Heliyon 10 (2024) e24230

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Knowledge structure and global trends of machine learning in stroke over the past decade: A scientometric analysis

Mingfen Wu, Kefu Yu, Zhigang Zhao **, Bin Zhu *

Department of Pharmacy, Beijing Tiantan Hospital, Capital Medical University, Beijing, 100070, China

ARTICLE INFO

Keywords:
Machine learning
Stroke
Deep learning
Global trends
Algorithm

ABSTRACT

Objective: Machine learning (ML) models have been widely applied in stroke prediction, diagnosis, treatment, and prognosis assessment. We aimed to conduct a comprehensive scientometrics analysis of studies related to ML in stroke and reveal its current status, knowledge structure, and global trends.

Methods: All documents related to ML in stroke were retrieved from the Web of Science database on March 15, 2023. We refined the documents by including only original articles and reviews in the English language. The literature published over the past decade was imported into scientometrics software for influence detection and collaborative network analysis.

Results: 2389 related publications were included. The annual publication outputs demonstrated explosive growth, with an average growth rate of 63.99 %. Among the 90 countries/regions involved, the United States (729 articles) and China (636 articles) were the most productive countries. Frontiers in Neurology was the most prolific journal with 94 articles. 234 highly cited articles, each with more than 31 citations, were detected. Keyword analysis revealed a total of 5333 keywords, with a predominant focus on the application of ML models in the early diagnosis, classification, and prediction of "acute ischemic stroke" and "atrial fibrillation-related stroke". The keyword "classification" had the first and longest burst, spanning from 2013 to 2018. 'Upport vector machine' got the strongest burst strength with 6.2. Keywords such as 'mechanical thrombectomy', 'expression', and 'prognosis' experienced bursts in 2022 and have continued to be prominent.

Conclusion: The applications of ML in stroke are increasingly diverse and extensive, with researchers showing growing interest over the past decade. However, the clinical application of ML in stroke is still in its early stages, and several limitations and challenges need to be addressed for its widespread adoption in clinical practice.

1. Introduction

Stroke is a collective term for a group of diseases that occur when blockage of blood flow to the brain or blood vessels suddenly rupture, leading to cerebral ischemia or resulting in damage to brain tissue. It can be categorized into ischemic stroke and hemorrhagic

* Corresponding author. Department of Pharmacy, Beijing Tiantan Hospital, Capital Medical University, No.119 South Fourth Ring West Road, Fengtai District, Beijing, 100070, China.

** Corresponding author. Department of Pharmacy, Beijing Tiantan Hospital, Capital Medical University, No.119 South Fourth Ring West Road, Fengtai District, Beijing, 100070, China.

E-mail addresses: 1022zzg@sina.com (Z. Zhao), zbtcm@163.com (B. Zhu).

https://doi.org/10.1016/j.heliyon.2024.e24230
Received 18 April 2023; Received in revised form 23 November 2023; Accepted 4 January 2024
Available online 12 January 2024

stroke according to pathology. In today's world, stroke continues to be a major public health concern, causing significant mortality, disability, and imposing a substantial economic burden. According to statistics from the World Health Organization, the global annual incidence of stroke is approximately 16 million, with a mortality rate as high as 6.5 %, accounting for approximately 1.1 million deaths [1]. Stroke has become the second leading cause of death after ischemic heart disease and the third leading cause of disability-adjusted life years (DALYs) lost globally [2]. According to the 2013 policy statement of the American Heart Association, the total annual stroke-related medical cost in the United States will reach $240.67 billion by 2030 [3]. For ischemic stroke, acute treatment is highly dependent on timely diagnosis. According to the current guidelines for the treatment of ischemic stroke, there is a strict time window for the treatment of ischemic stroke, and patients can be treated with intravenous thrombolysis within 4.5 h after the onset of symptoms [4]. For hemorrhagic stroke, timely diagnosis and assessment of the type and cause of hemorrhage using imaging techniques are essential to guide acute treatment decisions. Therefore, timely diagnosis, urgent treatment decisions, and accurate prediction are the cornerstones of acute stroke management.

Machine learning (ML) is a branch of artificial intelligence (AI) that utilizes statistical algorithms to learn from extensive historical data, generating empirical models to guide various tasks [5]. The clinical information accumulated by medical practice is vast and intricate, making it challenging for researchers to efficiently determine the predictive relevance of this multidimensional medical data for clinical decision-making. The advances in ML have presented opportunities to harness these massive medical datasets to inform medical practice across various domains. In recent years, ML models have been widely used to solve various complex challenges in stroke, such as early stroke detection and thrombolysis decision-making [6,7] neuroimaging analysis [8,9], stroke diagnosis and severity assessment [10,11], candidate selection for therapeutic intervention [12,13], prediction of short- and long-term functional outcomes and prognosis [[14–17]]. Early detection of stroke is a crucial step in ensuring effective treatment and ML has demonstrated significant value in facilitating this process [18]. Numerous applications of ML/DL in stroke have been reported, such as brainwaves being investigated for stroke prediction [19], electroencephalography (EEG) signals utilized to develop explainable AI models for stroke prediction [20], EEG features used for quantitative evaluation of task-induced neurological outcome after stroke [21], electrocardiogram (ECG) used to identify atrial fibrillation (AF) related stroke [22], electromyography (EMG) applied for prediction of myoelectric biomarkers in post-stroke gait [23], biosignals being investigated for stroke prediction [24], among others. Moreover, some ML models have been developed into automated applications for various clinical tasks, including identifying large vessel occlusions (LVOs), diagnosing ischemic and hemorrhagic stroke, and assessing salvageable brain tissue [25].

Based on the level of manual involvement required for categorization or labeling in the training corpus, ML can be categorized into four subtypes: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Deep learning (DL), which is based on neural networks, is a significant branch of ML [26]. Supervised learning stands in contrast to unsupervised learning, where we have observations of measurements but no associated responses, and the main aim of unsupervised learning approaches is to understand relationships between observations. The supervised learning algorithm refers to ML methods that have target variables or predict goals, including classification and regression, whose responses can be quantitative or qualitative. Fig. 1 shows the conceptual map of ML frameworks in stroke.

Most of the ML models used in stroke studies predominantly fall under the category of supervised learning, and the commonly employed algorithms for supervised learning encompass linear/logistic regression (LR), support vector machine (SVM), decision trees (DT), random forests (RF), k-nearest neighbor (kNN), and deep neural networks (DNN), among others. Classic examples of supervised learning algorithms extensively utilized in stroke studies include DT, SVM, DNN, RF, and LR. For instance, a DL model based on the

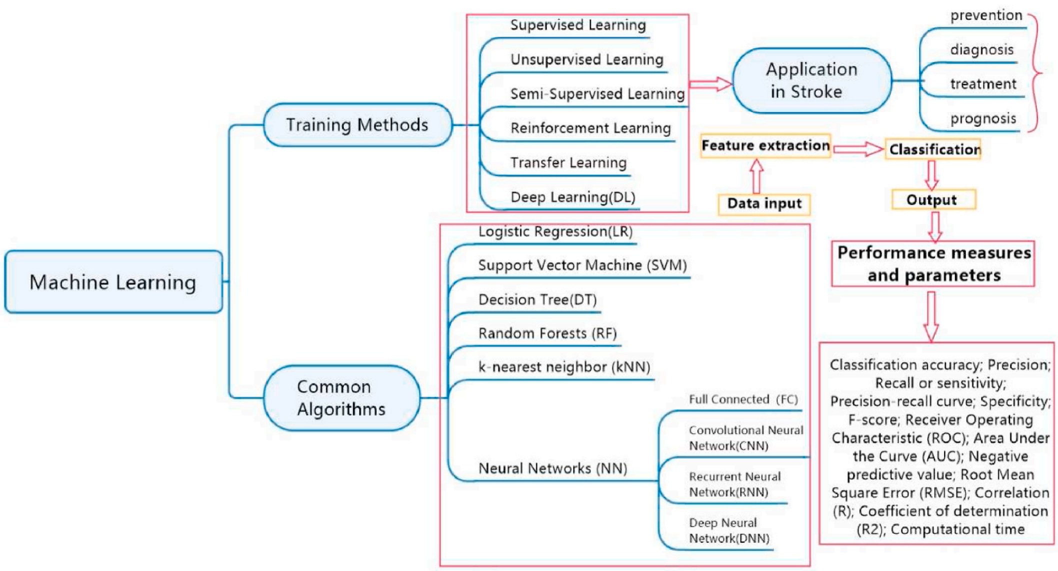

Fig. 1. The conceptual map of ML frameworks in stroke.

DNN algorithm has been trained to identify patients at risk for AF-related strokes [22]. The prediction of large vessel occlusion for ischemic stroke has been accomplished using the RF algorithm in an ML model [12]. Additionally, an ML model constructed using the SVM algorithm has been trained for stroke prediction [24]. Moreover, in a study predicting aneurysmal subarachnoid hemorrhage, seven ML models, including LR, SVM, DT, and DNN, were employed [27]. Each of these algorithms possesses its distinct strengths and weaknesses. Typically, a study necessitates training multiple models using various algorithms and subsequently selecting the optimal model through performance testing. When choosing an ML algorithm, factors such as data type, problem type, data volume and dimension, interpretability, reliability, algorithm complexity, and computing resources need to be taken into consideration.

Driven by AI and computer technology, ML has piqued the interest of global scholars, and the number of related publications has increased rapidly in recent years. Further systematic research on the knowledge structure, research hotspots, and frontiers of ML in stroke will help researchers to have a comprehensive understanding of this field. But as far as we know, no previous study has specifically analyzed the overview and global trends in the field of ML in stroke yet. Thus, this study is the first attempt to fill this research gap.

Scientometric analysis is a technique that can provide a macro perspective of a large number of academic literature. It can map the scientific development of a given research field by using text-mining technology for quantitative and qualitative analysis. In this way, countries, institutions, authors, and journals with great influence in the area can be identified, and research hotspots and development frontiers can also be detected. Our study aims to overview the research status of ML in stroke, map the knowledge structure, identify key influences, analyze the cooperation network, and identify the main research hotspots and frontiers.

1.1. Contribution of this study

- We conducted a comprehensive analysis of ML studies in stroke, summarizing the current research status. We identified the major countries, prolific institutions, active authors, and collaborative networks, and explored the main research trends and hotspots in this field.
- Our findings indicate that ML in stroke is a rapidly growing research area, with increasing attention from researchers over the past decade. The application of ML in the early diagnosis, classification, and prediction of 'acute ischemic stroke' and 'AF-related stroke' are particularly prominent research hotspots.
- Despite the rapid evolution of machine learning applications in stroke, its clinical implementation is still in its early stages and faces certain challenges and limitations.
- These findings serve as a valuable starting point, providing useful insights into future research directions and prospects in this rapidly evolving discipline.

2. Materials and methods

2.1. Data collection

The Web of Science Core Collection (WoSCC) was selected as the main data source. All documents related to ML in stroke were retrieved from the Science Citation Index Expanded (SCI-Expanded) of WoSCC on March 15, 2023. The period was set to 10 years. To more focus on this topic, we combined the MeSH terms and their entry terms in the literature search. The details of the retrieval strategies and refinement procedure are provided in Table 1.

Inclusion criteria: The publications were related to ML studies in stroke, the document types were original articles and reviews, the publication language was English, and the published time was from January 1, 2013, to March 15, 2023.

Exclusion criteria: non-English language papers, publications other than original articles and reviews, and publications published before 2013 were excluded.

The full records and cited references of the selected documents were downloaded from the WoSCC database, and then the data files were imported into the scientometrics tools of CiteSpace and VOSviewer for visualization analysis. The Journal Impact Factor (IF) was

Table 1
Details of the retrieval strategies and refinement procedure.

Search Number	Retrieval strategies and refinement procedure	Results
#1	TS = ("stroke*" OR "cerebrovascular accident*" OR "cerebrovascular apoplexy" OR "brain vascular accident*" OR "cerebral infarction*" OR "cerebral infarct*" OR "brain infarction*" OR "cerebral thrombosis" OR "cerebral embolism" OR "brain thrombosis" OR "brain embolism" OR "brain ischemia" OR "cerebral ischemia" OR "cerebral vascular occlusion" OR "cerebrovascular occlusion" OR "cerebral hemorrhage" OR "intracerebral hemorrhage" OR "intracranial hemorrhage" OR "brain hemorrhage" OR "subarachnoid hemorrhage" OR "encephalorrhagia" OR "hematencephalic")	537314
#2	TS = ("machine learning" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "adversarial learning")	279085
#3	#1 and #2	2628
#4	Time = 2013 to 2023	2584
#5	Article and review article (Excluding: online publication = 98, proceedings papers = 11, book chapters = 2, data papers = 1)	2393
#6	Publication language = English (Excluding: Spanish = 2, German = 1, Russian = 1)	2389

Note: TS = Topic, including title, abstract, author keywords and keywords plus; * = any ending to the word.

obtained from the Journal Citation Report (JCR) 2021. All data used in this study were downloaded from a public database, therefore, ethics committee approval or informed consent was not required.

2.2. Data analysis

CiteSpace (Version 6.1.R2), VOSviewer (Version 1.6.18), and the online analysis platform (http://bibliometric.com/) were used for visual analysis. Among the many scientometrics tools, Citespace and VOSviewer are the most widely used data analysis software. Citespace has advantages in data collection, data processing, visualization, and interpretation, while VOSviewer has the advantages of strong graphic display ability, can analyze large-scale data, and strong universality, which is suitable for multiple databases and data in various formats. Therefore, the combination of Citespace and VOSviewer in this study can better display the knowledge structure and development trend in the field of ML in stroke.

From publications that meet the criteria, key information such as publication years, number and distribution of publications, total citations (TC), average citations (AC) per item, countries/regions, organizations, authors, sources, references, titles, and keywords were extracted, and then, Microsoft Excel 2021 and visualization tools were used to quantify and visualize the above variables. Microsoft Excel 2021 was applied to analyze the publication trend and draw all trend charts, bar charts, stacked bar charts, and pie charts in the study. VOSviewer was used for the co-authorship analysis of countries/regions, authors, citation analysis for journals, and co-occurrence analysis of the keywords. CiteSpace was employed to analyze the cooperation network of countries/regions and institutions, the timeline clusters, and citation bursts of the keywords.

The parameters of CiteSpace were set as follows: the period is from 2013 to 2023, the time slice is set to 1 year, and the threshold (Top N) is set to 50, which means extracting the 50 nodes with the highest frequency every year.

The growth rate of publications was calculated by Equation (1) [28]:

$$\text{Growth rate} = \left(\sqrt[t_2-t_1]{N_2 \div N_1} \text{-}1 \right) \times 100 \tag{1}$$

t_1: First year; t_2: Last year; N_1: Number of publications in the first year; N_2: Number of publications in the last year.

3. Results

3.1. Analysis trends of publication outputs and citations

According to our screening criteria, a total of 2389 publications including 2198 articles and 191 reviews related to ML in stroke were identified for further analysis. The trends of publications and citations from 2013 to 2022 are plotted in Fig. 2 (2023 is excluded due to less than 1 year). Research on ML in stroke could be divided into two stages: the first stage (2013–2017) and the second stage (2018–2022). In the first stage, the number of publications increased gradually from 9 in 2013 to 52 in 2017, with an average publication volume of 28.6. The second stage showed explosive growth at a rapid linear rate, and the publications exceeded 100 in 2018 and reached 772 in 2022, with an average publication volume of 427.2. The second stage contributed 93.73 % (2136/2279) of all publications. The model fitting curve displayed that the annual publications had a significant exponential growth trend in the past ten years ($y = 5.2995e^{0.5168x}$, $R^2 = 0.9609$). The average growth rate from 2013 to 2022 was 63.99 %, indicating that the research of ML in stroke had received attention and the value of mining was getting higher and higher. Moreover, trends in publication citations were similar to the publication outputs, and both the most productive and cited year was 2022 with 772 publications and 12201 citations,

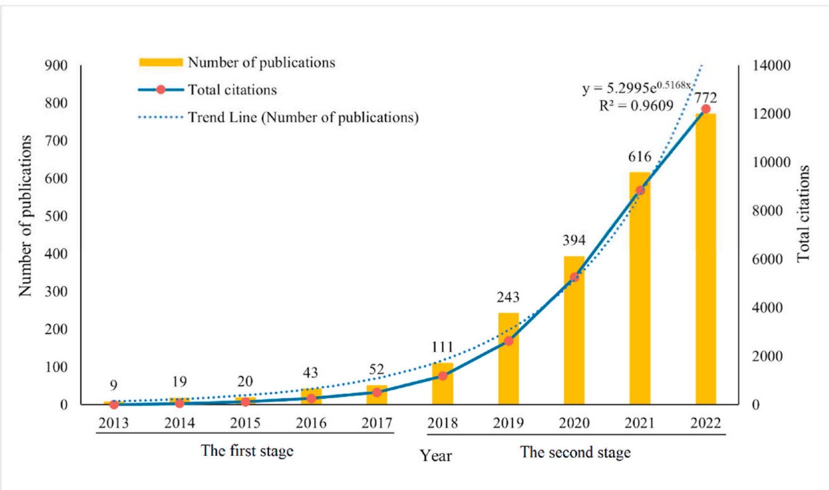

Fig. 2. Trend analysis of publication outputs and citations from 2013 to 2022.

respectively. The cumulative total number of citations of 2389 publications was 32,618, including 4787 records of self-citations by 15 March 2023, with an AC of 13.65 per article.

3.2. Analysis of major countries/regions and their cooperation

Based on WoSCC, there were 90 countries/regions that contributed to the field of ML in stroke over the past decade. The publications were distributed mainly concentrated in North America, East Asia, and Europe, as shown in Fig. 3A. Among the 90 countries/regions, the USA published the most articles, reaching 729 records, followed by China (636 records), England (252 records), Germany (190 records), India (184 records), South Korea (173 records), Canada (144 records), Australia (119 records), Italy (113 records), and Japan (99 records), see Fig. 3B and Table 2. The total number of publications in the top 10 countries/regions was 2639, which was more than the total number of publications we included, indicating that some articles were completed by cooperation between multiple countries. Annual publications in the 10 most productive countries/regions from 2013 to 2022 are shown in Fig. 3C and Table S1. As can be seen, the number of annual publications in China has increased by nearly 100 in each of the past two years and surpassed the United States for the first time in 2021. Moreover, the number of publications in China was 274 in 2022, 80 more than in the United States (194 publications), demonstrating that China is an active and rapidly developing country in this field, and the attention of Chinese scholars on this topic continues to surge.

Furthermore, Fig. 4A, B, and 4C display the cooperation relationship, co-occurrence network, and co-authorship overlay

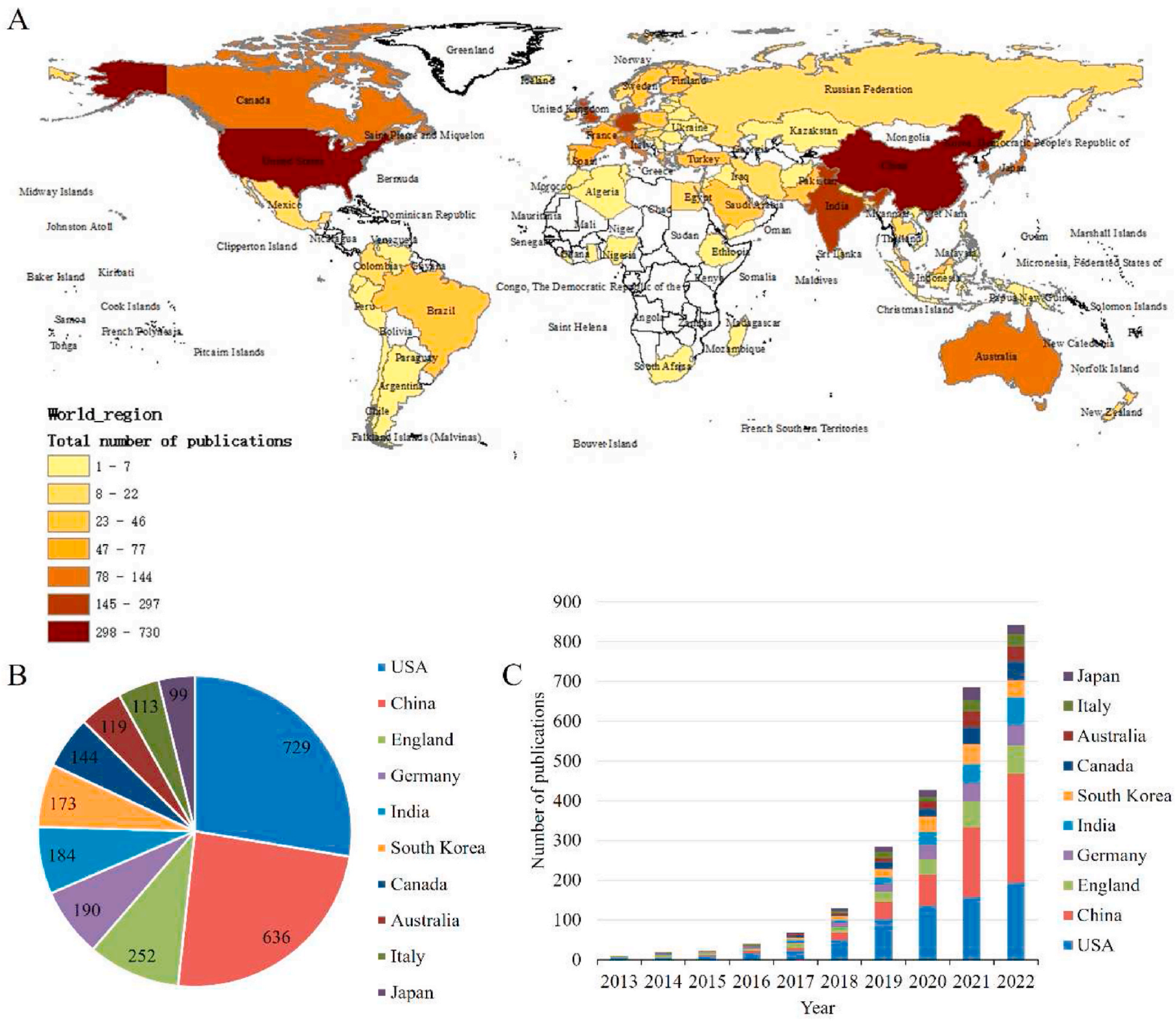

Fig. 3. (A) Geographic distribution of publications related to ML in stroke (The darker the color, the greater the number of publications); (B) Number of publications in the top 10 most productive countries/regions; (C) Annual publications in the top 10 most productive countries/regions from 2013 to 2022. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Top 10 productive countries/regions contributed to publications on ML associated with stroke research.

Rank	Country	Output [n (%)]	TC	AC	H-index	Centrality	TLS
1	USA	729 (30.51)	14216	19.50	53	0.04	832
2	China	636 (26.62)	6139	9.65	35	0.17	274
3	England	252 (10.55)	6237	24.75	37	0.10	588
4	Germany	190 (7.95)	2735	14.39	29	0.42	333
5	India	184 (7.70)	2630	14.29	29	0.17	375
6	South Korea	173 (7.24)	2115	12.23	25	0.21	82
7	Canada	144 (6.03)	1607	11.16	20	0.13	349
8	Australia	119 (4.98)	1213	10.19	18	0.13	236
9	Italy	113 (4.73)	2448	21.66	28	0.91	303
10	Japan	99 (4.14)	1267	12.80	19	0.00	137

Note: TC, total citations; AC, average citations per document; TLS, total link strength.

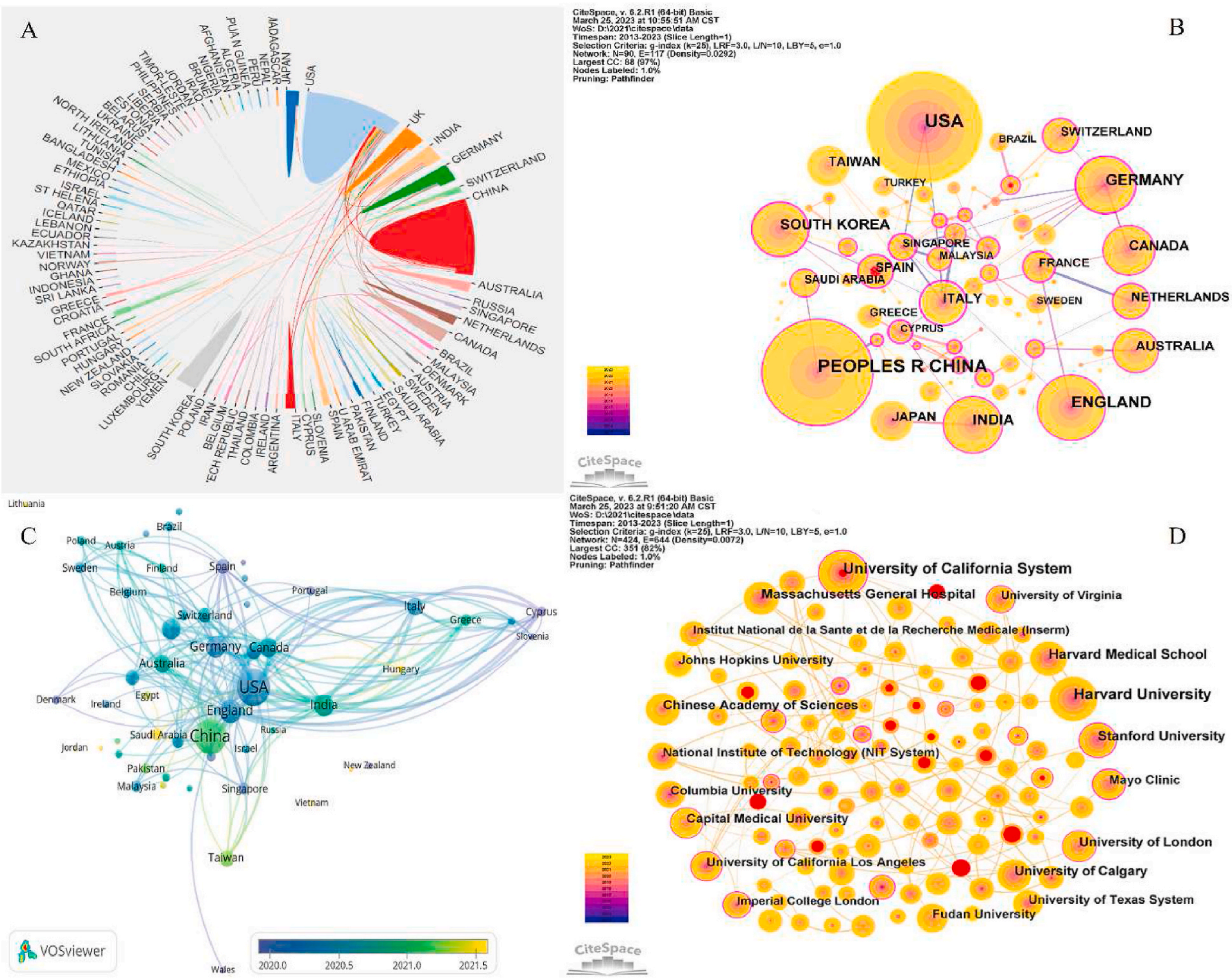

Fig. 4. (A) Map of cooperation between countries/regions contributed to publications in the field of ML in stroke using bibliometrics online platform; (B) Co-occurrence network map of countries/regions by CiteSpace (Each node represents a country/region, and the larger the node, the more papers; the nodes with higher centrality (>0.1) are highlighted with purple rings; the lines between nodes represent cooperative relationships of the two countries/regions, the thicker the line, the closer the cooperation); (C) Overlay visualization map of countries/regions by VOSviewer (The color of the node indicates the average publication time of a country/region); (D) Co-occurrence network map of institutions contributed to publications in the field of ML in stroke by CiteSpace (Each node represents an institution, other definitions is same to Fig. 3B). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

visualization of countries/regions using the bibliometrics online platform, CiteSpace, and VOSviewer, respectively. Fig. 4B and **C** specifically highlight the 47 countries/regions that have published more than 10 articles in this field. The purple ring surrounding each node represents a higher centrality, indicating its key role in the connected nodes. Notably, the United States has the largest node, signifying its significant contribution to ML in stroke. It ranks first in the number of publications (729 records), total citations (14216 times), H-index (53), and total link strength (TLS, 832), as shown in Table 2. England ranked third in the number of articles (252 records) but took the lead in terms of AC (24.75 times), indicating the greater influence of England papers. In terms of the centrality of published papers at each node, Italy held the highest value at 0.91, followed by Germany (0.42), Korea (0.21), and the United States (0.04). This suggests that despite being prolific, the academic communication capacity and influence of the United States are relatively weak, and there is a need to strengthen academic exchanges and cooperation among high-yield countries (nodes with centrality <0.1 indicate relatively weak connections and influence within the network).

3.3. Analysis of major institutions and their cooperation

Of these 2389 documents analyzed, a total of 424 institutions have contributed to the research on ML in stroke. Among them, both the University of California System and Harvard University have emerged as the top contributors, with 91 records each (as shown in Table 3). The University of California system, with an average publication year of 2013, leads in terms of total citations (1792 times), centrality (0.36), and H-index (21). Notably, six out of the top 10 prolific institutions are from the USA, further solidifying its position as a leading country in this field. It is worth highlighting that the Mayo Clinic, with an average publication year of 2019, stands out with the highest average citations (36 times), indicating its recent and highly active involvement in conducting high-quality research in this area. Fig. 4D provides a visual representation of the collaboration networks among 117 institutions that have published more than 10 articles. The figure demonstrates active collaboration among these prolific institutions, particularly showcasing the close partnership between the University of California System, Harvard University, and Harvard Medical School.

3.4. Analysis of major journals and discipline distribution

The 2389 publications on ML in stroke refer to 675 journals, including 95 journals that published more than 5 articles. According to the WOS category, these journals are placed in 137 categories, of which the top five were Neurosciences (394 papers), Clinical Neurology (378 papers), Engineering Electrical Electronic (291 papers), Engineering Biomedical (273 papers), and Radiology Nuclear Medicine Medical Imaging (263 papers).

The overlay visualization of the journal co-citation network is shown in Fig. S1, and the top 10 fruitful journals are listed in Table 4. The top 10 journals published 531 articles, accounting for approximately 22.23 % of the total. Frontiers in Neurology was the most prolific outlet with 94 articles, followed by IEEE Access (76 articles), Sensors (68 articles), Scientific Reports (62 articles), and Stroke (49 articles). Stroke was the most influential journal in this field, ranking first in IF (10.17), TC (1235 times), AC (25.65 times), H-index (22), and TLS (442). Encouragingly, two articles [29,30] about ML in stroke were published in the Lancet journal (Q1, 202.731). The publication of these articles in the corresponding high-level journals is enough to show that the "ML in Stroke" study is significant and groundbreaking.

3.5. Analysis of major authors and their cooperation

Over the last decade, a comprehensive study on ML in stroke involved the participation of 13,525 authors, resulting in the publication of 2389 articles. The top 10 contributors played a significant role, publishing 223 articles and accumulating a total of 6351 citations. This accounts for 9.33 % and 19.47 % of the total number of articles and citations, respectively. Table S2 provides an overview of the top 10 active authors, their affiliated organizations, countries, publication outputs, TC, AC, and TLS. The most prolific author in this field was Suri JS from AtheroPoint in the USA, with 36 publications and 1184 citations. Following closely behind was Saba L from Italy, with 35 publications and 1125 citations, and Laird JR from the USA, with 25 publications and 840 citations. Acharya UR from Australia had the highest AC (48.74), but their TLS value was 0, indicating a lack of collaboration with other scholars. Fig. S2

Table 3
Top 10 prolific institutions contributed to publications on ML associated with stroke research.

Rank	Institution	Country	Output [n (%)]	TC	AC	H-index	Centrality	Year
1	University of California System	USA	91 (3.81)	1792	19.69	21	0.36	2013
2	Harvard University	USA	91 (3.81)	1535	16.87	21	0.02	2016
3	Harvard Medical School	USA	58 (2.43)	871	15.02	18	0.01	2016
4	Stanford University	USA	51 (2.13)	1746	34.23	20	0.25	2015
5	Massachusetts General Hospital	USA	51 (2.13)	889	17.43	17	0.10	2016
6	Chinese Academy of Sciences	China	48 (2.01)	528	11.00	10	0.05	2014
7	University of London	UK	45 (1.88)	1069	23.76	16	0.19	2017
8	University of Calgary	Canada	44 (1.84)	594	13.5	12	0.02	2018
9	Capital Medical University	China	41 (1.72)	1212	29.56	10	0.11	2017
10	Mayo Clinic	USA	39 (1.63)	1404	36.00	16	0.23	2019

Note: TC, total citations; AC, average citations per document.

Table 4
Top 10 fruitful journals in the field of ML in stroke by the number of publications.

Rank	Journal	IF	Output [n (%)]	TC	AC	H-index	APY	TLS
1	Frontiers in Neurology	4.086	94 (3.94)	724	7.68	16	2020.81	435
2	IEEE Access	3.476	76 (3.18)	1028	13.67	15	2020.22	155
3	Sensors	3.847	68 (2.85)	686	10.25	13	2020.85	153
4	Scientific Reports	4.997	62 (2.60)	683	11.06	16	2020.55	155
5	Stroke	10.17	49 (2.05)	1235	25.65	22	2020.02	442
6	Diagnostics	3.992	38 (1.59)	197	5.18	7	2021.42	183
7	IEEE Journal of Biomedical and Health Informatics	7.021	38 (1.59)	728	18.25	15	2019.71	92
8	Computers in Biology and Medicine	6.698	37 (1.55)	837	22.89	14	2020.70	182
9	Applied Sciences Basel	2.838	36 (1.51)	226	6.47	8	2020.92	74
10	Journal of Stroke Cerebrovascular Diseases	2.677	33 (1.38)	211	6.67	8	2020.82	152

Note: IF, impact factor (2021); TC, total citations; AC, average citations per document; APY: average publication year; TLS, total link strength.

presents an overlay visualization of the co-authorship network map, focusing on authors who have published a minimum of 10 articles. Ultimately, 30 authors met this threshold. As depicted in Fig. S2, these productive authors have established close cooperative relationships, which have further facilitated the generation of academic achievements in this field.

3.6. Analysis of highly cited articles

Highly cited articles are regarded as one of the most important symbols to evaluate the influence of the general development trend and research frontiers, which have great advantages as an indicator of scientometrics analysis [31]. Based on the data from 2389 publications, this study uses Price's law to identify highly cited articles, and M is used to identify the lowest citation frequency of highly cited articles (Equation (2)).

$$M = 0.749\sqrt{N_{max}} \tag{2}$$

N_{max} is the citation frequency of highly cited articles.

According to the above formula, publications with more than 31 citations in our study are considered highly cited articles. Finally, a total of 234 highly cited articles were detected. Moreover, the top 10 highly cited articles, including first authors (FA), title, published year, source, document type, IF, TC, and AC per year, are listed in Table 5. The Co-occurrence map of 111 highly cited articles with the minimum number of citations of 50 is shown in Fig. S3. Based on the abstracts of these papers, the research topics of the highly cited articles in the field of ML in stroke focused on the diagnosis of brain lesions by 3D magnetic resonance imaging, prediction of risk factors for cerebrovascular disease, ML model based on convolutional neural network (CNN) algorithm for early identification of

Table 5
The top 10 highly cited articles in the field of ML in stroke.

Rank	FA	Title	Year	Source	Type	IF	TC	AC
1	Kamnitsas K et al. [60]	Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation	2017	Medical Image Analysis	Article	8.545	1730	247.14
2	Jiang F et al. [61]	Artificial intelligence in healthcare: past, present and future	2017	Stroke and Vascular Neurology	Review	4.081	928	132.57
3	PoplinR et al. [62]	Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning	2018	Nature Biomedical Engineering	Article	25.671	635	105.83
4	Hsieh, C et al. [63]	Taiwan's National Health Insurance Research Database: past and future	2019	Clinical Epidemiology	Review	4.790	481	96.2
5	Faust O et al. [64]	Deep learning for healthcare applications based on physiological signals: A review	2018	Computer Methods and Programs in Biomedicine	Review	5.428	476	79.33
6	Attia Z et al. [29]	An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction	2019	Lancet	Article	79.323	396	79.2
7	Chen M et al. [65]	Disease Prediction by Machine Learning Over Big Data From Healthcare Communities	2017	IEEE Access	Article	3.367	394	56.29
8	Dou Q et al. [66]	Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks	2016	IEEE Transactions on Medical Imaging	Article	10.048	377	47.13
9	Chilamkurthy S et al. [30]	Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study	2018	Lancet	Article	79.323	354	59
10	Siegel, J et al. [67]	Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke	2016	PNAS	Article	11.205	316	39.5

Note: FA, First Author; IF, impact factor (2021); TC, total citations; AC, average citations per year; PNAS, Proceedings of the National Academy of Sciences of the United States of America.

patients with AF, prediction of impairment after stroke and DL algorithm for CT image analysis. Specifically, 3 of the 10 articles are review articles. The top two highly cited articles were published in 2017, and the others were published between 2016 and 2019. The article titled "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation" got the highest TC with 1730 times and AC with 247.14 times, which suggests it is the most influential article. It is also observed that half of the top 10 highly cited articles had an IF higher than 10, of which two were published in the *Lancet* with an IF of 79.323, which highlights its importance in the field of ML in stroke.

3.7. Analysis of keywords

3.7.1. Co-occurrence analysis of keywords

By analyzing the results of keyword co-occurrence, we can identify the hotspots and frontiers of ML in stroke. A total of 5333 author keywords were detected in this study, of which 112 terms occurred at least 10 times, and 50 terms occurred at least 20 times. In the CiteSpace atlas, centrality represents the influence of the corresponding item in the network, and nodes with centrality exceeding 0.1 are considered key nodes [32]. The top 10 keywords with the highest occurrence and centrality are shown in Table 6. In terms of co-occurrence frequency, 'machine learning' topped the list with 879 occurrences, followed by 'deep learning' (496), 'stroke' (427), 'artificial intelligence' (187), and 'ischemic stroke' (122). The other keywords in the top 10 appeared between 66 and 99 times. From the high-frequency keywords, it can be seen that the application of the ML method based on the CNN algorithm in the prediction and classification of acute ischemic stroke and AF-related stroke was the current research hotspot. In terms of centrality, 'computed tomography' and 'coronary artery disease' tied for first place with 0.26, followed by 'association' (0.21), 'intima-media thickness' (0.20), and 'acute ischemic stroke' (0.19). The rest of the top 10 keywords had centralities ranging from 0.15 to 0.18. The centralities of the top 10 keywords are all more than 0.1, indicating that the above keywords are important hot words in this research field. Computed tomography (CT) imaging plays a pivotal role in the diagnosis and evaluation of various health conditions and their associated risk factors, including evaluation of the structure and potential blockages of coronary arteries, the association between intima-media thickness and the risk of developing acute ischemic stroke, diagnosing stroke and AF, aiding in the identification of these conditions and guiding appropriate treatment strategies. Although 'machine learning', 'deep learning', and 'stroke' are high-frequency keywords, they are not statistically analyzed in the following co-occurrence analysis because they are search terms. The keywords co-occurrence network visual map provided by VOSviewer is shown in Fig. 5A, and node size is determined by the frequency of keywords. Therefore, the larger the node, the higher frequency of these keywords in all the publications. Through the analysis of visual results, keywords such as 'artistic intelligence', 'ischemic stroke', 'atmospheric filtration', 'classification', 'prediction', 'acute ischemic stroke', and 'transactional neural network' have a high frequency of occurrence in these publications.

3.7.2. Cluster timeline analysis of keywords

The cluster timeline map shows the development path of keywords in each cluster. Cluster ranking is based on the number of keywords contained in each cluster. Clusters with a large number of keywords are ranked first, while clusters with a high ranking represent their strong influence in the field. In other words, the cluster number is inversely proportional to the cluster size, the smaller the cluster number, the higher the importance. The cluster timeline map of keywords conducted by CiteSpace is displayed in Fig. 5B. A total of 9 clusters are formed in this field, with cluster labels named # 0 to # 8. Details of each cluster of keywords are shown in Table S3. The nodes on each clustering timeline represent keywords that have burst in different years. The colors on the annual rings range from blue to red, with blue indicating 2013 and red indicating 2023. Cluster #0 was the largest one containing 122 keywords and the research topic is about using ML to predict risk factors for AF-related stroke [33,34]. Cluster # 7 was the earliest cluster with the mean year of 2014. At this time, ML was used to analyze the image of stroke patients and establish an ML model to predict stroke mortality [[35–37]]. Cluster # 6 was the latest one, with the mean year of 2020. The study focuses on outcome prediction in aneurysmal subarachnoid hemorrhage [27,38]. Keyword burst refers to the rapid explosion of high-frequency keywords within a specific period. By detecting and analyzing keyword bursts, we can gain insights into the development, changes, and emerging trends in a particular research field. In this study, we employed CiteSpace to detect keyword bursts and analyze their patterns. Fig. 6 presents the keyword burst map, revealing interesting findings. The keyword 'classification' exhibited the earliest burst in citations and the

Table 6
Top 10 keywords with the highest co-occurrence frequency and centrality in the field of ML in stroke.

Rank	Keyword	Occurrence	TLS	Rank	Keyword	Centrality	Occurrence
1	machine learning	879	1506	1	computed tomography	0.26	98
2	deep learning	496	878	2	coronary artery disease	0.26	16
3	stroke	427	893	3	association	0.21	75
4	artificial intelligence	187	432	4	intima-media thickness	0.20	32
5	ischemic stroke	122	248	5	acute ischemic stroke	0.19	160
6	atrial fibrillation	99	216	6	stroke	0.18	367
7	classification	69	137	7	atrial fibrillation	0.18	129
8	prediction	69	169	8	diagnosis	0.16	95
9	acute ischemic stroke	66	115	9	health	0.16	32
10	convolutional neural network	66	127	10	risk factors	0.15	82

Note: TLS, total link strength.

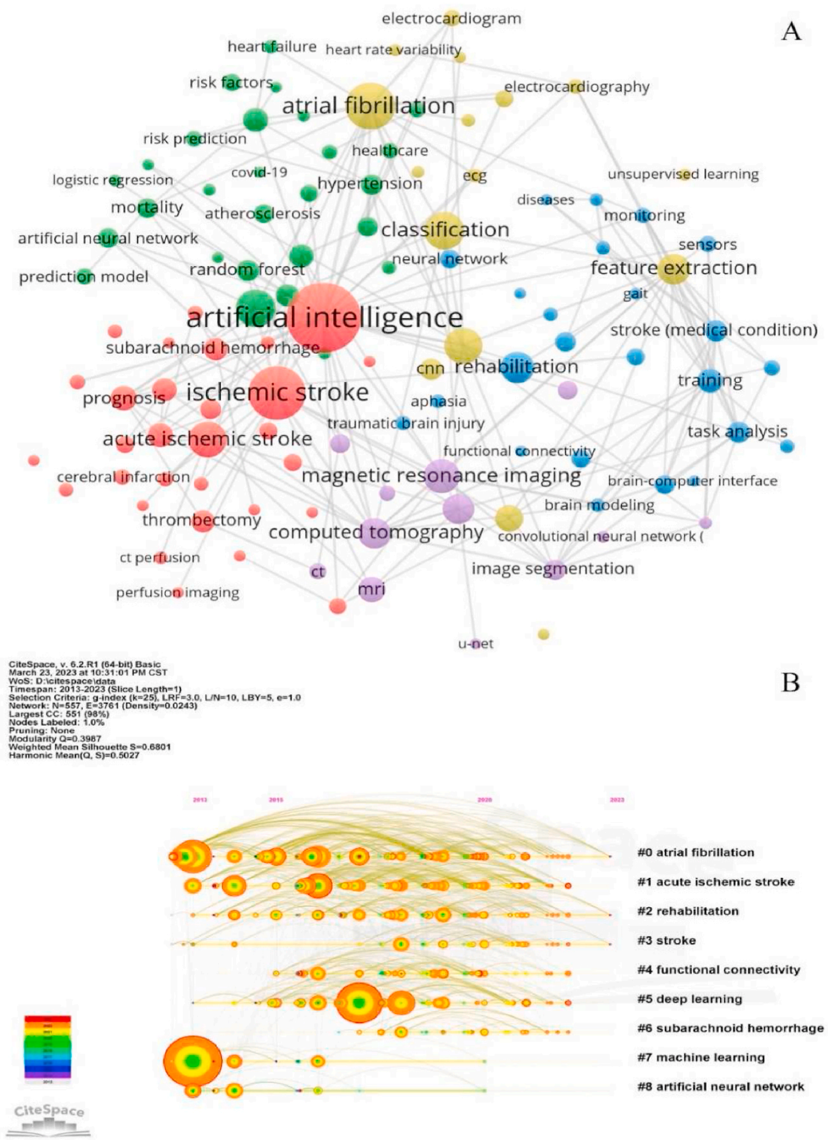

Fig. 5. (A) Co-occurrence map of keywords by excluding search terms (The size of the node represents the frequency of keyword occurrence, the nodes of the same color belong to the same cluster, and the line between the nodes represents the association between the two keywords); (B) Time line and clustering view of keywords (Each circle represents a keyword, and the position of the circle corresponds to the year at the top). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

longest duration, spanning from 2013 to 2018. On the other hand, 'Support Vector Machine' demonstrated the strongest burst strength, reaching 6.2. Furthermore, several keywords have recently experienced a burst surge and are expected to remain popular until 2023. These keywords include 'mechanical thrombectomy', 'expression', and 'prognosis', which represent the current research frontiers in the field. By considering 2018 as a boundary, we can broadly divide the variation of high-frequency keywords into two stages: the initial research stage (2013–2017) and the in-depth research stage (2018–2023). During the initial research stage, ML models were employed for stroke risk classification and prediction. Accelerometer data, capturing patients' movement patterns and behaviors, emerged as a commonly used input data type. Classification algorithms such as SVM and ANN were utilized to categorize patients' accelerometer data into different classes, enabling the assessment of stroke risk. These algorithms learned patterns and associations from known training data to classify and predict new, unlabeled data. As a result, keywords such as 'classification', 'accelerometer', 'Support Vector Machine', 'prediction', 'segmentation', 'Artificial Neural Network', and 'stroke risk' experienced bursts during this period. These applications provided novel perspectives and approaches in stroke research, holding the potential to enhance stroke prevention and treatment strategies. With the advancement of computer technology and the development of AI, the second phase of interest in stroke research has primarily focused on the extensive and in-depth application of ML. ML algorithms are being applied to various aspects of stroke, including cerebral small vessel disease, hemorrhagic stroke, mechanical thrombectomy, and stroke prognosis

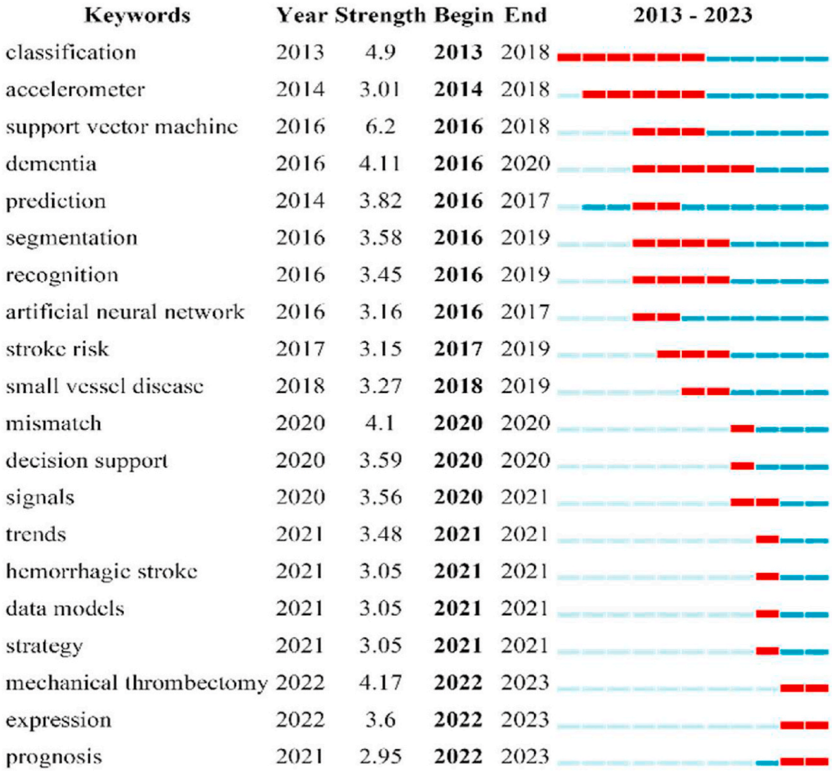

Fig. 6. The top 20 keywords with the strongest citation bursts. The red segment represents the starting and ending year of the burst duration. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

evaluation. As a result, keywords such as ‘small vessel disease’, ‘mismatch’, ‘decision support’, ’signals’, ‘hemorrhagic stroke’, ‘data models’, ‘strategy’, ‘mechanical thrombectomy’, ‘expression’, and ’prognosis’ have emerged as significant during this period. During this stage, ML-powered decision support systems have been developed to assist clinicians in making informed decisions regarding stroke management. These systems integrate diverse data sources, such as patient demographics, medical history, imaging, and laboratory results, to provide personalized recommendations and treatment strategies. These advancements hold immense potential for improving stroke prevention, treatment, and patient outcomes.

4. Discussion

4.1. Main finding

Our study found that research on ML in stroke has drawn increasing attention among scholars, with an explosive increase in the number of publications from 9 in 2013 to 772 in 2022. Moreover, 93.73 % of the total papers were published in the last 5 years, indicating that the field will continue to be of great interest to researchers in the next years. There may be several reasons for the explosive growth. First, the richness of medical data. In recent years, with the continuous development and application of medical technology, we can collect more stroke-related data, such as medical images [39], physiological signals [40,41], genomics data [42], etc. The availability of medical data provides more opportunities and resources for the training and application of ML algorithms. Second, the advancements in ML algorithms. With the development and improvement of ML algorithms, especially the emergence and application of DNN algorithms [25,43], the application coverage of ML in stroke continues to widen, and the effectiveness is getting better and better, which attracted more researchers to join this field. Third, driven by clinical practical needs. Stroke is a common disease that poses a serious threat to people’s health and lives. Therefore, strengthening the early prediction, diagnosis, and treatment of stroke is particularly important. ML models can help doctors predict [43,44], diagnose [45], and treat [46] stroke more accurately and quickly, thus improving the treatment outcomes and reducing morbidity and disability. In a word, the need for early detection, accurate diagnosis, and timely treatment has promoted the increasing application of ML in stroke care. Finally, the contribution of data sharing and cooperation. In recent years, some large-scale international data sharing and collaboration projects, such as ADNI [47], UK Biobank [48], BRAINS Image Bank [49], etc., have made it more convenient for scientists to access large-scale stroke-related data, and to collaborate and communicate with international peers. This has also promoted the application and development of ML in stroke.

Judging from the geographical distribution map, the global distribution of publications was mainly concentrated in North America, East Asia, and Europe. The USA was the largest contributor, accounting for 30.51 %, followed by China (26.62 %) and England (10.55

%). The top three countries/regions contributed 67.68 % of the total number of articles. Centrality refers to the degree to which a particular country, author, or journal is central or influential within a network of related publications, and it is an important measure of influence and impact. It is noteworthy that the centrality (0.04) of the USA, the most productive country, was indeed low. The possible reason is that although the USA has many high-level research institutions and scholars, the dispersion among these research institutions and scholars and the lack of cooperation and communication with other countries may reduce the centrality of the USA in the field of ML in stroke. This can also be seen from the distribution of institutions and authors. Six of the top 10 institutions were from the USA, while only two of the top 10 authors were from the USA, which again proved the dispersion of researchers in the United States. Notably, China was the most active and fastest-growing country, ranking first in terms of annual publication in 2021 and 2022. There are several important reasons for the rapid growth of China in the field of ML in stroke, including the support of national policies [50], active cooperation between various parties [51], a large number of data resources [52,53] and the rapid development of scientific and technological level [50], etc. These above factors jointly promote the rapid development and progress of China in this field.

4.2. Analysis of the research hotspots and frontiers

Analysis of high-frequency keywords, we found that ML and DL have emerged as prominent areas in stroke studies. Various ML/DL models have been widely used in the early diagnosis, classification and prediction of 'acute ischemic stroke' and 'AF-related stroke'. Previous studies have shown that ML algorithms can be used for early diagnosis of AF based on normal sinus rhythm electrocardiographs, allowing for early intervention to reduce stroke risk [29]. At present, ML models have been widely used to predict stroke risk and outcome [54]. For patients with AF, ML models may help clinicians identify high-risk patients and may consequently improve patient outcomes by reducing thromboembolic complications. Han's research team applied ML to develop a classification model to predict the short-term probability of stroke, which has better predictive performance than the traditional CHA_2DS_2-VASc score [55]. Leveraging ML/DL algorithms, researchers aim to enhance the diagnosis and prognosis prediction of stroke by utilizing large datasets. CNN is a commonly used technique as well as highly efficient to map data in a stroke field, and has become the current research hotspot. Additionally, these approaches have the potential to integrate multiple data sources, such as clinical data, imaging data, and genetic data, to discover novel biomarkers and risk factors associated with acute ischemic stroke, offering new insights for stroke prevention and treatment.

The keywords with high centrality are 'computed tomography', 'coronary artery disease', 'association', 'intima-media thickness', and 'acute ischemic stroke', indicating that the application of ML model based on neuroimaging has a significant impact on stroke caused by atherosclerosis and coronary artery disease. CT scan image is the most frequently used data in a stroke field. Neuroimaging is widely used in the diagnosis and evaluation of stroke, and patients suspected of having a stroke are often evaluated with CT [56]. ML algorithms can efficiently and accurately interpret neuroimaging results. For example, a cohort study [13] built a computed tomography perfusion-based ML model to predict follow-up infarction in patients with acute ischemic stroke, confirming that the model was better than current approaches.

From the burst keywords, it can be inferred that 'mechanical thrombectomy', 'expression', 'prognosis', and 'artificial neural network' are the research frontiers. Studies [50,57,58] have proved that ML can be used to predict clinical prognosis before reperfusion therapy, which can provide support for the choice of mechanical thrombectomy. Nishi et al. [58] used ML to predict the clinical outcome of large vessel occlusion before mechanical thrombectomy, demonstrating that ML models had significantly better performance than previously developed pretreatment scoring methods. Yao et al. [57] developed an interpretable ML model that can predict the outcome of ischemic stroke after mechanical thrombectomy in real time and accurately, which has been translated into an online calculator, freely available to the public.

4.3. Applications of ML/DL in clinical practice

In the past decade, ML/DL models have made significant breakthroughs in the field of stroke. Numerous studies have focused on the prediction, diagnosis, treatment, and outcome evaluation of stroke, resulting in improved efficiency and quality of stroke treatment. Currently, there is a greater emphasis on ML in stroke diagnosis compared to treatment [59]. Notably, ML has made remarkable progress in the application of acute ischemic stroke imaging [8]. Techniques such as SVM, RF, and CNN models have proven effective in stroke prevention, diagnosis, treatment, and outcome prediction. The strengths of ML/DL models lie in their powerful capabilities for automatic data processing, strong generalization abilities, self-learning capabilities, rapid diagnosis and prediction capabilities, and precise decision-making abilities. In comparison to traditional statistical inference, ML/DL models, particularly DNN algorithms, are generally more efficient and accurate in predicting outcomes. This opens up possibilities for advanced diagnostic and therapeutic applications in the near future and serves as a key driver for the emergence of precision medicine. Moreover, these approaches have the potential to integrate multiple data sources, including clinical, imaging, and genetic data, to discover novel biomarkers and risk factors associated with acute ischemic stroke, offering new insights for stroke prevention and treatment. Overall, the application of ML/DL models in the field of stroke holds great promise for future research and clinical advancements.

Although there has been some progress in the application of ML/DL in the field of stroke, it is still in its early stages and has not yet been widely adopted in clinical practice. Several challenges and limitations remain. Concerns exist regarding the quality and reliability of data, interpretability of models and results, reliability of clinical outcomes, as well as ethical and moral considerations. Medical care requires not only professional knowledge but also the accumulation of practical experience. ML models cannot replace human-computer interaction or a doctor's comprehensive judgment in making final decisions. These concerns have contributed to some physicians being resistant to the use of AI for decision-making. Addressing these concerns and building trust in ML/DL models among physicians is crucial for successful adoption. This can be achieved through education, transparency in ML/DL algorithms, involving physicians in the development and validation process, and demonstrating the benefits and added value that ML/DL can bring to clinical decision-making. Therefore, further research and validation are still necessary to ensure the reliable and effective application of ML in stroke. It is also critical to promote physician understanding and engagement with ML to fully leverage its role in clinical decision-making.

In conclusion, ML algorithms serve as powerful predictive tools when applied to large medical data sets. They have the potential to greatly enhance stroke prevention, diagnosis, and treatment through their predictive capabilities, automation, ability to discover associations, and personalized approach. ML has brought about groundbreaking innovations in recent years, introducing new ideas and methods for the diagnosis, prediction, and treatment of stroke. This progress is expected to pave the way for individualized precision treatment and drive advancements in scientific research and clinical practice.

5. Limitation

To the best of our knowledge, this is the first scientometric analysis of ML in stroke; however, there are some limitations to this study. First, the WOSCC database was selected as the only source of literature; thus, some literature from other databases was missed. Second, our search was conducted until March 15, 2023, meaning that newer publications may have been excluded. Third, only English articles and reviews were included in our study, and inevitably some literature in other languages, such as Chinese, was omitted, which may affect our conclusions to some extent. Nevertheless, we believe that this work still provides valuable insights into the knowledge structure and development trends in this field. We hope that future research will employ a more comprehensive analysis to further contribute to the advancement of ML in stroke.

6. Conclusion

Through quantitative and qualitative analysis of scientometrics and visual examination of network knowledge maps, this study has comprehensively summarized the current research status of ML in stroke. It has identified the major countries, prolific institutions, active authors and collaborative networks, while also exploring the main research trends and hotspots. The findings reveal that this is a growing research area, with increasing attention from researchers over the past decade. Researchers from various countries and institutions have contributed to this field and will likely continue to do so in the coming years. The global distribution of publications was primarily concentrated in North America, East Asia, and Europe, with the United States and China being the most prolific countries. The applications of ML in the early diagnosis, classification and prediction of 'acute ischemic stroke' and 'AF-related stroke' have emerged as significant research hotspots. Additionally, the utilization of ML models based on CNN in "mechanical thrombectomy" and "prognosis evaluation" of stroke represents the cutting edge of research in this field. Despite these advancements, the clinical application of ML in stroke is still in its early stages, and several limitations and challenges need to be addressed for its widespread adoption in clinical practice. Nevertheless, the findings of this study serve as a valuable starting point, providing useful insights into future research directions and prospects in this rapidly evolving discipline.

Ethics approval and consent to participate

Not Applicable.

Funding

This study was supported by the Nature Foundation of Capital Medical University (Number: PYZ23122).

Data availability statement

The data involved in this study are available in supplementary materials. For further information, please contact the authors.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Mingfen Wu: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Visualization, Writing – original draft. **Kefu Yu:** Methodology, Software, Visualization. **Zhigang Zhao:** Funding acquisition, Supervision, Validation, Writing – review & editing. **Bin Zhu:** Conceptualization, Data curation, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to express their appreciation to Professor CM Chen who invented Citespace, and Professor Van Eck and Waltman who invented VOSviewer, which are free to use. This study was supported by the Nature Foundation of Capital Medical University (Number: PYZ23122).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24230.

References

[1] World Stroke Organization, The Atlas of Heart Disease and Stroke, 2021. Available online: https://www.who.int/cardiovascular_diseases/resources/atlas/en/.
[2] B. Campbell, P. Khatri, Stroke, Lancet. 396 (10244) (2020) 129–142.
[3] B. Ovbiagele, L.B. Goldstein, R.T. Higashida, V.J. Howard, S.C. Johnston, O.A. Khavjou, D.T. Lackland, J.H. Lichtman, S. Mohl, R.L. Sacco, J.L. Saver, J. G. Trogdon, Forecasting the future of stroke in the United States: a policy statement from the American heart association and American stroke association, Stroke 44 (8) (2013) 2361–2375.
[4] W.J. Powers, A.A. Rabinstein, T. Ackerson, O.M. Adeoye, N.C. Bambakidis, K. Becker, J. Biller, M. Brown, B.M. Demaerschalk, B. Hoh, E.C. Jauch, C.S. Kidwell, T.M. Leslie-Mazwi, B. Ovbiagele, P.A. Scott, K.N. Sheth, A.M. Southerland, D.V. Summers, D.L. Tirschwell, Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the american heart Association/American stroke association, Stroke 50 (12) (2019) e344–e418.
[5] A. Rajkomar, J. Dean, I. Kohane, Machine learning in medicine, N. Engl. J. Med. 380 (14) (2019) 1347–1358.
[6] N. Kamal, K. Lakshminarayan, Simulation and machine learning provide new approaches to examine quality of acute stroke management, Stroke 53 (9) (2022) 2768–2769.
[7] H. Shao, W. Chan, H. Du, X.F. Chen, Q. Ma, Z. Shao, A new machine learning algorithm with high interpretability for improving the safety and efficiency of thrombolysis for stroke patients: a hospital-based pilot study, Digit. Health. 9 (2023) 579793240.
[8] S.A. Sheth, L. Giancardo, M. Colasurdo, V.M. Srinivasan, A. Niktabe, P. Kan, Machine learning and acute stroke imaging, J. Neurointerv. Surg. 15 (2) (2023) 195–199.
[9] W. Qiu, H. Kuang, E. Teleg, J.M. Ospel, S.I. Sohn, M. Almekhlafi, M. Goyal, M.D. Hill, A.M. Demchuk, B.K. Menon, Machine learning for detecting early infarction in acute stroke with non-contrast-enhanced CT, Radiology 294 (3) (2020) 638–644.
[10] M.R. Arbabshirani, B.K. Fornwalt, G.J. Mongelluzzo, J.D. Suever, B.D. Geise, A.A. Patel, G.J. Moore, Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration, NPJ Digit. Med. 1 (2018) 9.
[11] W.X. Yang, F.F. Wang, Y.Y. Pan, J.Q. Xie, M.H. Lu, C.G. You, Comparison of ischemic stroke diagnosis models based on machine learning, Front. Neurol. 13 (2022) 1014346.
[12] J. Wang, J. Zhang, X. Gong, W. Zhang, Y. Zhou, M. Lou, Prediction of large vessel occlusion for ischaemic stroke by using the machine learning model random forests, Stroke Vasc. Neurol. 7 (2) (2022) 94–100.
[13] H. Kuang, W. Qiu, A.M. Boers, S. Brown, K. Muir, C. Majoie, D. Dippel, P. White, J. Epstein, P.J. Mitchell, A. Davalos, S. Bracard, B. Campbell, J.L. Saver, T. G. Jovin, M. Rubiera, A.V. Khaw, J.J. Shankar, E. Fainardi, M.D. Hill, A.M. Demchuk, M. Goyal, B.K. Menon, Computed tomography Perfusion-Based machine learning model better predicts Follow-Up infarction in patients with acute ischemic stroke, Stroke 52 (1) (2021) 223–231.
[14] S. Campagnini, C. Arienti, M. Patrini, P. Liuzzi, A. Mannini, M.C. Carrozza, Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation: a systematic review, J. NeuroEng. Rehabil. 19 (1) (2022) 54.
[15] S.K. Jang, J.Y. Chang, J.S. Lee, E.J. Lee, Y.H. Kim, J.H. Han, D.I. Chang, H.J. Cho, J.K. Cha, K.H. Yu, J.M. Jung, S.H. Ahn, D.E. Kim, S.I. Sohn, J.H. Lee, K.P. Park, S.U. Kwon, J.S. Kim, D.W. Kang, Reliability and clinical utility of machine learning to predict stroke prognosis: comparison with logistic regression, J. Stroke. 22 (3) (2020) 403–406.
[16] J. Heo, J.G. Yoon, H. Park, Y.D. Kim, H.S. Nam, J.H. Heo, Machine Learning-Based model for prediction of outcomes in acute stroke, Stroke 50 (5) (2019) 1263–1265.
[17] J. Lv, M. Zhang, Y. Fu, M. Chen, B. Chen, Z. Xu, X. Yan, S. Hu, N. Zhao, An interpretable machine learning approach for predicting 30-day readmission after stroke, Int. J. Med. Inform. 174 (2023) 105050.
[18] S. Mainali, M.E. Darsie, K.S. Smetana, Machine learning in action: stroke diagnosis and outcome prediction, Front. Neurol. 12 (2021) 734345.
[19] Y.A. Choi, S.J. Park, J.A. Jun, C.S. Pyo, K.H. Cho, H.S. Lee, J.H. Yu, Deep Learning-Based stroke disease prediction system using Real-Time bio signals, Sensors 21 (13) (2021) 4269.
[20] M.S. Islam, I. Hussain, M.M. Rahman, S.J. Park, M.A. Hossain, Explainable artificial intelligence model for stroke prediction using EEG signal, Sensors 22 (24) (2022) 9859.
[21] I. Hussain, S.J. Park, Quantitative evaluation of task-induced neurological outcome after stroke, Brain Sci. 11 (7) (2021) 900.
[22] S. Raghunath, J.M. Pfeifer, A.E. Ulloa-Cerna, A. Nemani, T. Carbonati, L. Jing, D.P. Vanmaanen, D.N. Hartzel, J.A. Ruhl, B.F. Lagerman, D.B. Rocha, N.J. Stoudt, G. Schneider, K.W. Johnson, N. Zimmerman, J.B. Leader, H.L. Kirchner, C.J. Griessenauer, A. Hafez, C.W. Good, B.K. Fornwalt, C.M. Haggerty, Deep neural networks can predict New-Onset atrial fibrillation from the 12-Lead ECG and help identify those at risk of atrial Fibrillation-Related stroke, Circulation 143 (13) (2021) 1287–1298.
[23] I. Hussain, S.J. Park, Prediction of myoelectric biomarkers in post-stroke gait, Sensors 21 (16) (2021) 5334.
[24] R. Pitchai, B. Dappuri, P.V. Pramila, M. Vidhyalakshmi, S. Shanthi, W.B. Alonazi, K. Almutairi, R.S. Sundaram, I. Beyene, An artificial Intelligence-Based Bio-Medical stroke prediction and analytical system using a machine learning approach, Comput. Intell. Neurosci. 2022 (2022) 5489084.
[25] I.R. Chavva, A.L. Crawford, M.H. Mazurek, M.M. Yuen, A.M. Prabhat, S. Payabvash, G. Sze, G.J. Falcone, C.C. Matouk, A. de Havenon, J.A. Kim, R. Sharma, S. J. Schiff, M.S. Rosen, J. Kalpathy-Cramer, G.J. Iglesias, W.T. Kimberly, K.N. Sheth, Deep learning applications for acute stroke management, Ann. Neurol. 92 (4) (2022) 574–587.
[26] J.G. Greener, S.M. Kandathil, L. Moffat, D.T. Jones, A guide to machine learning for biologists, Nat. Rev. Mol. Cell Biol. 23 (1) (2022) 40–55.
[27] N.F. Dengler, V.I. Madai, M. Unteroberdorster, E. Zihni, S.C. Brune, A. Hilbert, M. Livne, S. Wolf, P. Vajkoczy, D. Frey, Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinico-radiological scores, Neurosurg. Rev. 44 (5) (2021) 2837–2846.
[28] A. Shao, K. Jin, Y. Li, L. Lou, W. Zhou, J. Ye, Overview of global publications on machine learning in diabetic retinopathy from 2011 to 2021: bibliometric analysis, Front. Endocrinol. 13 (2022) 1032144.

[29] Z.I. Attia, P.A. Noseworthy, F. Lopez-Jimenez, S.J. Asirvatham, A.J. Deshmukh, B.J. Gersh, R.E. Carter, X. Yao, A.A. Rabinstein, B.J. Erickson, S. Kapa, P. A. Friedman, An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction, Lancet 394 (10201) (2019) 861–867.
[30] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N.G. Campeau, V.K. Venugopal, V. Mahajan, P. Rao, P. Warier, Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study, Lancet 392 (10162) (2018) 2388–2396.
[31] J. Wang, F. Shahzad, A visualized and scientometric analysis of health literacy research, Front. Public Health 9 (2021) 811707.
[32] P. Deng, S. Wang, X. Sun, Y. Qi, Z. Ma, X. Pan, H. Liang, J. Wu, Z. Chen, Global trends in research of gouty arthritis over past decade: a bibliometric analysis, Front. Immunol. 13 (2022) 910400.
[33] C.M. Liu, M.E. Hsieh, Y.F. Hu, T.Y. Wei, I.C. Wu, P.F. Chen, Y.J. Lin, S. Higa, N. Yagi, S.A. Chen, V.S. Tseng, Artificial Intelligence-Enabled model for early detection of left ventricular hypertrophy and mortality prediction in young to Middle-Aged adults, Circ. Cardiovasc. Qual. Outcomes. 15 (8) (2022) e8360.
[34] Y.S. Lou, C.S. Lin, W.H. Fang, C.C. Lee, C.L. Ho, C.H. Wang, C. Lin, Artificial Intelligence-Enabled electrocardiogram estimates left atrium enlargement as a predictor of future cardiovascular disease, J. Pers. Med. 12 (2) (2022) 315.
[35] C.M. Lineback, R. Garg, E. Oh, A.M. Naidech, J.L. Holl, S. Prabhakaran, Prediction of 30-Day readmission after stroke using machine learning and natural language processing, Front. Neurol. 12 (2021) 649521.
[36] Chen, Y., Qin, C., Chang, J., Lyu, Y., Zhang, Q., Ye, Z., Li, Z., Tian, F., Ma, W., Wei, J., Feng, M., Yao, J., Wang, R. A machine learning approach for predicting perihematomal edema expansion in patients with intracerebral hemorrhage. Eur. Radiol. 33(6): 4052-4062..
[37] N. Farzaneh, C.A. Williamson, C. Jiang, A. Srinivasan, J.R. Bapuraj, J. Gryak, K. Najarian, S. Soroushmehr, Automated segmentation and severity analysis of subdural hematoma for patients with traumatic brain injuries, Diagnostics 10 (10) (2020) 773.
[38] D. Yu, G.W. Williams, D. Aguilar, J.M. Yamal, V. Maroufy, X. Wang, C. Zhang, Y. Huang, Y. Gu, Y. Talebi, H. Wu, Machine learning prediction of the adverse outcome for nontraumatic subarachnoid hemorrhage patients, Ann. Clin. Transl. Neurol. 7 (11) (2020) 2178–2185.
[39] G. Wang, T. Song, Q. Dong, M. Cui, N. Huang, S. Zhang, Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks, Med. Image Anal. 65 (2020) 101787.
[40] D.S. Tian, C. Qin, L.Q. Zhou, S. Yang, M. Chen, J. Xiao, K. Shang, D.B. Bosco, L.J. Wu, W. Wang, FSAP aggravated endothelial dysfunction and neurological deficits in acute ischemic stroke due to large vessel occlusion, Signal Transduct. Target Ther. 7 (1) (2022) 6.
[41] M. Bacigaluppi, G. Martino, FABP4 a novel therapeutic target in ischaemic stroke, Eur. Heart J. 41 (33) (2020) 3181–3183.
[42] J.N. Acosta, S.C. Brown, G.J. Falcone, Genetic underpinnings of recovery after stroke: an opportunity for gene discovery, risk stratification, and precision medicine, Genome Med. 11 (1) (2019) 58.
[43] Y. Yu, S. Christensen, J. Ouyang, F. Scalzo, D.S. Liebeskind, M.G. Lansberg, G.W. Albers, G. Zaharchuk, Predicting hypoperfusion lesion and target mismatch in stroke from diffusion-weighted MRI using deep learning, Radiology 307 (1) (2023) e220882.
[44] D.Y. Kim, K.H. Choi, J.H. Kim, J. Hong, S.M. Choi, M.S. Park, K.H. Cho, Deep learning-based personalised outcome prediction after acute ischaemic stroke, J. Neurol. Neurosurg. Psychiatry 94 (5) (2023) 369–378.
[45] T. Cai, H. Ni, M. Yu, X. Huang, K. Wong, J. Volpi, J.Z. Wang, S. Wong, DeepStroke: an efficient stroke screening framework for emergency rooms with multimodal adversarial deep learning, Med. Image Anal. 80 (2022) 102522.
[46] A. Nielsen, M.B. Hansen, A. Tietze, K. Mouridsen, Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning, Stroke 49 (6) (2018) 1394–1401.
[47] C.J. Weber, M.C. Carrillo, W. Jagust, C.J. Jack, L.M. Shaw, J.Q. Trojanowski, A.J. Saykin, L.A. Beckett, C. Sur, N.P. Rao, P.C. Mendez, S.E. Black, K. Li, T. Iwatsubo, C.C. Chang, A.L. Sosa, C.C. Rowe, R.J. Perrin, J.C. Morris, A. Healan, S.E. Hall, M.W. Weiner, The Worldwide Alzheimer's disease neuroimaging initiative: ADNI-3 updates and global perspectives, Alzheimers Dement (N Y). 7 (1) (2021) e12226.
[48] L.C. Rutten-Jacobs, S.C. Larsson, R. Malik, K. Rannikmae, C.L. Sudlow, M. Dichgans, H.S. Markus, M. Traylor, Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants, BMJ 363 (2018) k4168.
[49] Brain Images of Normal Subjects, 2023. Available online: http://www.brainsimagebank.ac.uk/ (Accessed 6 April 2023).
[50] L. Ding, C. Liu, Z. Li, Y. Wang, Incorporating artificial intelligence into stroke care and research, Stroke 51 (12) (2020) e351–e354.
[51] S. Wu, B. Wu, M. Liu, Z. Chen, W. Wang, C.S. Anderson, P. Sandercock, Y. Wang, Y. Huang, L. Cui, C. Pu, J. Jia, T. Zhang, X. Liu, S. Zhang, P. Xie, D. Fan, X. Ji, K. L. Wong, L. Wang, Stroke in China: advances and challenges in epidemiology, prevention, and management, Lancet Neurol. 18 (4) (2019) 394–405.
[52] W. Wang, B. Jiang, H. Sun, X. Ru, D. Sun, L. Wang, L. Wang, Y. Jiang, Y. Li, Y. Wang, Z. Chen, S. Wu, Y. Zhang, D. Wang, Y. Wang, V.L. Feigin, Prevalence, incidence, and mortality of stroke in China: results from a nationwide population-based survey of 480 687 adults, Circulation 135 (8) (2017) 759–771.
[53] T. Guan, J. Ma, M. Li, T. Xue, Z. Lan, J. Guo, Y. Shen, B. Chao, G. Tian, Q. Zhang, L. Wang, Y. Liu, Rapid transitions in the epidemiology of stroke and its risk factors in China from 2002 to 2013, Neurology 89 (1) (2017) 53–61.
[54] F.K. Wegner, L. Plagwitz, F. Doldi, C. Ellermann, K. Willy, J. Wolfes, S. Sandmann, J. Varghese, L. Eckardt, Machine learning in the detection and management of atrial fibrillation, Clin. Res. Cardiol. 111 (9) (2022) 1010–1017.
[55] L. Han, M. Askari, R.B. Altman, S.K. Schmitt, J. Fan, J.P. Bentley, S.M. Narayan, M.P. Turakhia, Atrial fibrillation burden signature and Near-Term prediction of stroke: a machine learning analysis, Circ Cardiovasc Qual Outcomes 12 (10) (2019) e5595.
[56] A.L. Czap, S.A. Sheth, Overview of imaging modalities in stroke, Neurology 97 (20 Suppl 2) (2021) S42–S51.
[57] Z. Yao, C. Mao, Z. Ke, Y. Xu, An explainable machine learning model for predicting the outcome of ischemic stroke after mechanical thrombectomy, J. Neurointerv. Surg. 15 (11) (2023) 1136–1141.
[58] H. Nishi, N. Oishi, A. Ishii, I. Ono, T. Ogura, T. Sunohara, H. Chihara, R. Fukumitsu, M. Okawa, N. Yamana, H. Imamura, N. Sadamasa, T. Hatano, I. Nakahara, N. Sakai, S. Miyamoto, Predicting clinical outcomes of large vessel occlusion before mechanical thrombectomy using machine learning, Stroke 50 (9) (2019) 2379–2388.
[59] M.S. Sirsat, E. Ferme, J. Camara, Machine learning for brain stroke: a review, J. Stroke Cerebrovasc. Dis. 29 (10) (2020) 105162.
[60] K. Kamnitsas, C. Ledig, V. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med. Image Anal. 36 (2017) 61–78.
[61] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, Artificial intelligence in healthcare: past, present and future, Stroke Vasc. Neurol. 2 (4) (2017) 230–243.
[62] R. Poplin, A.V. Varadarajan, K. Blumer, Y. Liu, M.V. Mcconnell, G.S. Corrado, L. Peng, D.R. Webster, Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning, Nat. Biomed. Eng. 2 (3) (2018) 158–164.
[63] C.Y. Hsieh, C.C. Su, S.C. Shao, S.F. Sung, S.J. Lin, Y.Y. Kao, E.C. Lai, Taiwan's national health insurance research database: past and future, Clin. Epidemiol. 11 (2019) 349–358.
[64] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: a review, Comput. Methods Programs Biomed. 161 (2018) 1–13.
[65] M. Chen, Y. Hao, K. Hwang, L. Wang, L. Wang, Disease prediction by machine learning over big data from healthcare communities, IEEE Access (2017) 8869–8879.
[66] D. Qi, C. Hao, Y. Lequan, Z. Lei, Q. Jing, W. Defeng, V.C. Mok, S. Lin, H. Pheng-Ann, Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks, IEEE Trans. Med. Imaging. 35 (5) (2016) 1182–1195.
[67] J.S. Siegel, L.E. Ramsey, A.Z. Snyder, N.V. Metcalf, R.V. Chacko, K. Weinberger, A. Baldassarre, C.D. Hacker, G.L. Shulman, M. Corbetta, Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke, Proc. Natl. Acad. Sci. USA 113 (30) (2016) E4367–E4376.