BAREKENG: Journal of Mathematics and Its Applications
June 2023 Volume 17 Issue 2 Page 1103-1112
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss2pp1103-1112

BIBLIOMETRIC ANALYSIS OF NEURAL BASIS EXPANSION ANALYSIS FOR INTERPRETABLE TIME SERIES (N-BEATS) FOR RESEARCH TREND MAPPING

Dewi Retno Sari Saputro[1]*, Heri Prasetyo[2], Antoni Wibowo[3], Fadiah Khairina[4], Krisna Sidiq[5], Gusti Ngurah Adhi Wibowo[6]

[1,4,5]*Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Sebelas Maret*
Ir. Sutami Street No. 36 Kentingan, Surakarta, Central Java, 57126 Indonesia

[2]*Department of Informatics, Faculty of Informatics and Data Science, Universitas Sebelas Maret*
Ir. Sutami Street No. 36 Kentingan, Surakarta, Central Java, 57126 Indonesia

[3]*Master of Information Technology, Bina Nusantara University*
K. H. Syahdan Street No. 9, Kemanggisan, Palmerah Jakarta 11480, Indonesia

[6]*Department of Mathematics, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Halu Oleo,*
H.E.A Mokodompit Street, Anduonouhu Kendari, Southeast Sulawesi, 93232, Indonesia

*Corresponding author's e-mail: * **dewiretnoss@staff.uns.ac.id***

Article History:

Received: 24th February 2023
Revised: 8th May 2023
Accepted: 10th May 2023

Keywords:

Bibliometrics;
N-BEAT;
Publication;
Map research;
VOSviewer

ABSTRACT

Bibliometrics is the statistical analysis of articles, books, and other forms of publication. The bibliometrics analysis is performed with data on the number and authorship of scientific publications and articles, and citations to measure the work of individuals or groups of researchers, organizations, and countries to identify national and international networks and map developments in new multidisciplinary fields of science and technology. In addition, bibliometrics assesses and maps the research, organization, and country of researchers at a given time period. The Bibliometric analysis also has advantages which include mapping relationships between concepts, mapping research directions or trends, mapping state of the art (the novelty of the results of research conducted), and providing insights related to fields, topics, and research problems for future works. This study aims to determine the growth and development of N-BEATS publications, their distribution, variable keywords, and author collaboration using a bibliometric network. The research method used in this paper, through screening of articles obtained from the Scopus database page in 2008-2022, is used for citations in the form of metrics. At the same time, they are visualizing the metadata with VOSviewer. Data was collected from the direct science database with the keyword N-BEATS. The results show that 2022 has the highest number of publications, reaching 310 publications (14.90%). The distribution of research publications on N-BEATS shows a perfect distribution. Terms in the N-BEATS variable often appear and are associated with other variables.

How to cite this article:

D. R. S. Saputro, H. Prasetyo, A. Wibowo, F. Khairina, K. Sidiq and G. N. A. Wibowo., "BIBLIOMETRIC ANALYSIS OF NEURAL BASIS EXPANSION ANALYSIS FOR INTERPRETABLE TIME SERIES (N-BEATS) FOR RESEARCH TREND MAPPING," *BAREKENG: J. Math. & App.*, vol. 17, iss. 2, pp. 1103-1112, June, 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

Bibliometrics comes from the words Biblio or bibliography and metric. Biblio means book or bibliography, and metric is related to measurement. Bibliometrics is a quantitative method used to measure, track, and analyze scientific literature [1]. According to [2], bibliometrics is divided into two groups, namely groups that study publication distribution and that discuss citation analysis. The first group analyzes documents quantitatively so that there are 3 (three) bibliometric laws, which are described in the following paragraphs.

The first is Lotka's law (1926**)** [3], to calculates the distribution of results of different authors. It contains a mathematical relationship between the number of works produced and the number of authors. Lotka studied this phenomenon in 1926, but it was not until 1949 that Lotka's law of author productivity was declared. Lotka's law is one of the main research subjects in bibliometrics. According to Lotka, if a hundred people make one work, then a quarter of every 100 people will make two works, one in nine out of 100 will make three works, and there will be a third. Sixteen out of 100 people do the work. Four works, etc. This is Lotka's inverse square law of productivity.

The second is Zipf's law (1933) [4] [5], to classify words and their frequency of occurrence in literature, namely the frequency of occurrence of words in the document (words appear or frequency) turns out to have a specific pattern. It can be used as a parameter during indexing. Studies like this have yet to be developed in Indonesia, but the research is known as Zipf's Law abroad.

The third is Bradford's law (1934), which estimates the exponential decline in search results for references in scientific journals. One formula stated that if the number of articles sorted journals in a field into three groups, each containing about a third of all articles, the number of journals in each group would be equal to $1:n:n^2$ [6] [7], where n is the "Bradford multiplier" and depends on the particular journal collection. There are several formulations related to this principle. In many disciplines, this pattern is referred to as the Pareto distribution.

According to [8], bibliometrics has three components: bibliometrics for bibliometrics; bibliometrics for science (information science); and bibliometrics for science policy and management. The Bibliometrics analysis (BA) is used to quantitatively analyze and measure various specific indicators in the published literature in a particular field to obtain knowledge maps based on large databases [9]. It allows researchers to summarize publication information regarding article distribution by author, institution, year, journal, and specialty, author-organization collaboration, and split word analysis [10]. The BA is based on relevant information about scientific publications/documents. The necessary information that can be used is to identify the source (journal/document, volume, page), author's name, institution/institutional address, references, type of document, title, keyword, abstract and subject. Based on some publication/documentary information, the researcher used the author's name of the document type of the journal in Conciencia as the basic unit in the BA.

The current research results are the results of research conducted and published in journals, proceedings, seminars, books, and the like. Bibliometrics is widely used to identify subject trends, such as; identifying key journals and library usage patterns. Bibliometrics is also used to build research models of scientific communication. Most of these models are tested and used mainly at the institutional level to explain scientific productivity; explain publication growth, identify critical journals; and document screening.

This article is a bibliometric analysis from the direct science page, which discusses Neural Basis Expansion Analysis for Interpretable Time Series (N-BEATS) using VOSviewer. N-BEATS is a deep learning architecture that connects backward and forwards residuals to a stack of connected classes commonly used for univariate time series prediction [11]. The N-BEATS deep learning architecture has the property of being independent of any particular specification of the time series or input rate [12]. The advantage of using N-BEATS architecture over other deep learning architectures is interpretability and does not depend on feature engineering or input scaling in a particular time series [13]. Using the N-BEATS architecture, it is possible to process time series data without preprocessing or decomposition [11].

N-BEATS architecture is divided into two basic block parts. First, the basic block has four fully connected (FC) layers with nonlinear ReLu. This part produces two outputs, a forward predictor or "forecast" (θ^f) and a backward predictor or backcast (θ^b). Second, the basic block has a fully connected (FC) layer without nonlinear ReLu and base layers g^f and g^b. The base layer is a linear projection layer. In the block

ℓ,there is an input x_ℓ, and two output vectors $\widehat{x_\ell}$, and $\widehat{y_\ell}$. $\widehat{x_\ell}$ is the block's best estimate of x_ℓ or backcast, and $\widehat{y_\ell}$, is a forward forecast on the block of horizon length (H),which has a range of values from 2H to 7H.

The N-BEATS architecture uses two configurations: generic architecture (N-BEATS-G) and interpretable architecture (N BEATS-I). Some Scopus-indexed N-BEATS research in the last ten years needs to be mapped to obtain the latest research trends. Therefore, this article examines N-BEATS for research trend mapping with bibliometric analysis.

2. RESEARCH METHODS

This research is a literature review by summarizing and evaluating the N-BEATS dataset using bibliometric methods based on Co-word analysis. Co-word analysis is based on the assumption that the keywords of a paper or article are an adequate description or can make a representation of the content of its content. The research was conducted by screening articles from the Scopus database site from 2008 to 2022 with the keyword N-BEATS **[14]**. The data obtained is then called metadata. Metadata in the form of the number of publications per year, journals that contain articles in the field of N-BEATS, authors, author origin, and subject.

Meanwhile, the development trend of international publications was analyzed with VosViewer software. The number of related documents for the last 14 years from 2008 to 2022 in 647 documents with reference curation: Mendeley, ProQuest RefWorks, Zotero (RIS), EndNote (RIS). The research steps are shown in **Figure 1**.

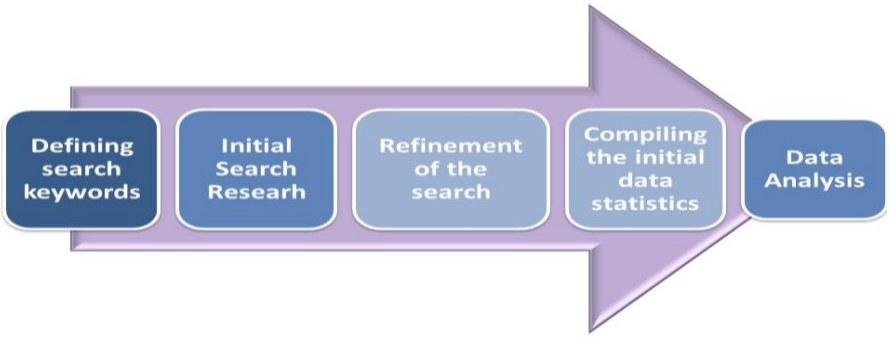

Figure 1. Research steps

Based on **Figure 1**, the steps taken are defining search keywords, initial search research, refinement of the search, compiling the initial data statistics, and data analysis. Refinement of the search is done by adding filters and restrictions so that the data obtained will be more specific and specific. Compiling the initial data statistics is done by looking at the distribution of each research using various graphical visualizations such as histograms and line graphs.

3. RESULTS AND DISCUSSION

3.1. Development of N-BEATS Publications in Scopus Metadata

Scopus is the world's most extensive collection of literature abstracts, with citations that provide summaries of a wide range of peer-reviewed scientific publications and research. Scopus can help researchers effectively track, analyze and visualize research. More than 22,000 high-quality abstracts published by 5,000 publishers worldwide are provided in the Scopus database from science, technology, medicine, social sciences, arts, and literature. Scopus has 55 million records since 1823, 84% of which are from Elsevier reference citations **[15]**.

N-BEATS: Neural Basis Expansion Analysis is used for interpretable Time Series estimates. N-BEATS uses a simple but powerful architecture of ensembled feed-forward networks with stacked residual blocks of forecasts and ‘backcasts’. History of previous research related to the N-BEATS algorithm is shown by **Figure 2**.

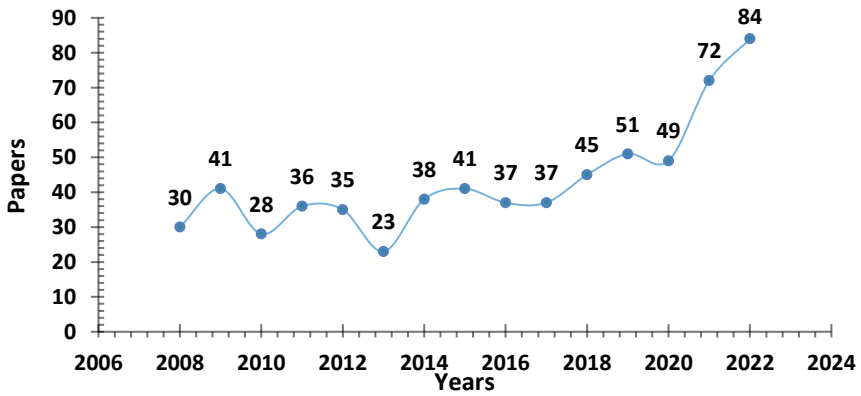

Figure 2. Visualization of the trend in the number of papers

Figure 2 shows that the development of N-BEATS publications from 2008 to 2022 has increased significantly, and the highest Scopus index occurred in 2022, reaching 84 publications (14.90%). Based on **Figure 2**, the articles processed in this study included 647 data and showed an upward trend except in 2010 and 2013, which showed a decrease. **Figure 3** shows the number of studies by country.

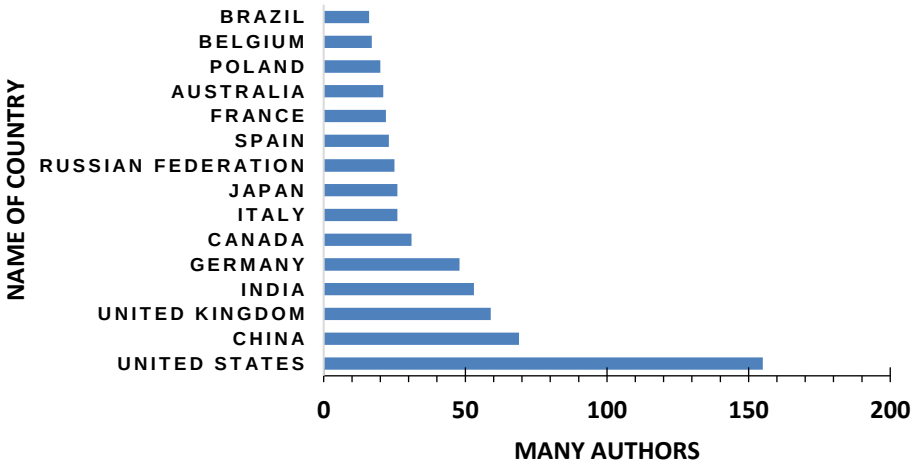

Figure 3. Visualization of author affiliation

Figure 3 shows the 15 most countries as author affiliations, author affiliations are dominated by the United States, China, United Kingdom, India, Germany, and so on. **Figure 4** shows the document types in the metadata.

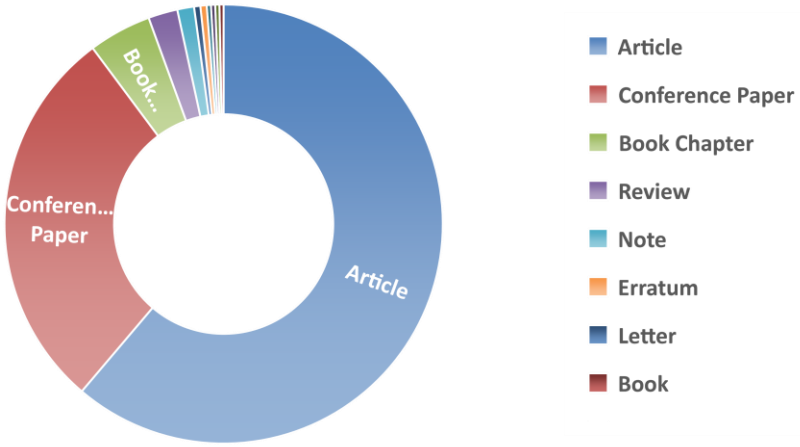

Figure 4. Visualization of document types

Based on **Figure 4**, the most common type of documents are articles, conference papers, and book chapters. The other document types have yet to evolve significantly.

3.2. Bibliometrics with VosViewer

The earliest known bibliometric research dates back to the early 20th century. In the early stages of its development, the bibliometric method was widely applied in the medical field [16]. Around the 1940s, the development of bibliometrics had strong foundations from the formation and maturity of Brad Ford's Law, Zipf's Law, and Lotka's Law [3] [17] [18]. Since then, this method has been widely used in natural sciences [19], subjects such as mathematics [20], physics [21], and chemistry [22].

Visualization of Similarities Viewer (VOSviewer) is a computer program with free access to explore bibliometric knowledge network maps [23]. The algorithm used in VOSviewer is similar to the Multi-Dimensional Scaling (MDS) algorithm. VOSviewer is capable of generating clusters that automatically display different colors on the map. The clustering algorithm works with a parameter (γ) that can be modified to get more or fewer clusters. Still, according to [23], cluster density and color can be visualized using VOSviewer. The advantage of VOSviewer over other analysis applications is that it uses text mining to identify mapping-related noun phrase combinations and built-in clustering methods to test the data [24] [25]. Although there are various programs that analyze the similarity of text units and matrices, VOSviewer has a strength that lies in its intuition [26]. VOSviewer also has various interactive features and options that make it easy for users to access and explore bibliographical data, such as the co-occurrence relationships between key terms and concepts and the number of citations therein [24] [26].

On the metadata, bibliometric analysis was carried out using the VOSviewer application, with the analysis carried out including Co-occurrence analysis, used to reveal research topics/variables statistically, and co-authorship analysis used to find the relationship between several researchers based on research documents produced by researchers. The bibliographic analysis explores research fields that change yearly due to researchers' findings and attention to different research problems.

Bibliometric analysis is done by visualizing networks, overlays, and densities to identify bibliometric networks between articles or online publications from downloaded metadata. A bibliometric network consists of vertices in the form of circles that represent keywords, while the edges or nodes of the network represent the relationship between pairs of vertices. Mapping and clustering in bibliometric analysis through VOSviewer software are complementary. Mapping can be used to obtain a detailed picture of the structure of the bibliometric network [23]. In addition, clustering provides an overview or preview of bibliometric clustering. The approaches used in the bibliometric analysis are citation analysis to see one article cited by other articles and co-citation analysis to find two or more articles cited by one article.

3.3. Bibliometrics Study Coverage

Scientific concepts in a document can be seen through the words (co-words) used. Co-occurrence analysis is the basis of collingual analysis. Keywords that originate from two or more different documents serve as indices for these documents [27]. Co-words analysis aims to analyze various documents' content, patterns, and trends by measuring the terms' strength [28] [29] [30] [31]. Synonym analysis calculates the number of research document keywords that co-occur in the articles under study. Previously, these keywords were set by the author. The relationship between documents is stronger when many keywords exist in a particular document group.

The co-words analysis map is based on the appearance of the same keyword in the article, the occurrence of unique or important terms in the article, and the title or abstract. These keywords and terms represent a concept from the subject analysis. Using denormalized keywords can lead to inconsistent terms, and normalization requires using a thesaurus. A thesaurus is a list of terms that cover one particular field so that the terms used are more specific. A thesaurus differs from a list of subject headings, which are usually general and cover all fields of science. Indexing uses descriptors that represent each concept. Standardizing keywords with a thesaurus aim to make the words used consistent so that one term is used for concepts presented in different articles and has the same meaning.

Furthermore, the metadata was analyzed using the VOSviewer application by selecting the data option 'create a map based on text data,' which aims to create a term network or term relationship (term) based on text data. **Figure 5** shows the results of this visualization through extraction in VOSviewer with title and abstract fields with complete counting and 10 number of term occurrences and obtained eight groups.

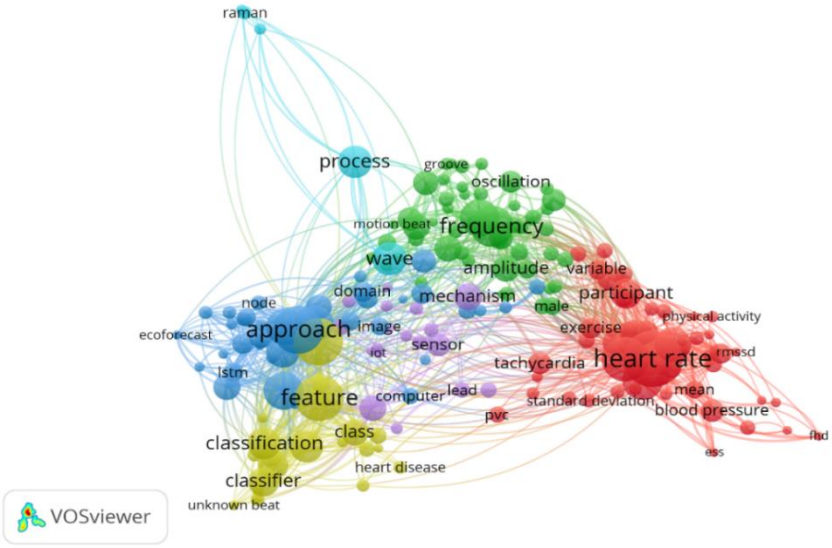

Figure 5. Visualization with Title and abstract restriction

Figure 5 describes what variables frequently appear associated with N-BEATS. The larger the circle shown for each variable indicates that the variable is used more frequently. The terms dominate this: approach, feature, heart rate, frequency, computer, mechanism, wave, and application. There are 7 clusters for co-authorship analysis with an author analysis unit and complete counting method with a restriction of a maximum number of authors ten and authors minimum of 2 documents. It can be seen that Zhang, y; Wang, h; Wang, y; li, y; Xu, x; Zhou, y; and li, j dominate each cluster, and Zhang, y contributes the most as an author. **Figure 6** shows that some authors partnered with other authors to publish.

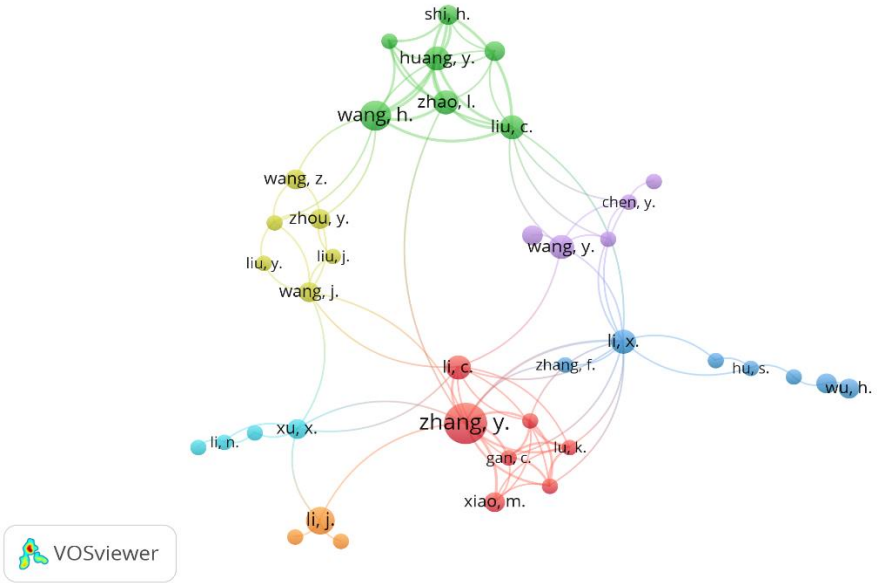

Figure 6. Visualization of Network Author and Co-authorship

Based on **Figure 6**, the larger the circle shown for each author's name, the more frequently the author publishes, such as Zhang. Next, there are 17 clusters for co-occurrence analysis type with keywords analysis unit and complete counting method with a minimum restriction of two keywords. Of the 1700 keywords obtained that meet the criteria, 172 documents result from their network co-occurrence with other keyword links, as shown in **Figure 7**.

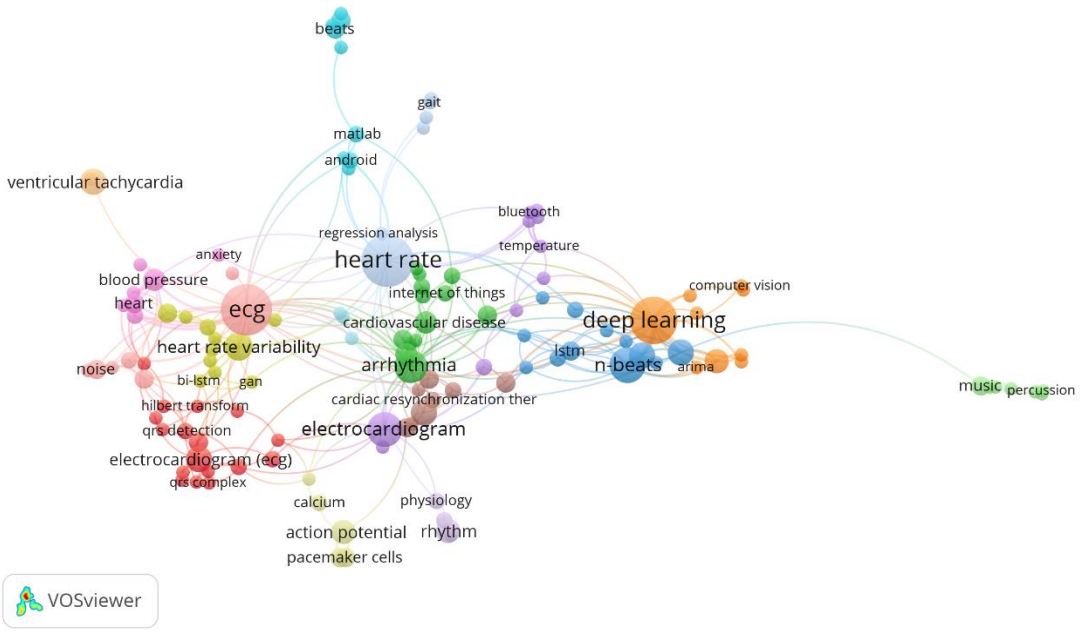

Figure 7. Network keywords and co-occurrence visualization

Based on **Figure 7**, much research has been conducted regarding deep learning, heart rate, and ecg. The development of N-BEATS-based research is closely related to LSTM, Deep learning, Neural networks, unsupervised domain, and several other deep learning applications. This can be seen in the visualization of **Figure 8**.

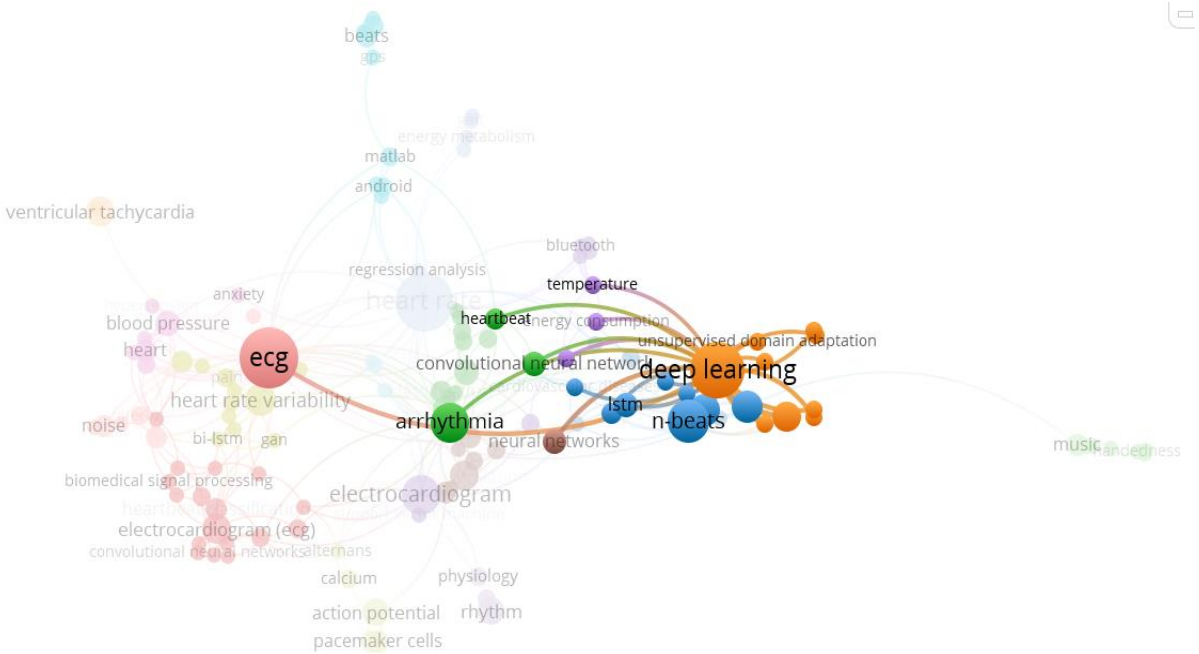

Figure 8. Visualization of Networking with Deep Learning term

Based on **Figure 8**, the N-BEATS algorithm has a close relationship with deep learning. Apart from that, other related algorithms are LSTM and CNN, as well as several fields of application such as temperature, heartbeat, and arrhythmia.

4. CONCLUSIONS

Based on the results and discussion, it can be concluded that the development of documents related to N-BEATS in 2008 - 2022 indexed by Scopus was highest in 2022, reaching 310 publications (14.90%).

China's Ministry of Education is the institution that publishes the most N-BEATS research, and the United States is the most significant contributor. Zhang, Y is the most prolific author with the most subjects in Medicine and Engineering. The N-BEATS development map is based on co-words clusters into seven and co-authors clusters into 9 clusters. The direction of N-BEATS research is growing over time, supported by collaboration, given the rapid development of interpretation in N-BEATS. In addition, the bibliometric mapping identifies authors who take the theme of the N-BEATS field and have a collaborative author relationship with each other. The last mapping is the development of the field of N-BEATS science based on co-occurrence (keywords). This mapping identifies the relationship between scientific concepts and nine dominant groups. Thus, the direction of future research is to serve as a basis for further research on archives and see the variety of topics that can be researched, not only in the national scope but also in the international realm.

REFERENCES

[1] R. C. Roemer and R. Borchardt, "A 21st Century Librarian's Guide to Bibliometrics, Altmetrics, and Research Impact," in *Meaningful Metrics*, Chicago, The Association of College and Research Libraries, 2015.

[2] S. Basuki, "Dari Bibliometrika Hingga Informetrika," *Jurnal Media Pustakawan : Media Komunikasi Antar Pustakawan*, vol. 23, no. 1, pp. 7-14, 2016.

[3] A. J. Lotka, "The frequency distribution of scientific productivity," *J. Wash. Acad. Sci.*, vol. 16, p. 317–323, 1926.

[4] C. Samuelsson, "Relating Turing's Formula, and Zipf's Law," in *4th Workshop on Very Large Corpora*, Sussex, 1996.

[5] H. Guiter and M. Arapov, Studies on Zipf's Law, Bochum: Studienverlag Brockmeyer, 1982.

[6] H. Birger and N. Jeppe, "Bradford's law of scattering: Ambiguities in the concept of "subject"," in *COLIS*, 2015.

[7] Musthafa, "Dalam Kumpulan Makalah Kursus Bibliometrika," Pusat Studi Jepang UI: Masyarakat Informatika Indonesia, 2002.

[8] W. Glänzel, "Bibliometrics as a research field: A course on theory and application of bibliometric indicators," 2003.

[9] O. Ellegaard and J. Wallin, "The bibliometric analysis of scholarly production: How great is the impact?," *Scientometrics*, vol. 105, no. 3, p. 1809–1831, 2015.

[10] R. Zeng and A. Chini, "A review of research on the embodied energy of buildings using bibliometric analysis," *Energy and Building*, vol. 155, p. 172–184, 2017.

[11] B. Oreshkin, G. Dudek, P. Pełka and E. Turkina, "N-BEATS neural network for mid-term electricity load forecasting," *Apply Energy*, vol. 293, pp. 1-33, 2021.

[12] B. Oreshkin, D. Carpov, N. Chapados and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," pp. 1-31, 2019.

[13] A. Bulatov, "Forecasting Bitcoin Prices Using N-BEATS Deep Learning Architecture," Sheridan College, Institute of Technology and Advanced Learning, 2020.

[14] "Scopus," Scopus, 2023. [Online]. Available: https://www.scopus.com.

[15] Elsevier, "content policy and selection," 2017. [Online]. Available: https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection.

[16] E. Humlme, "Statistical Bibliography in Relation to the Growth of Modern Civilization," University of Cambridge, London, 1923.

[17] S. C. Bradford, "Classic paper: Sources of Information on Specific Subjects," *Collect. Manag.*, vol. 1, p. 95–104, 1976.

[18] G. K. Zipf, The Psycho-Biology of Language, New York: Houghton Mifflin, 1935.

[19] P. O. Seglen, "Use of citation analysis and other bibliometric methods in evaluation of the quality of research," *Tidsskr. Den. Nor. Legeforen.*, vol. 109, p. 3229–3234, 1989.

[20] C. Campos and F. Redondo, "Bibliometrics and clinical chemistry," *Clin. Chem.*, vol. 37, p. 303–304, 1991.

[21] H. Behrens and P. Luksch, "Mathematics 1868–2008: A bibliometric analysis," *Scientometrics* , vol. 86, p. 179–194, 2010.

[22] S. Ahn, J. Kang and H. J. Lee, "Research Trends in Condensed Matter Physics Based on Bibliometric Analysis," *J. Inf. Int. Interdiscip.*, vol. 14, p. 1745–1760, 2011.

[23] E. W. Rice, "Bibliometric evaluations of modern Clinical Chemistry are needed," *Clin. Chem.*, vol. 29, p. 1858–1859, 1983.

[24] L. Waltman, N. J. V. Eck and E. C. M. Noyons, "A unified approach to mapping and clustering of bibliometric networks," *Journal of Informetrics*, vol. 4, no. 4, p. 629–35, 2010.

[25] N. V. Eck and L. Waltman, "Text mining and visualization using VOSviewer," *ISSI Newsletter*, vol. 7, no. 3, pp. 50-54, 2011.

[26] N. V. Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, p. 523–538, 2010.

[27] V. Diodato, Dictionary of Bibliometrics, New York: The Haworth Press, 1994.

[28] L. Leydesdorff and I. Rafols, "Interactive overlays: A new method for generating global journal maps from Web-of-Science data," *Journal of Informetrics*, vol. 6, no. 2, p. 318–332, 2012.

[29] M. D. Looze and J. Lemarie, "Corpus relevance through co-word analysis : an application to plants," *Scientometrics*, vol. 39, no. 3, p. 267–280, 1997.

[30] N. Coulter, "Software engineering as seen through its research literature: A study in co-word analysis," *Journal of American Society for Information Science*, vol. 49, no. 13, p. 1206–1223, 1998.

[31] J. Du and Y. Wu, "A bibliometric framework for identifying “Princes” who wake up the “sleeping beauty in challenge-type scientific discoveries," *Journal of Data and Information Science*, vol. 1, no. 1, pp. 50-68, 2016.