Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

Bibliometric Analysis of Deep Learning Applications in Diabetes

Arash Salehpour

Department of Computer Engineering, Rasht Branch, Islamic Azad University, Rasht, Gilan, Iran

E-mail: arash.salehpour4@gmail.com

Abstract

This study provides a bibliometric review of deep learning applications in diabetes between 2018 and 2022, with an analysis of the 2201 publications. This review highlights the influential aspects of deep learning in diabetes research from a bibliometric perspective. Deep learning has drawn significant interest from researchers, particularly those working in diabetes. Two well-known databases: Web of Science and Scopus, each of which having its own data format, are combined into a single format using the R programming language in R Studio, and the duplicates are removed. The Bibliometrix package is used to conduct quantitative analysis, which includes highlighting the primary journals, the works that have been referenced the most, the authors, nations, and institutions that have produced the most, as well as keyword clustering, paper split into sub-periods to track theme progression, and top trend analysis. The findings demonstrate a notable increase in publications since 2018. A plethora of studies are conducted on the practical applications of deep learning to treat diabetes, which is dramatically rising. IEEE Access, Scientific Reports, and Computers in Biology and Medicine are the top three most relevant journals. China is most productive and its publications are highly cited, while the USA comes second. Accuracy, atrial fibrillation, and heart infarction have recently been the hot topics. The most frequently used words are human, article, and diabetes mellitus. The findings help academics better understand the study area in this related field, which is one of the hottest research fields in Artificial Intelligence.

Keywords : Bibliometrics, Diabetes, Deep Learning, Scient metrics, Trends

1. Introduction

Diabetes is one of the current, much-discussed issues in the world. Governments are allocating a large portion of their GDP to social programs, and policies like immunization

DOI: https://doi.org/10.36548/jtcsst.2022.4.006
Received: 29.11.2022, received in revised form: 27.12.2022, accepted: 12.01.2023, published: 24.01.2023

have increased the average lifespan of citizens [1, 2]. This research is unique in that it applies deep learning to the study of diabetes, one of the hottest topics in Artificial Intelligence (AI) research, and focuses on bibliometric analysis. Today, about 450 million people have one of the several kinds of diabetes, in which the body either cannot generate insulin (Type 1) or cannot effectively utilize what it produces (Type 2) [3]. While traditional treatments have been less effective, technological advancements in intelligent devices have played an essential role in controlling diabetes. Considering disease conditions such as diabetes that require 24-hour care and monitoring, existing technologies have many advantages in personal and automatic care. Considering that many use mobile apps to monitor their health, many devices are trying to create new solutions for people with type 1 and type 2 diabetes to prevent this condition.

Insulin is a hormone generated by the pancreas that facilitates the transfer of glucose from meals into the cells for the use as energy. Sometimes human body may not make enough or any insulin or use insulin correctly. Glucose consequently lingers in the circulation rather than reaching the cells. Too much glucose in blood could lead to health risks over time. Diabetes is a chronic disease affecting many of the world's population today. Diabetes-specific technologies, such as nutrition and mobility control applications, insulin pumps, etc., play a significant role in diabetes control. Even though there is no treatment for diabetes, it can be managed to maintain health. Diabetes is frequently referred to as "borderline diabetes" or "a touch of sugar." These phrases indicate that a person does not have diabetes or has a lesser form of the disease. However, diabetes is hazardous in all forms [4].

The NHS [5] Diabetes Programs understands that one-on-one interventions do not always work for all patients. On the other hand, digital technologies offer new ways to improve the patient experience and outcomes, and make services more efficient and less stressful for clinicians and the public. Bibliometric analysis is a well-known and reliable approach for sifting through enormous amounts of scientific data and analyzing such data. Diagnosing a health issue is an essential and crucial element of healthcare. One of the trickiest tasks for medical professionals is determining the kind of diabetes an individual has. Nevertheless, assessing many criteria at the moment of diagnosis might sometimes provide erroneous findings. Consequently, interpreting and classifying diabetes is a tricky endeavor. The healthcare business has benefited tremendously from recent technology developments, particularly machine learning approaches. Numerous classification methods for diabetes have been described in the scientific literature [2].

Deep learning is a branch of AI that aims to mimic how humans access particular types of knowledge. The use of modeling and statistics in prediction is one of the critical components of data science. Deep-learning computer programmers go through a process similar to how a young child learns to recognize a dog. Each deep learning algorithm uses a nonlinear transformation system on its input and uses what it learns to produce a statistical model as its output as part of its learning hierarchy. In traditional machine learning, a programmer must instruct the computer what to look for to determine whether an image contains a dog. This is because the learning process is supervised. The computer's success rate entirely depends on the programmer's ability to precisely define a dog's feature set. This laborious process is known as "feature extraction." Deep learning allows the program to construct this set of features independently and without human intervention. Unsupervised learning is generally more accurate and faster. Deep learning programs need enormous amounts of training data and a lot of processing power to be accurate. Neither of these things was easily accessible to programmers before the big data and cloud computing eras.

In recent years, machine learning and deep learning methods have been utilized to predict diabetes and its effects. As a result, several researchers have worked to create instruments for forecasting the start of diabetes. Early models were created utilizing methods of statistical learning that are widely used [7]. These techniques allow for the forecasting of potential occurrences by analyzing historical training data for relevant trends [8, 9]. Based on the number of citations, publications, and connections with the most fruitful nations, institutions, and journals, deep learning applications in diabetes have been described in this paper. Moreover, Diabetes classification, Diabetes prediction and Real-time healthcare data analysis have been analyzed.

2. Research Methodology

A quantitative research model with a bibliometric approach to analyze the impact of deep learning in diabetes research has been conducted. The data for bibliometric analysis are retrieved from two of the best-known and leading libraries, Web of Science (WoS) and Scopus, by merging them into one file with the help of the R programming language, removing duplicates, and restricting the search to English-language documents published between 2018 and 2022. The search has been performed based on the title, abstract, and keywords. This review follows the guidelines for conducting systematic literature reviews. By Kithcenham and charters [10] and bibliometric guideline by [11], and the R bibliometrix

package by [6] for reports have been utilized. A quick overview of the entire research field can be obtained, and knowledge voids and gaps can be described using bibliometrics.

2.1 Data Collection and processing

Data cleaning is an essential step. The search was performed on December 10, 2022, and a total of 134 search strings composed of two keywords were defined, where the first keyword was the term synonyms of "Diabetes," and the second keyword corresponded with deep learning as listed in the ACM Computing Classification System (CCS) [12]. Figures 1, 2, and 3 depict the process with the refinement as, document type: article, language: English, period: 2018-2022, and excluded in press articles.

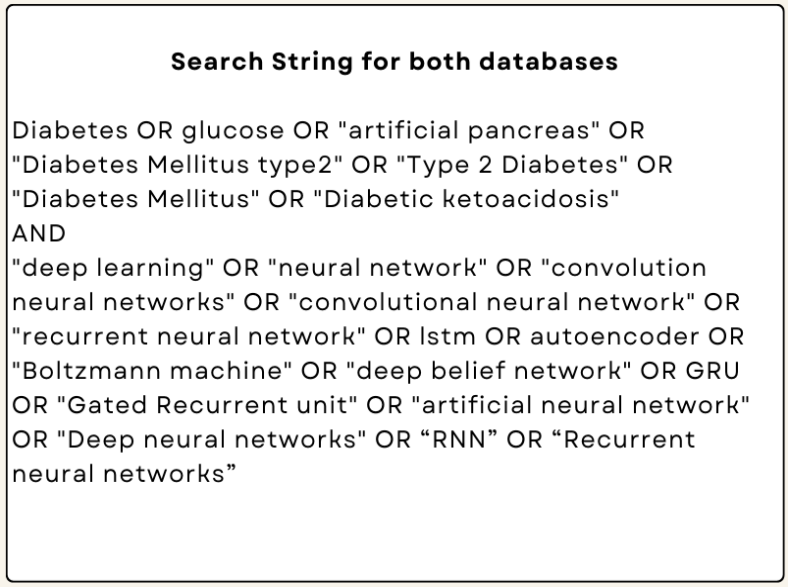


Figure 1. Search Strategy

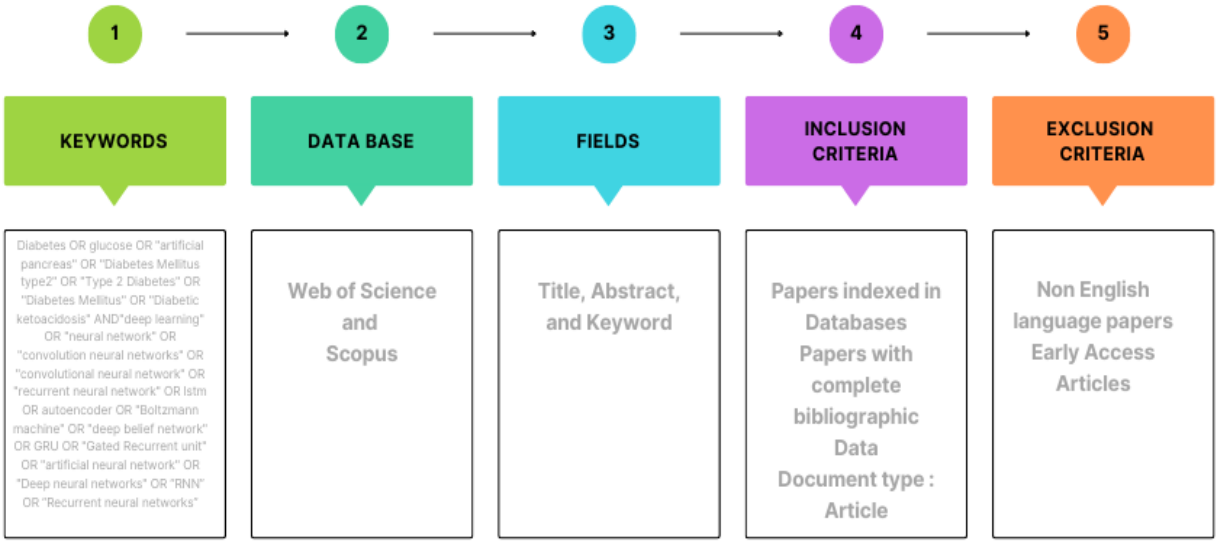


Figure 2. Step by Step data retrieval

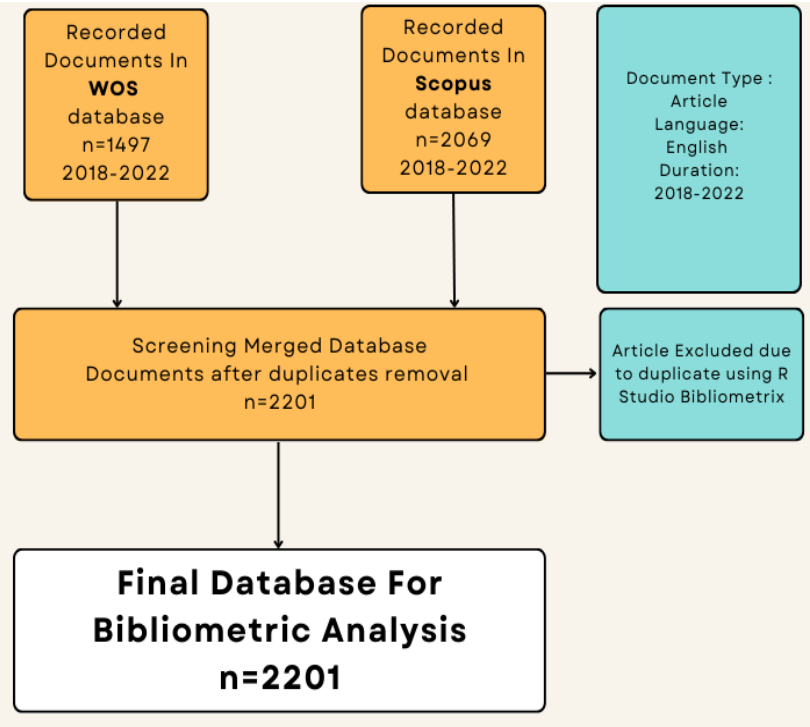


Figure 3. Data retrieval process

2.2 Research questions

The research questions formulated according to the objective of this systematic review are as follows:

RQ.1 How has the field of Deep Learning applications in diabetes research evolved between 2018-2022?

RQ.2 What is the average citation per doc?

RQ.3 What is the average citation per year?

RQ.4 Which scholars are the most locally cited in deep learning research in diabetes?

RQ.5 What are the most relevant sources?

RQ.6 Who are the most productive authors?

RQ.7 What are the most productive countries?

RQ.8 What are the most cited countries?

RQ.9 What are the top ten institutions in the field of diabetes research?

RQ.10 What are the most global cited documents?

RQ.11 What are the most local cited references?

RQ.12 What are the trend topics?

RQ.13 How are the co-occurrence and network approaches?

3. Bibliometric Analysis

3.1 Dataset overview

Table 1 shows the general information of the data.

Table 1. Data Overview

Description	Results
Main Information	
Timespan	2018:2022
Sources (Journals, Books, etc.)	918
Documents	2201
Annual Growth Rate %	61.43
Document Average Age	1.25
Average citations per doc	9.908 (RQ2)
References	84311
Document Contents	
Keywords Plus (ID)	13466
Author's Keywords (DE)	5134
Authors	
Authors	9432
Authors of single-authored docs	55
Authors Collaboration	
Single-authored docs	56
Co-Authors per doc	6.37
International co-authorships %	2.862
Document Types	
Article	2198

3.2 Annual Scientific Production

The growth of the journal articles by years is depicted in Fig. 4. It can be seen from Fig. 5 that the number of papers about research fields has continuously increased from 2018 to 2022. This shows that interest in this study area has been growing over the last several years. (**RQ1)** Number of articles per year are, 2018=109, 2019=279, 2020=417, 2021=648, and 2022=747.

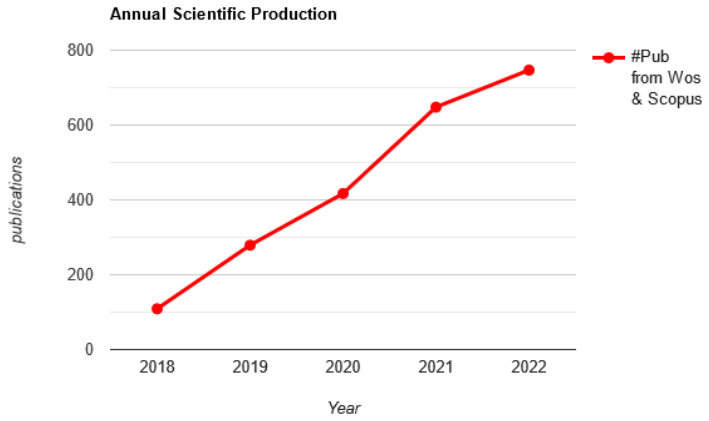
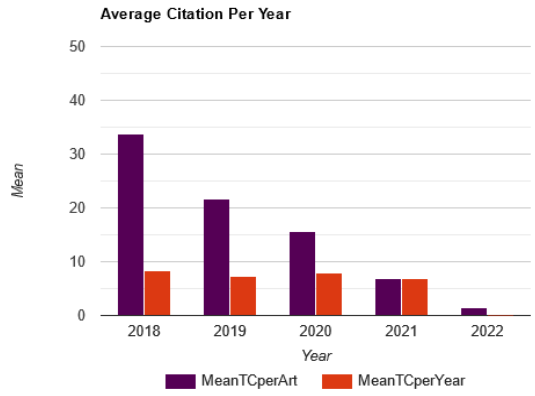


Figure 4. Article publication per year 2018-2022

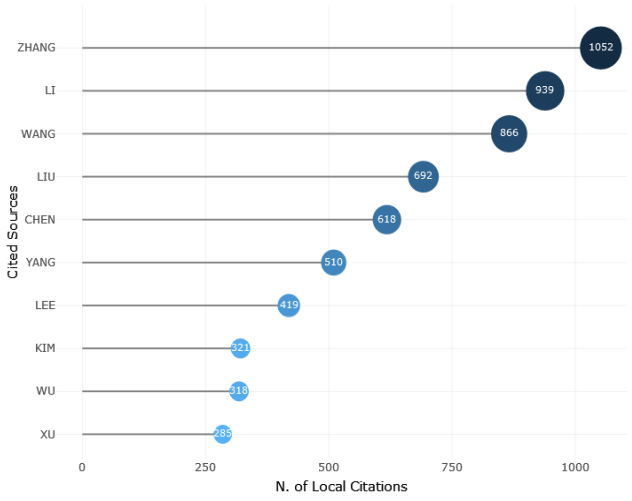
Table 2 shows average citations per year (**RQ3)**. Table 3 shows the top ten most locally cited documents. Zhang, Li, and Wang are the three best ranks (**RQ4).**

Table 2. Average Citation per Year



Year	N	MeanTC per Article	Mean TC per Year	Citable Years
2018	109	33.65	8.41	4
2019	279	21.59	7.2	3
2020	417	15.67	7.84	2
2021	648	6.93	6.93	1
2022	747	1.41	8.41	0

Table 3. Most Local Cited Sources (from Reference Lists)



Sources	Citation
ZHANG	1052
LI	939
WANG	866
LIU	692
CHEN	618
YANG	510
LEE	419
KIM	321
WU	318
XU	285

3.3 Relevant Sources

Figure 5 shows the top ten journals that published the most papers on deep learning in diabetes research. In total, 2201 journal articles about the research field were published in the WoS and Scopus databases between 2018 and 2022, with 356 papers published in the top ten most productive journals. **(RQ5)**

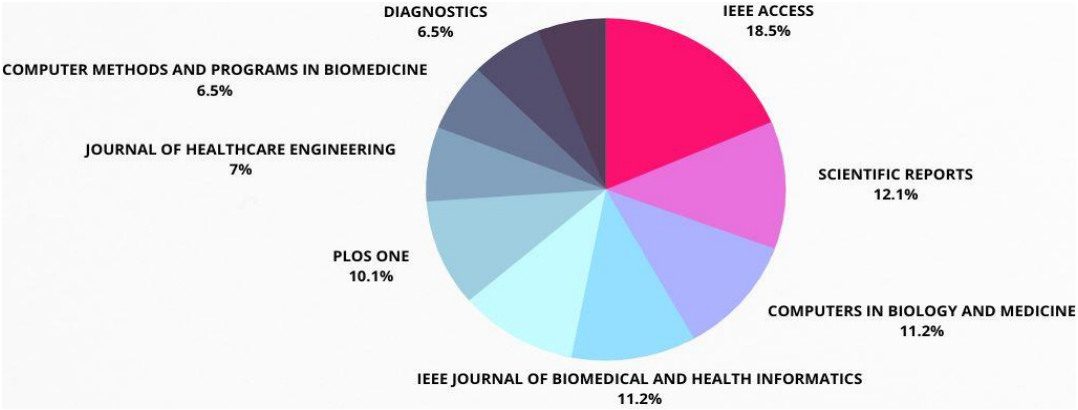


Figure 5. Top ten journal percentage in publication

IEEE Access, Scientific reports, and Computers in biology and medicine are the most locally cited sources as shown in table 4.

Table 4. Most Local Cited Sources (from Reference Lists)

Sources	Articles
IEEE ACCESS	66
SCIENTIFIC REPORTS	43
COMPUTERS IN BIOLOGY AND MEDICINE	40
IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS	40
SENSORS	38
PLOS ONE	36
JOURNAL OF HEALTHCARE ENGINEERING	25
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE	23
DIAGNOSTICS	23
APPLIED SCIENCES (SWITZERLAND)	22

3.4 Productive Authors

Figure 6 depicts the top ten most productive authors who have the most publications in diabetes deep learning research. Wang Y ranked first with 56 publications in this field, Wang J ranked second with 48 publications, and Liu Y ranked third with 48 publications. Fig. 7 shows the author's production over time. (**RQ6**)

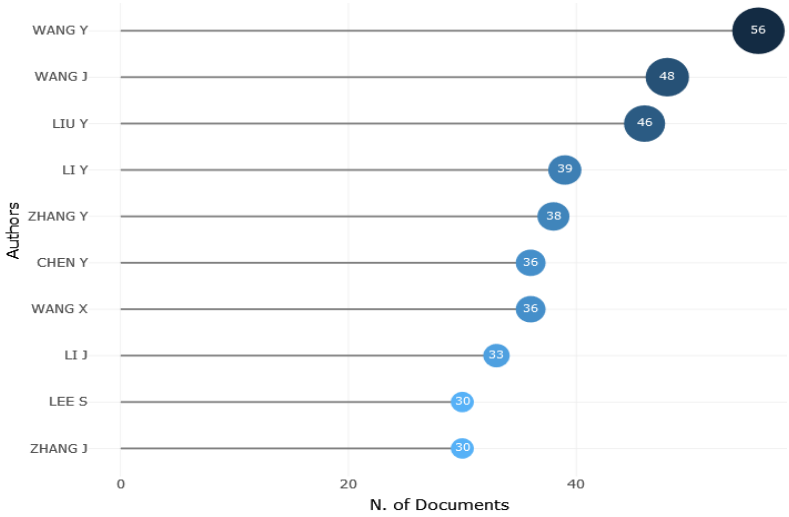


Figure 6. Most Productive Authors

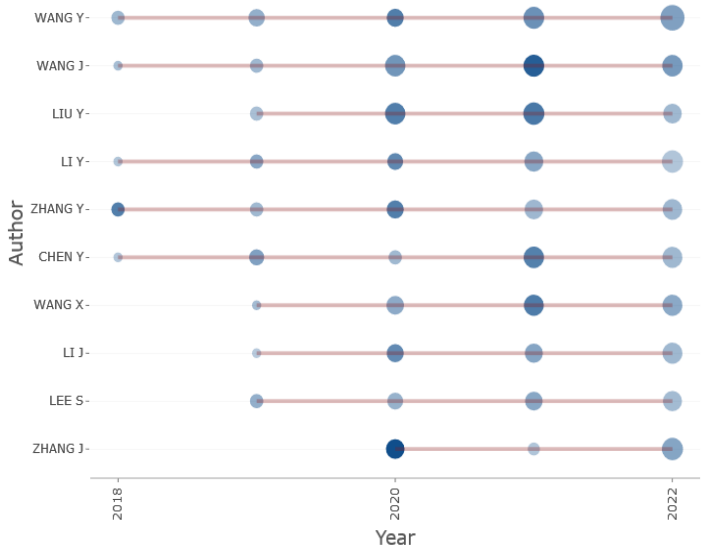


Figure 7. Authors' Production over Time

Table.5 shows the top 10 most productive countries in deep learning-aimed research in diabetes according to the number of publications. The first most productive country is China with 482 publications, and the United States comes next with a total of 287 publications. (**RQ7**)

Table 5. Corresponding Authors' Country

Country	Articles	SCP	MCP	Freq	MCP_Ratio
CHINA	482	472	10	0.219	0.021
USA	287	271	16	0.13	0.056
INDIA	246	244	2	0.112	0.008
KOREA	110	110	0	0.05	0
UNITED KINGDOM	68	62	6	0.031	0.088
IRAN	49	48	1	0.022	0.02
JAPAN	46	42	4	0.021	0.087

SAUDI ARABIA	40	39	1	0.018	0.025
ITALY	39	37	2	0.018	0.051

3.5 Most Cited Countries

The most-cited countries are tabulated in Table 6. China comes first with a total citation of 4256 and an average article citation of 8.83, and the United States comes second. Singapore has the best average document citation score (33.8). **(RQ8)**

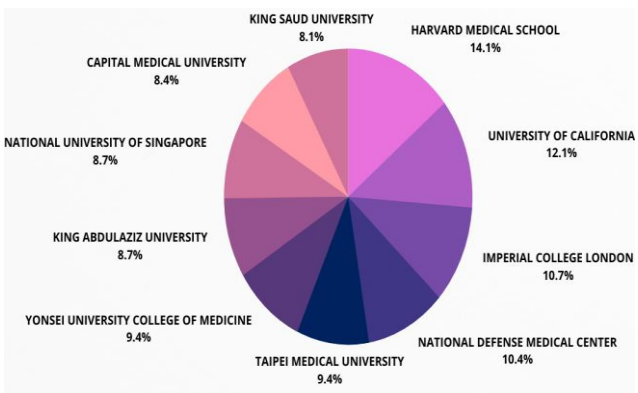
Table 6. Most Cited Countries

Country	Total Citation	Average Article Citations
CHINA	4256	8.83
USA	4074	14.2
INDIA	2061	8.38
UNITED KINGDOM	1105	16.25
KOREA	1019	9.26
SINGAPORE	507	33.8
ITALY	438	11.23
IRAN	420	8.57
CANADA	406	14
PAKISTAN	390	13.45

3.6 Relevant Affiliations

The affiliations of the top ten institutions in the field of diabetes research with deep learning are shown in Table 7. Harvard Medical School and the University of California, located in the United States, which ranks second among the most productive countries, are at the top of this list. **(RQ9)**

Table 7. Most Relevant Affiliations



Affiliation	Articles
HARVARD MEDICAL SCHOOL	42
UNIVERSITY OF CALIFORNIA	36
IMPERIAL COLLEGE LONDON	32
NATIONAL DEFENSE MEDICAL CENTER	31
TAIPEI MEDICAL UNIVERSITY	28
YONSEI UNIVERSITY COLLEGE OF MEDICINE	28
KING ABDULAZIZ UNIVERSITY	26
NATIONAL UNIVERSITY OF SINGAPORE	26
CAPITAL MEDICAL UNIVERSITY	25

3.7 Cited Documents

3.7.1 Most Global Cited Documents

Table 8 shows that Tunyasuvunakool K, 2021, in NATURE, is the most globally cited document, with 584 total citations and 292 per year. **(RQ10)**

Table 8. Most Global Cited Documents

Paper	Ref.	Total Citations	TC per Year	Normalized TC
TUNYASUVUNAKOOL K, 2021, NATURE	[13]	584	292	84.25
JIN Q, 2019, KNOWL BASED SYST	[14]	359	89.75	16.63
ZOU Q, 2018, FRONT GENET	[15]	318	63.6	9.45
WAN S, 2018, COMPUT ELECTR ENG	[16]	217	43.4	6.45
KUMAR PM, 2018, FUTURE GENER COMPUT SYST	[17]	206	41.2	6.12
LU D, 2018, SCI REP	[18]	176	35.2	5.23
ALAM KMR, 2020, NEURAL COMPUT APPL	[18]	164	54.67	10.46
HEIDARI AA, 2019, SOFT COMPUT	[19]	159	39.75	7.36
GADEKALLU TR, 2020, ELECTRONICS (SWITZERLAND)	[20]	157	52.33	10.02
JAJARMI A, 2019, CHAOS	[21]	136	34	6.3

3.7.2 Most Local Cited References

Automated identification of Diabetic retinopathy using deep learning [22], is the most frequently cited reference in the research field based on Table 9. **(RQ11)**

Table 9. Most Local Cited References

Ref.	Cited References	No.
[22]	Automated Identification of Diabetic Retinopathy Using Deep Learning	1
[23]	Exploring Deep Reinforcement Learning with Multi Q-Learning	2
[24]	Random Forests	3
[23]	Deep learning	4
[25]	Automated Diabetic Retinopathy Detection Based on Binocular Siamese-Like Convolutional Neural Network	5
[26]	Deep Residual Learning for Image Recognition	6
[27]	Very Deep Convolutional Networks for Large-Scale Image Recognition	7
[28]	Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs	8
[26]	Deep Residual Learning for Image Recognition	9

3.8 Trend Topics

Bibliometric analysis of the academic literature is in terms of the subject distributions of the academic material. This investigation provides information on the publishing patterns of authors who acknowledge that their work, as well as the phrases they use in their work are

being studied bibliometrically. Fig. 8 depicts the trending topics. Research trends analysis uses bibliometric techniques to analyze a specific topic. The findings of this bibliometric trend topic study can be useful for researchers. According to the bibliometric analysis, the trending research topics (accuracy, atrial fibrillation, heart infarction, human, article, and diabetes mellitus) have become important after 2020. **(RQ12)**

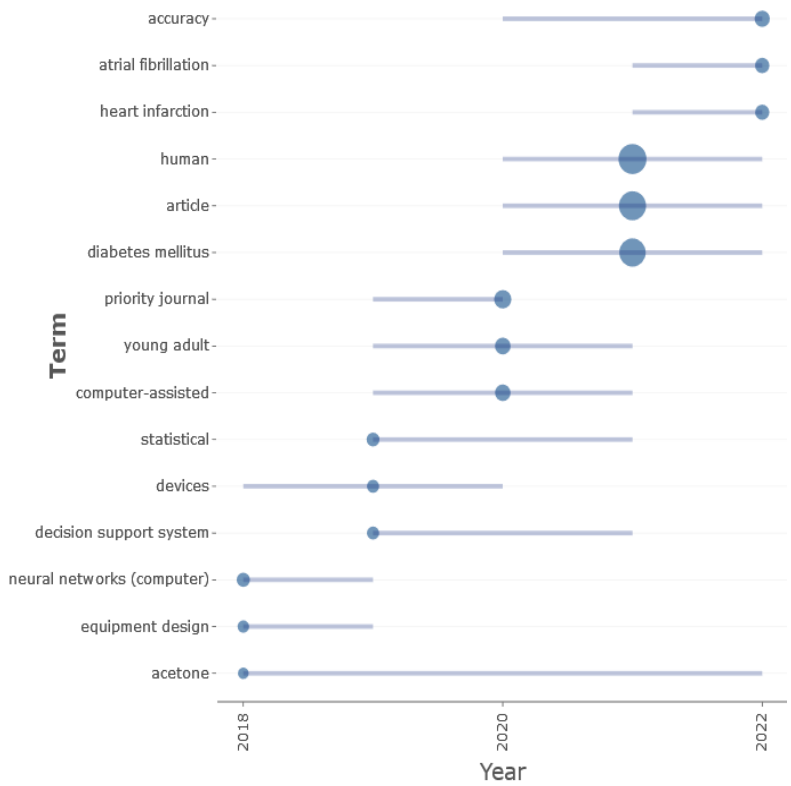


Figure 8. Trend Topics

3.9 Network Approaches

A co-occurrence network is an undirected graph constructed using a corpus of documents as its starting point. Each node in the graph represents a unique word in a vocabulary, and each edge represents the frequency with which two words co-occur in a document. In order to show the links between words in a corpus of texts and to extract information about those associations, co-occurrence network is used. For instance, a co-occurrence network can be used to figure 9 out the terms that are most often found along with a specific word. **(RQ13)**

Table 10. Co-Word Network Analysis

Node	Cluster	Betweenness	Closeness	PageRank
Deep Learning	1	0.521442	0.020408	0.028586
Diabetic Retinopathy	1	0.129958	0.020408	0.015171

Neural Networks	1	0.167131	0.020408	0.02125
Convolutional Neural Network	1	0.122686	0.020408	0.016458
Computer	1	0.104009	0.020408	0.018646
Diagnosis	1	0.080997	0.020408	0.011812
Forecasting	1	0.007109	0.02	0.010212
Diagnostic Imaging	1	0.080981	0.020408	0.015617
Algorithms	1	0.054773	0.020408	0.014113

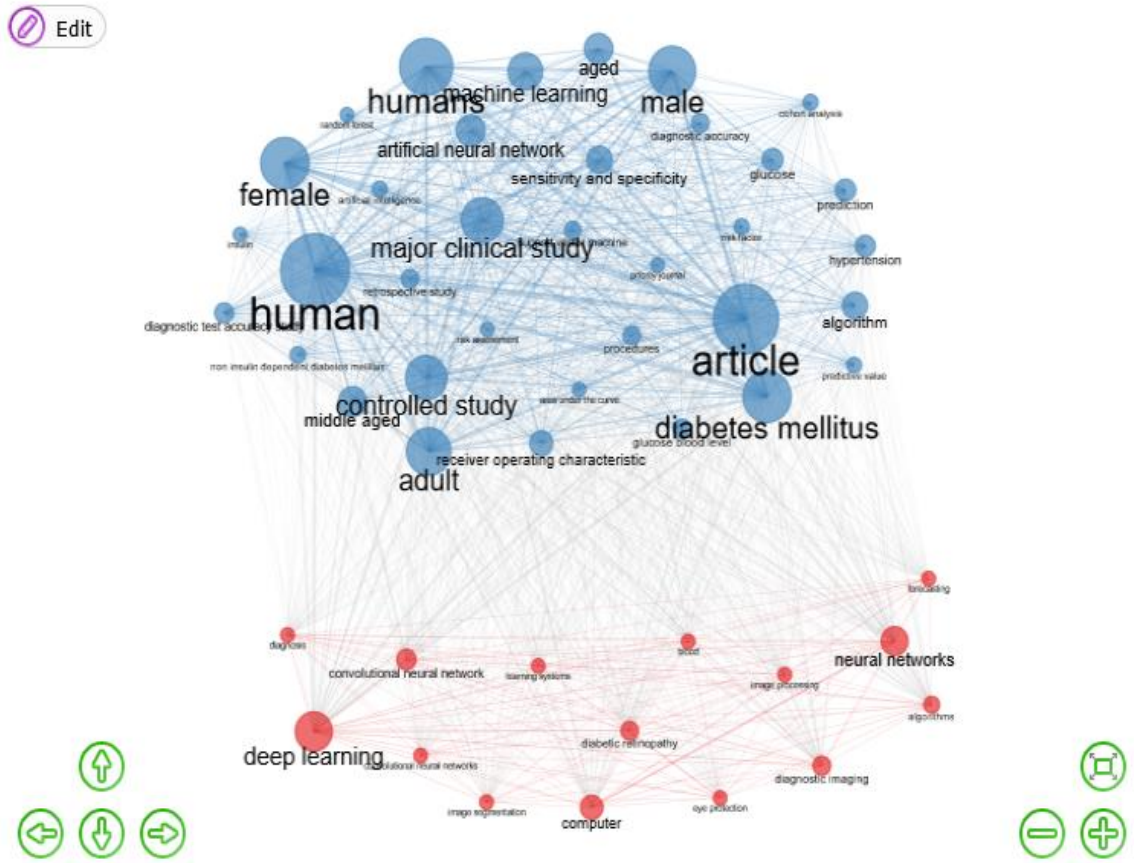


Figure 9. Co-occurrence Network

4. Conclusion and Limitation

A review of journal articles spanning 2018–2022, culled from the Scopus and Web of Science databases, provides an extensive data research landscape. In recent years, deep learning methods have been widely used to enhance all aspects of science. In this study, bibliometric data taken from these databases are comprehensively evaluated, and deep learning applications in diabetes research have been investigated. A wide range of groups, including academics, educators, and research institutions, give this research much attention. This paper uses Bibliometrix to perform bibliometric analysis which focuses on deep learning in diabetes-related journal articles. In this research, "human" and "deep learning" are the most

frequent words. The study reveals that China is the most productive and influential country. Deep learning has the greatest betweenness, with a betweenness of 0.52 in co-word network analysis. Word Accuracy, Atrial Fibrillation, and Heart Infection are the most trending topics by 2022. In 2018, this research field saw a significant increase in publications. Tunyasuvunakool K, (2021) in "Nature" is the most globally cited paper, with 584 total citations and 292 total citations per year. Harvard Medical School is the most productive institute, with 42 articles, and the University of California follows with 36. The most cited country is China, with 4256 total citations, and Singapore is leading in average article citation criteria (33.8). Wang Y is the most productive author with 56 publications, and IEEE Access is the most productive journal. Zhang is the most cited author in this research field, with a total citation of 1052, an annual growth rate of 64.43, and an average citation per document of 9. This study fills the gap in the literature by conducting a bibliometric analysis of papers dealing with deep learning in the study of diabetes.

Only scientific papers published in the Scopus and WoS databases were analyzed. The articles such as conference proceedings papers, non-English articles, and non-Scopus and WoS index papers have also published in this field which need to be analyzed. The primary study subjects can be categorized via the co-occurrence network analysis. The crucial aspect is that the interest in science has kept expanding. Future research should take a broader approach and incorporate more information from different databases to increase the impact of the analysis. Researchers, practitioners, and accounting experts who are interested in the use of deep learning in diabetes research may find the result of this research result to be a helpful resource.

References

[1] Williams, R., et al., Global and regional estimates and projections of diabetes-related health expenditure: Results from the International Diabetes Federation Diabetes Atlas. Diabetes research and clinical practice, 2020. 162: p. 108072.

[2] Butt, U.M., et al., Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications. Journal of Healthcare Engineering, 2021. 2021: p. 9930985.

[3] Zhu, T., et al., Deep Learning for Diabetes: A Systematic Review. IEEE J Biomed Health Inform, 2021. 25(7): p. 2744-2757.

[4] NIH. What is Diabetes? ; Available from: https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes.

[5] NHS. Digital innovations in diabetes. Available from: https://www.england.nhs.uk/diabetes/digital-innovations-to-support-diabetes-outcomes/.

[6] Aria, M. and C. Cuccurullo, bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 2017. 11(4): p. 959-975.

[7] Parlina, A., K. Ramli, and H. Murfi Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database. Information, 2020. 11, DOI: 10.3390/info11020069.

[8] Fregoso-Aparicio, L., et al., Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. Diabetology & Metabolic Syndrome, 2021. 13(1): p. 148.

[9] Rejeb, A., et al., The big picture on Instagram research: Insights from a bibliometric analysis. Telematics and Informatics, 2022. 73: p. 101876.

[10] Kitchenham, B. and S. Charters, Guidelines for performing systematic literature reviews in software engineering. 2007.

[11] Donthu, N., et al., How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 2021. 133: p. 285-296.

[12] System, A.C.C.; Available from: https://dl.acm.org/ccs.

[13] Tunyasuvunakool, K., et al., Highly accurate protein structure prediction for the human proteome. Nature, 2021. 596(7873): p. 590-596.

[14] Jin, Q., et al., DUNet: A deformable network for retinal vessel segmentation. Knowledge-Based Systems, 2019. 178: p. 149-162.

[15] Zou, Q., et al., Predicting Diabetes Mellitus With Machine Learning Techniques. Frontiers in Genetics, 2018. 9.

[16] Wan, S., Y. Liang, and Y. Zhang, Deep convolutional neural networks for diabetic retinopathy detection by image classification. Computers & Electrical Engineering, 2018. 72: p. 274-282.

[17] Kumar, P.M., et al., Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. Future Generation Computer Systems, 2018. 86: p. 527-534.

[18] Lu, D., et al., Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimerâ€™s Disease using structural MR and FDG-PET images. Scientific Reports, 2018. 8(1): p. 5697.

[19] Heidari, A.A., et al., An efficient hybrid multilayer perceptron neural network with grasshopper optimization. Soft Computing, 2019. 23(17): p. 7941-7958.

[20] Gadekallu, T.R., et al. Early Detection of Diabetic Retinopathy Using PCA-Firefly Based Deep Learning Model. Electronics, 2020. 9, DOI: 10.3390/electronics9020274.

[21] Jajarmi, A., B. Ghanbari, and D. Baleanu, A new and efficient numerical method for the fractional modeling and optimal control of diabetes and tuberculosis co-existence. Chaos, 2019. 29(9): p. 093111.

[22] Gargeya, R. and T. Leng, Automated Identification of Diabetic Retinopathy Using Deep Learning. Ophthalmology, 2017. 124(7): p. 962-969.

[23] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. Nature, 2015. 521(7553): p. 436-444.

[24] Breiman, L., Random Forests. Machine Learning, 2001. 45(1): p. 5-32.

[25] Zeng, X., et al., Automated Diabetic Retinopathy Detection Based on Binocular Siamese-Like Convolutional Neural Network. IEEE Access, 2019. 7: p. 30744-30753.

[26] He, K., et al. Deep Residual Learning for Image Recognition. in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

[27] Simonyan, K. and A. Zisserman Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014. arXiv:1409.1556.

[28] Gulshan, V., et al., Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. Jama, 2016. 316(22): p. 2402-2410.