Viewer you can produce the latest gaps and data analysis

Keywords: Bibliometric, Cardiovascular, Disease, Clusters, Network

1. INTRODUCTION

Cardiovascular disease, commonly known as heart disease, is a condition where blood vessels narrow or become blocked, potentially leading to heart attacks, angina, or strokes. Disorders affecting the heart muscle, valves, or rhythm are also considered forms of heart disease. According to the American Heart Association (2017), cardiovascular disease causes approximately 17.3 million deaths worldwide, with around 3 million of these occurring before the age of 60. Global statistics indicate that there are 9.4 million deaths annually due to cardiovascular disease, of which 45% are attributed to coronary heart disease [1][2]. Cardiovascular disease (CVD) encompasses a range of disorders affecting the heart and blood vessels, including coronary heart disease (CHD) [3], cerebrovascular disease, hypertension, and peripheral vascular disease. The term CVD also includes other conditions such as rheumatic heart disease and congenital heart disease [4][5]. The importance of accurate and timely diagnosis of heart disease is highlighted for the effective prevention and treatment of heart failure [6]. The World Health Organization (WHO) reports that 30% of deaths are caused by smoking, equating to 17.3 million people, and this figure is expected to rise to 23.3 million by 2030. Smoking [7][8][9][10][11], significantly increases the risk of cardiovascular disease, especially in low-income countries. In Indonesia, cardiovascular disease accounts for 80% of total deaths, making it the leading cause of mortality.

In previous research, the comparison between the K-NN method and Naive Bayes Classifier on datasets related to heart disease highlighted the relevance of different algorithms across various domains [12][13]. Machine learning methods are crucial in cardiology, particularly for the prediction and diagnosis of heart disease. Numerous studies have demonstrated the effectiveness of machine learning algorithms in predicting heart disease using various datasets [14][15]. The ensemble classifier, a type of approach in machine learning, has shown promise in improving the accuracy of predicting heart disease compared to traditional classifiers [16][17]. Those algorithms include decision trees, support vector machines, random forests, logistic regression, and ensemble learning methods. The use of machine learning in cardiology encompasses the diagnosis of coronary artery disease, heart failure, and improved risk assessment in cardiovascular diseases [18][19][20]. Researchers highlight the importance of feature selection and classification methods in improving the accuracy of heart disease prediction systems [21][22]. By leveraging machine learning techniques, researchers can identify important features to build efficient prediction models. Furthermore, machine learning is not only used to predict heart disease but also to estimate survival in cases of heart failure [23].

In the field of public health, age, smoking, alcohol consumption, dietary patterns, physical activity levels, obesity, and hypertension have been the subject of deep concern for numerous researchers, healthcare practitioners, and policymakers. As major triggers of several serious diseases, including cancer, heart disorders, and respiratory issues, smoking represents a global challenge that requires a multidisciplinary approach and collaborative efforts to address. Each year, approximately three million people die from cardiovascular diseases due to tobacco consumption, making it a leading cause of death worldwide. The etiology of functional post-stroke ischemic stroke is multifactorial. The pathophysiology of ischemic stroke involves two crucial stages, namely (i) supply disturbances such as hypertension, smoking, diabetes, and blood to the brain, and (ii) the development of hypoxia and hyperlipidemia, while factors that cannot be modified, and tissue necrosis in the brain. In ischemic stroke, modifiable factors such as genetics, age, gender, and early blood supply decline.

The purpose of bibliometric analysis using VOS Viewer is to identify patterns and trends in scientific research related to cardiovascular disease with previously conducted machine learning models. By utilizing bibliographic data, this analysis aims to depict and evaluate the progress of knowledge in the field of health, including research focus, collaboration among researchers, and

the impact of scientific publications in academic literature. Thus, this bibliometric analysis is expected to provide comprehensive insights into cardiovascular disease.

1.1 LITERATURE REVIEW

Bibliometric analysis is a method used to evaluate the distribution of publication and citation counts from various literature sources. Topics in bibliometric analysis can be described both qualitatively and quantitatively. Bibliometric indicators are used to measure the progress of science by considering characteristics and developments in the field. The reliability of bibliometric indicators is influenced by the selection of available databases and the identification of publications based on author affiliation. Bibliometric indicators are more effective when used on a large aggregation scale, but less suitable for evaluating small individual or research team.

Bibliometric data collection related to cardiovascular disease caused by smoking uses the Publish or Perish application. The Publish or Perish application is software useful for researchers in analyzing and evaluating their scholarly work. This tool accesses publication data from various sources, including Google Scholar, Crossref, PubMed, Open Alex, Scopus, Semantic Scholar, and Web of Science.

Bibliometric data analyzed is generally large-scale and objective, involving publication counts, citations, as well as keyword and topic frequencies. It often requires a balanced evaluation between objective and subjective aspects, depending on informed techniques and procedures.

2. RESEARCH METHOD

In this study, data was obtained from the Publish or Perish application by conducting searches using the keywords "cardiovascular disease," "OR," "And," and "bioinformatics." The data obtained from the Publish or Perish search was then compiled for bibliometric analysis, which consists of four stages: searching, filtering, examining bibliometric attributes, and bibliometric analysis (Julia et al., 2020). The stages of the research can be outlined as follows.

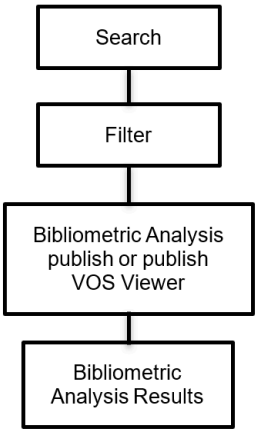

Figure 1. Research Method

Search Stage: The Publish or Perish application was chosen as the primary source for bibliographic searches due to its reputation as an application that provides peer-reviewed literature and publications. Filtering Stage: The journals to be analyzed are evaluated. The bibliography included article titles, abstracts, keywords, articles, or reviews. From the initial search using the Publish or Perish application, 377 bibliographies related to cardiovascular disease were found. The data obtained from the Publish or Perish search were then filtered to obtain 20 bibliographies for discussion.

Bibliometric Analysis: In this study, bibliometric analysis was conducted based on three aspects formulated as problem statements. It is hoped that through this analysis, the questions posed at the beginning of the research can be answered. To conduct this bibliometric analysis, the VOS Viewer application was used as a tool, allowing for visualization of the analysis results. VOS Viewer is a computer program specifically designed to visualize bibliometric maps, utilizing text-mining functions to depict networks or relationships in article citations.

The results of the bibliometric analysis are the filtered results from the search using Publish or Perish with articles relevant to the research theme. After filtering, they are exported in CSV Excel format. The exported file can then be used in the VOS Viewer application. The next stage involves selecting the 'create' option in VOS Viewer, followed by the creation of a map based on bibliographic data using the file data obtained from Publish or Perish. Once these steps are completed, the bibliometric analysis can begin.

3. RESULT AND ANALYSIS

Based on the search using the Publish or Perish application, out of 500 articles searched, it was found that there are 50 publications with article titles that match the predetermined topic and utilize machine learning models, as shown in Figure 2.

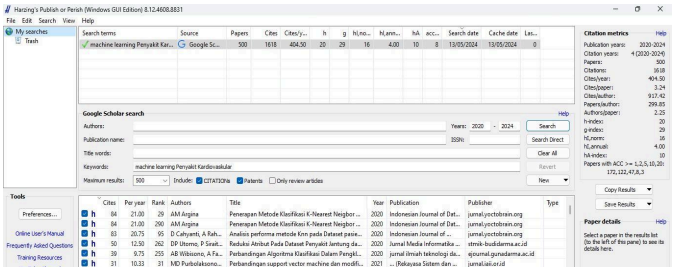

Figure 2. Search Results Publish or Perish

The publication development of articles on machine learning topics related to cardiovascular diseases between 2020-2024 in the Google Scholar database has significantly increased each year. Figure 3 shows that the highest number of publications regarding cardiovascular diseases was in 2023 with 177 articles or 38% of the total data.

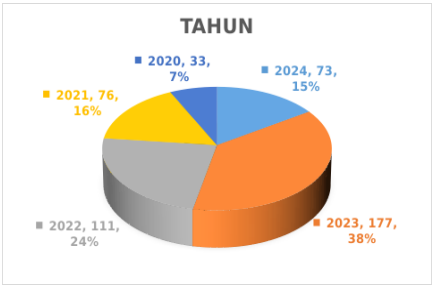

Figure 3. Data mapping of Publication Articles

Based on the percentage calculation shown in Figure 3, it can be concluded that the growth of machine learning publications on cardiovascular diseases has significantly increased, especially in 2023, which recorded the highest number of publications with 177 articles, accounting for 38%. This confirms that research on machine learning in cardiovascular diseases still has attraction and potential for further study and expansion. The data listed in the table indicates that the most cited article, titled "Attribute Reduction in Heart Disease Dataset and Classification Using C5.0 Algorithm" written by DP Utomo, P Sirait, R Yunis in 2020, recorded 50 citations. Further information about other scientific articles published between 2020 and 2024 is shown in Table 2.

Table 2. Research Based on Citations

Pref	Cites	Authors	Title	Year	Publication
1	50	DP Utomo, P Sirait, R Yunis	Attribute Reduction in Heart Disease Dataset and Classification Using C5.0 Algorithm	2020	stmik-budidarma.ac.id
2	45	Spencer, R., Thabtah, F., Abdelhamid, N., & Thompson, M. P	Exploring feature selection and classification methods for predicting heart disease.	2020	Digital Health
3	39	AB Wibisono, A Fahrurozi	Comparison of Classification Algorithms in Classifying Coronary Heart Disease Data	2020	ejournal.gunadarma.ac.id
4	29	D Derisma	Performance Comparison of Algorithms for Predicting Heart Disease with Data Mining Techniques	2020	jurnal.polibatam.ac.id
5	28	F Handayani	Comparison of Support Vector Machine, Logistic Regression, and Artificial Neural Network in Heart Disease Prediction	2021	JEPIN (Jurnal Edukasi Dan Penelitian Informatika)

In mapping the development of publications on machine learning in cardiovascular diseases, 7 articles were identified. It can be concluded that classification is the most commonly used machine learning model in cardiovascular diseases.

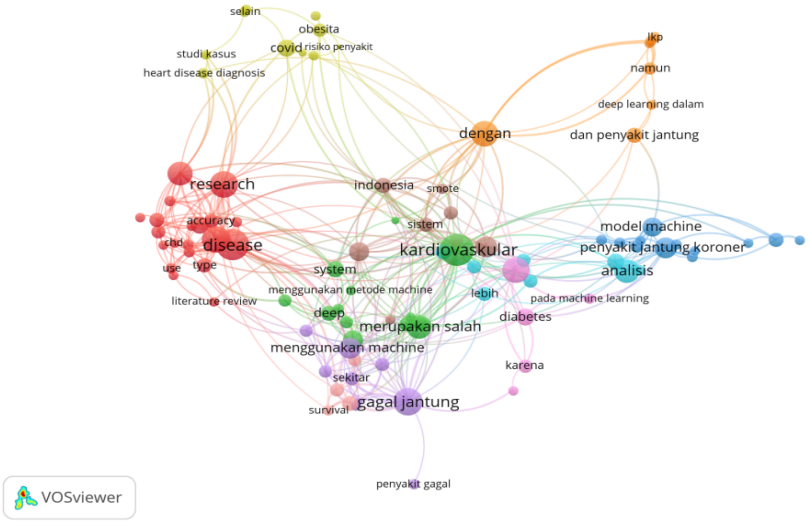

Figure 4. Visualization of machine learning networks in cardiovascular disease

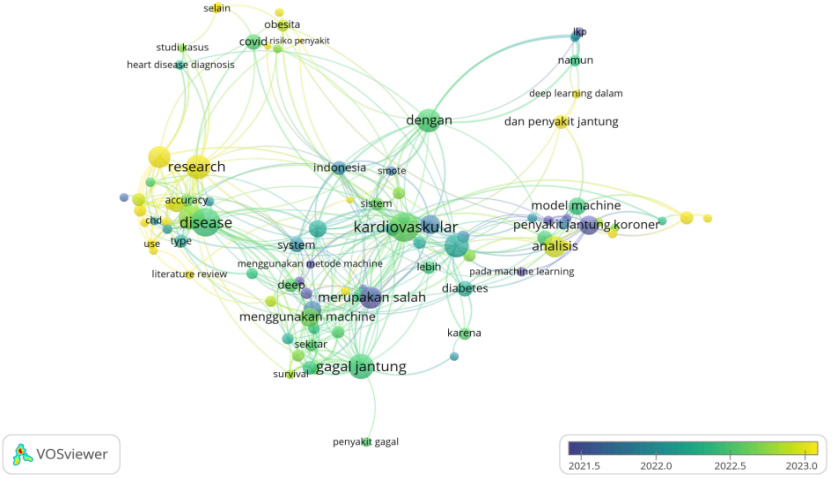

Figure 5. Overlay visualization of machine learning on cardiovascular disease

4. CONCLUSION

Based on the Bibliometric Analysis results regarding cardiovascular diseases caused by high blood pressure, smoking, dietary patterns, obesity, physical inactivity, and alcohol consumption, in cardiovascular disease research from 2020-2024, in 2023, there were 177 articles or 38% of the total data. The most commonly used machine learning method is classification. For further research on cardiovascular diseases, deep learning methods can be used, which have fewer connected variable pathways.

REFERENCES

[1] Dito Putro Utomo, Pahala Sirait, Roni Yunis, "Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0" 2020, Jurnal Media Informatika Budidarm, doi: https://doi.org/10.30865/mib.v4i4.235.

[2] AB Wibisono, A Fahrurozi, "Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner" 2020, ejournal.gunadarma.ac.id, doi: https://doi.org/10.35760/tr.2019.v24i3.2393.

[3] Jagentar Parlindungan Pane, Lindawati Simorangkir, Praska Indah Sari Br Saragih, "FAKTOR-FAKTOR RISIKO PENYAKIT KARDIOVASKULAR BERBASIS MASYARAKAT", Volume 4 Nomor 4, November 2022 e-ISSN 2715-6885; p-ISSN 2714-9757.

[4] D Derisma, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining" 2020, jurnal.polibatam.ac.id, doi: https://doi.org/10.30871/jaic.v4i1.2152.

[5] F Handayani, "Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung" 2021, EPIN (Jurnal Edukasi Dan Penelitian Informatika).

[6] Agus Wantoro, Admi Syarif, Khairun Nisa Berawi, Kurnia Muludi, Sri Ratna Sulistiyanti, Sutyarso, "Implementasi Metode Pembobotan Berbasis Aturan Dan Metode Profile Matching Pada Sistem Pakar Medis Untuk Prediksi Risiko Hipertens" 2021, Jurnal TEKNOINFO, doi: https://doi.org/10.33365/jti.v15i2.1523.

[7] Astuti, Febriana Capritasari, Rafiastiana Iskardyani, Delfi Alifiar, Ilham Irham,Muhammad "Identifikasi variasi dan Ekspresi Gen Pada Stroke".
[8] Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classiffier pada Data Set Penyakit Jantung", 2020, IJODAS.
[9] Liu, Wentao Xu, Yuxin Bai, Shengbin Liao, Libin."Bioinformatics analysis of key biomarkers for bladder cancer", Biomedical Reports. 2023, doi: https://doi.org/10.3892/br.2022.1596.
[10] Suhartono, Totok C, Syahiduz Z, Analisis Bibliometrik Penelitian Pohon Keputusan untuk Prediksi Kanker Payudara, Journal of Documentation and Information Science, Vol. 7 No. 2 (2023), doi: https://doi.org/10.33505/jodis.v7i2.216
[11] Huang, Xiaowei Wu, Bian Zhang, Fangxue Chen, Fancheng Zhang, Yong Guo, Huizhi Zhang, Hongtao, "Epigenetic Biomarkers Screening of Non-Coding RNA and DNA Methylation Based on Peripheral Blood Monocytes in Smokers",Frontiers in Genetics. 2022, doi: https://doi.org/10.3389/fgene.2022.766553.
[12] Inggrianti Pratiwi Putri, "Analisis Performa Metode K- Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascula" 2021, IJODAS, doi: https://doi.org/10.33096/ijodas.v2i1.25.
[13] Shemuel, Josia Angelique, Priscilla Josephine, Evalina Ryan Fugaha, Daniel Gabriela, Vania Alhusain, Shaheer Parikesit, Arli Aditya "Comparative Gene Analysis of Squamous Cell Lung Carcinoma Between Smoking and Non-smoking Individuals", berita biologi. 2023, doi: https://doi.org/10.55981/beritabiologi.2023.908.
[14] Anderies, A., Tchin, J. A. R. W., Putro, P. H., Darmawan, Y. P., & Gunawan, A. A. S. (2022). Prediction of heart disease uci dataset using machine learning algorithms. Engineering, MAthematics and Computer Science (EMACS) Journal, 4(3), 87-93.
[15] Mondal, A., Mondal, B., Chakraborty, A., Kar, A., Biswas, A., & Majumder, A. B. (2023). Heart disease prediction using support vector machine and artificial neural network. Artificial Intelligence and Applications, 2(1), 59-65, doi: https://doi.org/10.47852/bonviewaia3202823.
[16] Sameer, S. and Sriramya, P. (2021). Improving the efficiency by novel feature extraction technique using decision tree algorithm comparing with svm classifier algorithm for predicting heart disease. Alinteri Journal of Agriculture Sciences, 36(1), 713-720.
[17] Yousefi, S. (2021). Comparison of the performance of machine learning algorithms in predicting heart disease. Frontiers in Health Informatics, 10(1), 99.
[18] Machine learning algorithms for prediction of heart disease. International Research Journal of Modernization in Engineering Technology and Science.
[19] Nakajima, K. and Maruyama, K. (2022). Nuclear cardiology data analyzed using machine learning. Annals of Nuclear Cardiology, 8(1), 80-85.
[20] Juárez-Orozco, L. E., Клен, P., Niemi, M., Ruijsink, B., Daquarti, G., Es, R. v., & Knuuti, J. (2022). Artificial intelligence to improve risk prediction with nuclear cardiac studies. Current Cardiology Reports, 24(4), 307-316.
[21] Spencer, R., Thabtah, F., Abdelhamid, N., & Thompson, M. P. (2020). Exploring feature selection and classification methods for predicting heart disease. Digital Health, 6, 205520762091477.
[22] Maini, E., Venkateswarlu, B., & Gupta, A. (2019). Determination of significant features for building an efficient heart disease prediction system. International Journal of Recent Technology and Engineering (IJRTE), 9(2), 4499-4504.
[23] Tak, A., Parihar, P. M., Mathur, S., Das, B., & Sawal, D. (2022). Survival prediction in heart failure using machine learning algorithms.