AJNR

Assessing the Emergence and Evolution of Artificial Intelligence and Machine Learning Research in Neuroradiology

Alexandre Boutet, Samuel S. Haile, Andrew Z. Yang, Hyo Jin Son, Mikail Malik, Vivek Pai, Mehran Nasralla, Jurgen Germann, Artur Vetkas, Farzad Khalvati and Birgit B. Ertl-Wagner

AJNR Am J Neuroradiol published online 23 March 2024
http://www.ajnr.org/content/early/2024/03/23/ajnr.A8252

Published March 23, 2024 as 10.3174/ajnr.A8252

ORIGINAL RESEARCH

ARTIFICIAL INTELLIGENCE

Assessing the Emergence and Evolution of Artificial Intelligence and Machine Learning Research in Neuroradiology.

Alexandre Boutet*, Samuel S. Haile*, Andrew Z. Yang, Hyo Jin Son, Mikail Malik, Vivek Pai, Mehran Nasralla, Jurgen Germann, Artur Vetkas, Farzad Khalvati** and Birgit B. Ertl-Wagner**

***These authors contributed equally.**
****These authors jointly supervised this work.**

ABSTRACT

BACKGROUND AND PURPOSE: Interest in artificial intelligence (AI) and machine learning (ML) has been growing in neuroradiology, but there is limited knowledge on how this interest has manifested into research and specifically, its qualities and characteristics. This study aims to characterize the emergence and evolution of AI/ML articles within neuroradiology and provide a comprehensive overview of the trends, challenges, and future directions of the field.

MATERIALS AND METHODS: We performed a bibliometric analysis of the *American Journal of Neuroradiology* (AJNR): the journal was queried for original research articles published since inception (Jan. 1, 1980) to Dec. 3, 2022 that contained any of the following key terms: “machine learning”, “artificial intelligence”, “radiomics”, “deep learning”, “neural network”, “generative adversarial network”, “object detection”, or “natural language processing”. Articles were screened by two independent reviewers, and categorized into Statistical Modelling (Type 1), AI/ML Development (Type 2), both representing developmental research work but without a direct clinical integration, or End-user Application (Type 3) which is the closest surrogate of potential AI/ML integration into day-to-day practice. To better understand the limiting factors to Type 3 articles being published, we analyzed Type 2 articles as they should represent the precursor work leading to Type 3.

RESULTS: A total of 182 articles were identified with 79% being non-integration focused (Type 1 n = 53, Type 2 n = 90) and 21% (n = 39) being Type 3. The total number of articles published grew roughly five-fold in the last five years, with the non-integration focused articles mainly driving this growth. Additionally, a minority of Type 2 articles addressed bias (22%) and explainability (16%). These articles were primarily led by radiologists (63%), with most of them (60%) having additional postgraduate degrees.

Copyright 2024 by American Society of Neuroradiology.

CONCLUSIONS: AI/ML publications have been rapidly increasing in neuroradiology with only a minority of this growth being attributable to end-user application. Areas identified for improvement include enhancing the quality of Type 2 articles, namely external validation, and addressing both bias and explainability. These results ultimately provide authors, editors, clinicians, and policymakers important insights to promote a shift towards integrating practical AI/ML solutions in neuroradiology.

ABBREVIATIONS: AI = artificial intelligence; ML = machine learning.

Received month day, year; accepted after revision month day, year.
From the Joint Department of Medical Imaging (A.B., M.N.), University Health Network, Toronto, ON, Canada; Temerty Faculty of Medicine (S.S.H., H.S., M.M.), Division of Neurosurgery, Department of Surgery, Toronto Western Hospital (A.Z.W., J.G., A.V.), Department of Medical Imaging (V.P., F.K., B.B.E.-W.), Department of Computer Science (F.K.), Department of Mechanical and Industrial Engineering (F.K.), University of Toronto, Toronto, ON, Canada; Division of Neuroradiology, Department of Diagnostic Imaging (V.P., B.B.E.-W.) The Hospital for Sick Children, Toronto, ON, Canada; Neurosciences and Mental Health Program, (F.K., B.B.E.-W.) SickKids Research Institute, Toronto, ON, Canada.

The authors declare no conflicts of interest related to the content of this article.

Please address correspondence to Alexandre Boutet, MD, PhD, Joint Department of Medical Imaging, University Health Network, University of Toronto 399 Bathurst St., Toronto, ON M5T 2S8, Canada; e-mail: alexandre.boutet@uhn.ca.

SUMMARY SECTION

PREVIOUS LITERATURE: Previous studies examining the trends in artificial intelligence (AI) and machine learning (ML) research in medicine has established medical imaging to be a major area of application, accounting for up to 20% of publications. Also, these studies have demonstrated neuroradiology to be a leading area of AI research. The studies that have looked at AI research trends in medical imaging have maintained a broad scope examining growth and research topics, but there is a lack of studies specifically investigating areas that can affect the future integration of this research such as article quality.

KEY FINDINGS: AI/ML publications in the AJNR have grown roughly five-fold in the last five years, mainly driven by developmental research work and only a minority of articles (21%) investigated end-user applications of AI. Several areas of improvement were identified including external validation, bias, and explainability.

KNOWLEDGE ADVANCEMENT: Although this study confirms the rapid growth of AI/ML research in neuroradiology, it has also identified areas for improvement in the field. These results ultimately provide authors, editors, clinicians, and policymakers important insights to promote a shift towards integrating practical AI/ML solutions in neuroradiology.

INTRODUCTION

Artificial intelligence (AI) is a field focusing on the development of tools that mimic human cognition, especially its learning, decision, and interpretation capabilities. In recent years, the field has rapidly grown in its knowledge base and complexity. One example of this growth is machine learning (ML)—a subfield of AI that uses algorithms and models to enable machines to independently learn from experience or data. The use of AI and ML has become widespread in its applications, especially with recent unprecedented advances in Large Language Models, such as OpenAI's ChatGPT. AI research in the field of medicine has exponentially grown with more than 60% of all articles that existed in 2018 having been published between

2008 and 2018.[1] The main factors driving this growth are recent increases in computing power and the availability of large, high-quality datasets. Additional factors include increased interest in AI's potential as a scalable solution to today's demands of service efficiency, intervention efficacy, and personalized healthcare.[2]

A recent bibliometric study by Kocak et al. that examined the trends in AI research in medical imaging showed medical imaging as a major application of AI/ML in medicine, accounting for 20% of AI in medicine publications.[3] Additionally, a narrative review by Pesapane et al. demonstrated that from among the different medical imaging subspecialties, neuroradiology has been consistently identified as leader in AI/ML publications accounting for 34% of the AI/ML articles published in the field of medical imaging.[4] Moreover, a recently published review conducted by Wagner et al. which outlined current topics in AI and neuroradiology found this to be one of the most heavily researched fields in medical imaging.[5] These studies however maintained a broad scope, mostly examining growth, citations, and topics of AI/ML articles rather than more specific indicators of AI/ML article quality.[5]

The popularity of AI/ML in neuroradiology can partly be explained by: (1) the prominent role of imaging for diagnosis, (2) the availability of brain-specific public datasets, (3) the use of complex multidimensional and multimodal data, and (4) the attraction of researchers to the subspeciality due to the leading role of neuroimaging in the field of Big Data, and linked to that, advanced data analysis and AI.[2,6] However, there is a paucity of studies focused on the emergence and evolution of AI/ML research in the field of neuroradiology. This could help us understand whether this increased interest in AI/ML is reflected in the amount of research published and also how the characteristics and quality of this research may influence its integration into neuroradiology practice.[3,4]

To provide an understanding of the emergence and evolution of AI/ML in neuroradiology, we bibliometrically analyzed the *American Journal of Neuroradiology* (AJNR). First, we assessed the number of research articles focusing on AI/ML integration into day-to-day practice. To gain a better understanding of the factors that may be limiting their publication, we then focused our analysis on the developmental research articles as they should represent the precursor work.

MATERIALS AND METHODS

Study Design

We used the bibliometric method to analyze original articles on AI/ML in AJNR. This method facilitates the examination of a particular field of research by providing a concise overview based on the published literature while also shedding light on details that drive the evolving nature and challenges of the field.

Search Strategy and Eligibility Criteria

We included original research articles published between journal inception (January 1, 1980) to the date the query was conducted (Dec. 3, 2023) that contained any of the following key terms: "machine learning", "artificial intelligence", "radiomic", "deep learning", "neural network", "generative adversarial network", "object detection", "natural language processing" and fit the definitions of AI and ML as described in Erickson et al. (2017) and Wagner et al. (2021) (refer to Supplementary Materials).[7,8] We used the advanced search function on the AJNR website and then queried the journal for original research articles containing the key terms, however it is limited by the fact that it retrieves articles that mention keywords anywhere in the article, irrespective of their use and context. Articles without available full-text and review articles were excluded.

Selection of Studies

The identified articles were then uploaded into Covidence (https://www.covidence.org/) to facilitate screening. Two reviewers (A.B., and M.N., trained academic neuroradiologists with 6 years' experience and neuroradiology fellows) independently screened titles and abstracts for relevancy based on the inclusion criteria. Conflicts between the two reviewers were resolved by a third independent reviewer (V.P., neuroradiology fellow). Articles that successfully passed this stage of the screening process moved on to full-text review. Two reviewers (A.B and M.N. with V.P. as adjudicator) independently performed the full-text reviews and determined whether these articles completely met the inclusion criteria with reasons for exclusion being recorded.

Data Extraction

The included articles were categorized into three types of AI/ML studies: statistical modelling (Type 1), AI/ML development (Type 2) both representing developmental research work but without a direct clinical integration, or end-user application (Type 3) which is the closest surrogate of potential AI/ML integration into day-to-day practice using the following definitions:

Statistical Modelling (Type 1): Article uses statistics to analyze sample data and make predictions. These articles are usually hypothesis-based, with a limited number of candidate biomarkers to be tested using statistical methods.

AI/ML Development (Type 2): Article develops and trains (may additionally test and validate) an AI/ML model.

End-user Applications (Type 3): Article applies pre-existing AI/ML product/software.

Additionally, the year of publication and number of authors were extracted from each article. As Type 2 AI/ML development articles typically precede Type 3 end-user application articles, we extracted more data such as the profession and degrees of first/corresponding author(s), geographic location of data source, source of funding, and neuroradiology subfield of the article. We also extracted data from these articles relevant to study quality such as the study design (pro- vs. retrospective;

multi- vs. single-center), types of AI/ML models used, whether testing or validation was done, type of validation (internal and/or external), and whether learning was supervised or unsupervised. Each article was also examined for whether it addressed bias or explainability—a concept relating to whether the features extracted, or predictions made by the authors' AI/ML models could be reasonably understood by humans. The data were summarized using descriptive statistical methods.

RESULTS

Our search identified 657 AJNR original articles of which 116 duplicates were removed. The remaining 541 articles were screened of which 349 were excluded. The main reason for exclusion was the lack of specificity of the search engine often retrieving key terms used without the study employing AI/ML methodology. 192 articles underwent full-text review to confirm their inclusion criteria eligibility. From these 192 articles, 10 articles were excluded as they did not fit the definitions of AI/ML.[7,8] The 182 remaining articles were included in our bibliometric analysis (Fig 1).

Type 3 end-user application articles remain a minority

The number of articles published has steadily grown since 2007 with a twenty-five- and five-fold increase in the last fifteen and five years, respectively. Developmental research articles without direct clinical integration represented 79% of all AI/ML articles (Type 1 – Statistical Modelling, n = 53; Type 2 – AI/ML Development, n = 90) with the remaining 21% of articles being integration-focused (Type 3 – End-user application, n = 39). The proportion of Type 1 articles published per year has greatly decreased to levels less than those of the Type 3 articles while the number of Type 2 articles has rapidly increased in recent years, now representing the majority of articles (Fig 2). The average number of authors per publication was large but similar across article types with Type 1, 2, and 3 articles having 9, 10, and 9 authors on average, respectively.

Analysis of Type 2 articles sheds insight into the lack of Type 3 articles

We found a consistent increase in the number of Type 2 articles published since 2007, with a four-fold increase in the last five years. The vast majority of the ninety Type 2 articles focused solely on the adult population (n = 72, 80%) with the remainder focusing on pediatric (n = 14, 16%) or adult and pediatric populations (n = 4, 4%). The most frequent topic representing 40% (n = 36) of all published Type 2 articles was oncology with the second being vascular disease (Fig 3b-c). Regarding the geographical demographics of Type 2 studies, we found that a large proportion (n = 42, 47%) were published by authors based in the United States, with South Korea (n = 12, 13%) coming in a distant second. In addition to a variety of countries being represented in the data, eight articles used public datasets and were not associated with a specific country (Fig 3a).

Two-thirds (n = 57, 63%) of all Type 2 publications had a radiologist as a first or corresponding/senior author and among these radiologist-led Type 2 publications, sixty percent (n = 34, 60%) were led by a radiologist with graduate degrees, many of them being PhD and MBA degree holders. Most articles (n = 71, 79%) received some type of funding, most commonly from governmental organizations (e.g., National Institutes of Health, Department of Defense, National Cancer Institute), foundations, and universities, with nineteen articles receiving no funding. Although there were only six industry-funded articles, all of them were recently published in 2022 and 2023.

In terms of the quality of Type 2 articles, we found more single-centered (58%) than multi-centered designs (42%). However, the vast majority were retrospective with only 17% (n = 15/90) being prospective studies (Fig 4a-b). Our full-text review also revealed that only 14% (n = 13/90) of Type 2 articles were externally validated (Fig 4c). Algorithm type was categorized into deep (e.g., convolutional neural networks, generative adversarial networks, multilayer perceptrons) or conventional learning (e.g., linear regression, random forest, support vector machine) techniques. We found that these publications favored the use of more complex algorithms with deep and conventional learning techniques representing 56% and 44% of articles respectively (Fig 4d).

Additionally, we found that only 22% (n = 20/90) of Type 2 articles mentioned bias as a point of discussion, either as something successfully avoided, a limitation, or recognized as an area for future improvement. Similarly, the concept of explainability or interpretability in AI was addressed in only 16% (n = 14/90) of articles. We also found that Type 2 articles published between 2020 and 2023 adequately addressed bias at higher rates (n = 17/68, 25%) than those published prior to 2020 (n = 3/22, 13.6%). Additionally, Type 2 articles published between 2020 and 2023 addressed explainability at slightly higher rates (n = 11/68, 16.2%) than those published prior to 2020 (n = 3/22, 13.6%).

DISCUSSION

Our results demonstrate that the number of AI/ML publications in the AJNR has been rapidly increasing, with Type 2 articles being the major drivers of this overall growth. Only a minority of these new articles focused on end-user application (Type 3), indicating AI/ML research in the field is still centered on research rather than clinical integration although this has been growing. Our data points to specific characteristics and deficiencies of the Type 2 articles, which provide potential explanations and solutions to improve the quality of research and accelerate its integration.

Our observed rapid increase in AI/ML publications in AJNR is consistent with the overall growth of AI research in medicine, one big exception being the drop in articles published in 2020, most likely due to the COVID-19 pandemic.[1,9] The decrease in Type 1 and stark increase in Type 2 articles is consistent with what we would expect of a relatively new and growing field. As more research is conducted and published, researchers gain more experience and knowledge and can build upon previous publications to produce more complex research. As expected, we found only a small number of Type 3 end-user application articles, but we anticipate that they will eventually become the leading type of AI article especially as industries become more interested and begin to increase their funding.[10] This shift in article focus over time suggests authors should begin to shift their focus toward research centered on implementing and integrating AI in a manner that is feasible and clinically relevant.

Our analysis of Type 2 articles showed a four-fold increase in the last 5 years compared to the five-fold increase of all AI/ML articles (Type 1, 2, and 3 combined) in the same time period. This trend is in line with the evolving complexity of the field which has moved away from its earlier use of conventional statistical or "shallow learning" methods and towards using deep learning methods. Should a shift to deep learning methods be sustained, we would anticipate seeing a similar shift towards more Type 3 end-user application studies as deep learning methods are more suited for tackling contemporary challenges.[10] The analysis of Type 2 articles also showed characteristics and weaknesses that may limit AI/ML research and its subsequent integration into day-to-day practice. Our results demonstrated a focus on the adult population as well as oncological and vascular diseases. The limited amount of sufficiently large pediatric datasets restricts the ability of researchers to train high quality AI models specific to this population. AI models focused on the pediatric population have also been shown to have lower rates of clinically significant outcomes relative to the adult population which can be explained by the limited amount of data.[11,12] It is plausible that the field may be slowed down in the "research phase" with an attempt to solve these issues, resulting in more Type 2 articles being published. Additionally, the focus on oncology as a topic of research can be explained by multiple reasons such as the contribution of cancers as the second-most leading cause of deaths worldwide and thus a wide availability of data, but also due to the fact that AI papers focusing on cancers tend to be highly cited and this can act as an additional motivating factor.[13] Nonetheless, these findings illustrate areas where institutions, authors, and journals can fill gaps in the literature, bringing the field closer to clinically significant and feasible AI/ML solutions which could ultimately enhance patient care as a whole, but particularly for pediatric and non-oncology patients.

Furthermore, our analysis across Type 2 articles demonstrated that data were derived from only 12 nations with about half of these articles stemming from the United States indicating geographical centralization. This is consistent with the fact that one-third of all publications on AI in medicine and half of all AJNR publications originate from the United States.[14] The second largest number of analyzed publications derived from South Korea, which differs from previous AI in medicine data.[1] The centralized nature of the articles and the minimal use of public data sources emphasize the need for increased exposure and education in AI and ML for medical trainees, researchers, and physicians around the world. This would hopefully increase international collaboration, propelling the field forward with an increased availability of new and diverse ideas and best practices, but also allow researchers to externally validate and test their models and increase the size of their datasets—increasing the chances of outcomes being replicable, reliable, generalizable, and clinically significant, ultimately promoting the translation of research into practice.

Our results showed an average of 9-10 authors per article which is larger than the 5 and 8 authors reported for publications on AI in medicine and AJNR respectively.[1,14] This can partly be explained by the need for larger datasets, external validation, and testing samples, therefore encouraging researchers to seek out collaborators at other institutions. It also highlights the increasing need for multidisciplinary collaboration amongst researchers. Additionally, we found that radiologists led 63% of the publications and interestingly, 60% held additional degrees such as PhDs and MBAs. These PhD and MBA degreeholding radiologists have a broader knowledge base and may indicate pre-existing technical skills in domains such as computer science, or an interest in entrepreneurial ventures and health systems leadership. These results demonstrate that although there is room for improvement, the field of neuroradiology is currently doing well with respect to authors' collaboration and diversity of expertise and training backgrounds.

Our evaluation of Type 2 articles revealed weaknesses that may slow down the advancement of the field, and thus implementation of AI/ML solutions into everyday practice. Results showed that only a minority of articles we analyzed were prospective, multi-centered, or externally validated, which are important indicators for an AI model's real-world clinical application.[5,15] The lack of external validation may be related to limited data access, lack of time or funding, but also potential misprioritization focused on proving a theory-of-concept rather than the production of a practical solution that has the potential to be implemented. Researchers may find it easier to conduct and ultimately publish Type 2 compared to Type 3 articles as publishing the latter would require a functional AI/ML tool that can be applied in some way. Going forward, this finding represents an area where journal publishers and editors can contribute to the field by encouraging the implementation of important indicators of generalizability into study methodologies.

The trust that is required for an effective therapeutic relationship between the patient and healthcare system is also a requirement for AI applications. In our assessment of potential bias and explainability, we found that only 22% and 16% of articles adequately addressed bias and explainability, respectively. We did however find that Type 2 articles published since 2020 were slightly more likely to have adequately addressed bias and explainability. Although attention towards these important topics has been increasing recently, there is marked room for improvement before trust in AI is established let alone widely integrated into our everyday healthcare.[16]

Our results and their interpretation are limited to AJNR publications. It is possible that findings across all relevant medical imaging journals, or other methodological designs such as a qualitative study examining the experiences and perceptions of AI/ML among neuroradiologists may yield different conclusions. Article inclusion and categorization relied upon published definitions the authors used, and this can ultimately affect the results and analysis of this study. The factors we found to be limiting or contributing to the advancement of the field towards more integration-focused research was primarily gleaned from our analysis of Type 2 articles. However, we acknowledge factors outside our scope of analysis such as regulatory challenges, funding constraints, and medico-legal issues may also hinder AI/ML integration.

CONCLUSIONS

Our investigation confirmed the rapid growth of AI research in neuroradiology. An in-depth analysis demonstrated that AI/ML developmental research work without direct clinical integration is the primary driver of publication growth. Improving the quality of this research work will be crucial to promote their integration into day-to-day practice. These results ultimately provide authors, editors, clinicians, and policymakers important insights to promote a shift towards integrating practical AI/ML solutions in neuroradiology.

ACKNOWLEDGMENTS

None.

REFERENCES

1. Tran BX, Vu GT, Ha GH, et al. Global Evolution of Research in Artificial Intelligence in Health and Medicine: A Bibliometric Study. *J Clin Med* 2019;8:360. DOI: https://doi.org/10.3390/jcm8030360

2. Tang A, Tam R, Cadrin-Chênevert A, et al. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Can Assoc Radiol J* 2018;69:120-135. DOI: https://doi.org/10.1016/j.carj.2018.02.002

3. Kocak, B., Baessler, B., Cuocolo, R. et al. Trends and statistics of artificial intelligence and radiomics research in Radiology, Nuclear Medicine, and Medical Imaging: bibliometric analysis. *Eur Radiol* 2023 June 14. DOI: https://doi.org/10.1007/s00330-023-09772-0

4. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur Radiol Exp* 2018;2(1):35. DOI: https://doi.org/10.1186/s41747-018-0061-6

5. Wagner DT, Tilmans L, Peng K, et al. Artificial Intelligence in Neuroradiology: A Review of Current Topics and Competition Challenges. *Diagnostics* 2023;13(16). DOI: https://doi.org/10.3390/diagnostics13162670

6. Lui YW, Chang PD, Zaharchuk G, et al. Artificial Intelligence in Neuroradiology: Current Status and Future Directions. *AJNR Am J Neuroradiol* 2020;(8):E52-E59. DOI: https://doi.org/10.3174/ajnr.A6681

7. Erickson BJ, Korfiatis P, Akkus Z, et al. Machine Learning for Medical Imaging. *Radiographics* 2017;37:505-515. DOI: https://doi.org/10.1148/rg.2017160130

8. Wagner MW, Namdar K, Biswas A, et al. Radiomics, machine learning, and artificial intelligence-what the neuroradiologist needs to know. *Neuroradiology* 2021;63:1957-1967. DOI: https://doi.org/10.1007/s00234-021-02813-9

9. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer* 2018;18:500-510. DOI: https://doi.org/10.1038/s41568-018-0016-5

10. Lu Y. Artificial Intelligence: A survey on evolution, models, applications and future trends. *Journal of Management Analytics* 2019;6:1-29. DOI: https://doi.org/10.1080/23270012.2019.1570365

11. Daldrup-Link H. Artificial intelligence applications for pediatric oncology imaging. *Pediatr Radiol* 2019;49:1384-1390. DOI: https://doi.org/10.1007/s00247-019-04360-1

12. Ramgopal S, Sanchez-Pinto LN, Horvat CM, et al. Artificial intelligence-based clinical decision support in pediatrics. *Pediatr Res* 2023;93:334-341. DOI: https://doi.org/10.1038/s41390-022-02226-1

13. Guo Y, Hao Z, Zhao S, et al. Artificial Intelligence in Health Care: Bibliometric Analysis. *J Med Internet Res* 2020;22(7):e18228. DOI: https://doi.org/10.2196/18228

14. Zohrabian VM, Staib LH, Castillo M, et al. Scientific Collaboration across Time and Space: Bibliometric Analysis of the American Journal of Neuroradiology, 1980-2018. *AJNR Am J Neuroradiol* 2020;41:766-771. DOI: https://doi.org/10.3174/ajnr.a6523

15. Kim DW, Jang HY, Kim KW, et al. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. *Korean J Radiol* 2019;20:405-410. DOI: https://doi.org/10.3348/kjr.2019.0025

16. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* 2021;23:18. DOI: https://doi.org/10.3390/e23010018

TABLES

None.

FIGURES

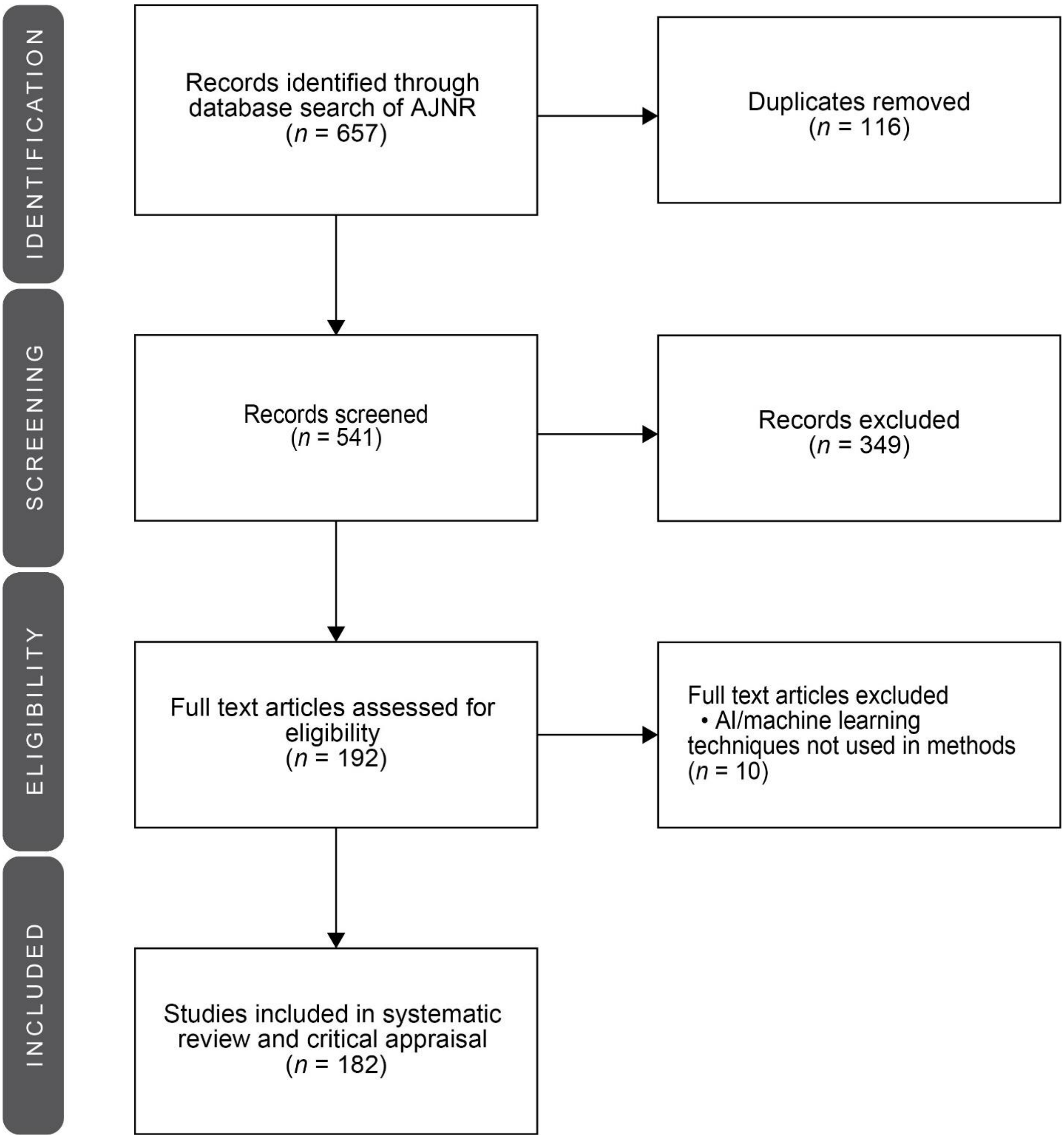

FIG 1. Flow chart of the article screening process.

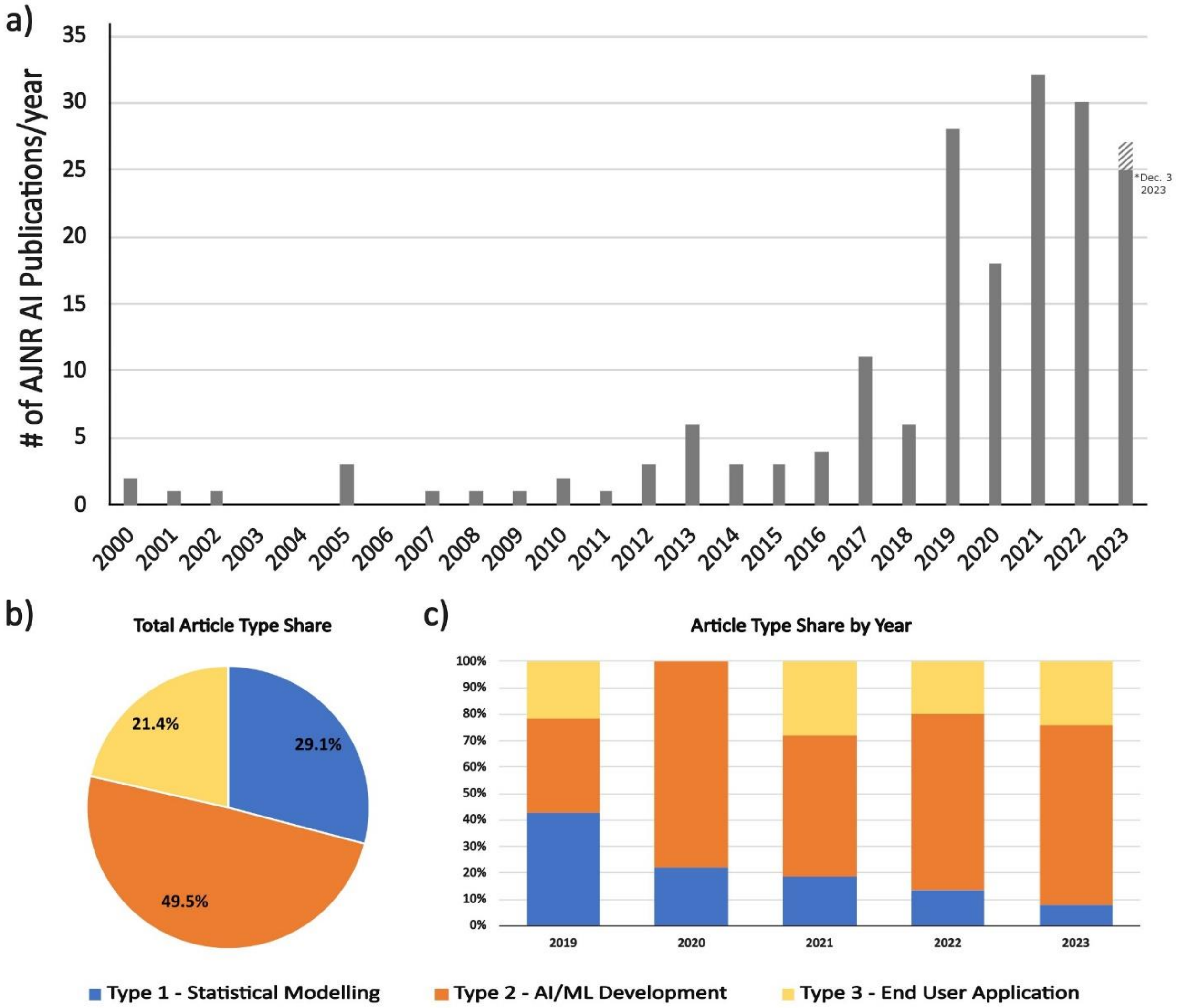

FIG 2. Global Results (Type 1, 2, and 3). Data characterizing the growth of all AI/ML articles (i.e., Type 1, 2, and 3). a) The number of all articles published from the first year an AI/ML article was published in the AJNR in 2000 to the article search date on Dec. 3rd, 2023. *Striped bar in 2023 indicates the extrapolated number of articles published in that year should the search have occurred at the end of the year. b) Pie chart breaking down the share each article type holds among all AI/ML publications. c) Percentile bar graph demonstrating article type share in the last five years.

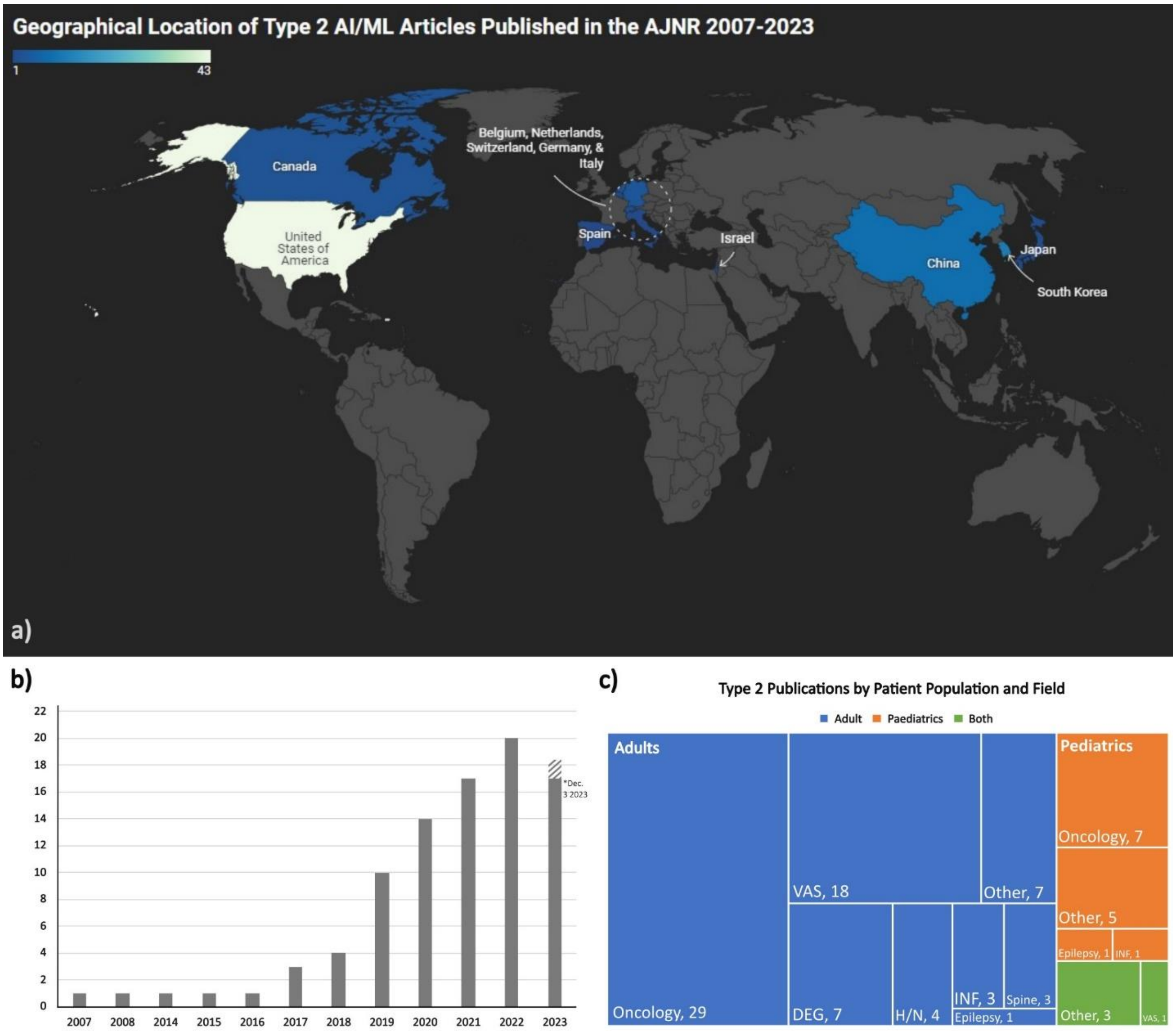

FIG 3. Qualitative Type 2 results. General overview of Type 2 publication characteristics (n = 90). a) A world map showing where studies were conducted. b) Bar graph demonstrating the number of Type 2 articles published in years with at least 1 article published. *Striped bar in 2023 indicates the extrapolated number of articles published in that year should the search have occurred at the end of the year. c) Graphic display of the patient populations and topics studied. VAS = Vascular, DEG = Degenerative, H/N = Head & Neck, INF = Inflammation.

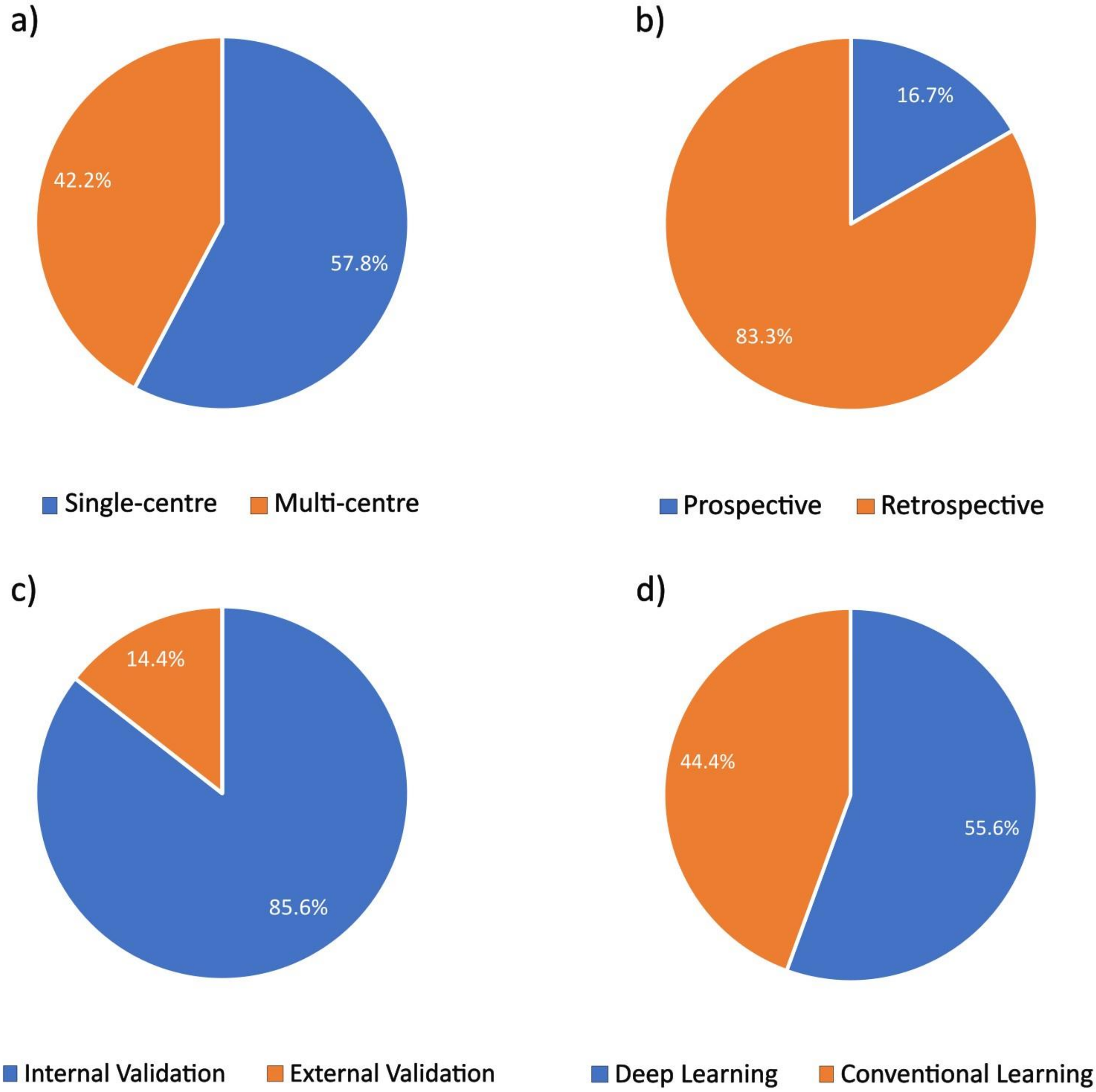

FIG 4. Evaluative Type 2 results. Methodological characteristics of Type 2 publications (n = 90). Pie charts demonstrating the percentage of Type 2 articles that were (a) single- vs. multi-centered, (b) prospective vs. retrospective, (c) externally validated or internally validated their models, and (d) utilized deep learning to train their models or only conventional learning techniques.

SUPPLEMENTAL FILES

Definitions Used for Screening Articles

AI: "…can learn the threshold (or patterns) through examination of a large number of images using parts or the whole image and contextual information depending on the technique used" (Wagner al., 2021). Note that images or other forms of input could be used (e.g. text).

ML: "…field of AI in which algorithms are trained using known datasets, from which the machine "learns." The developed algorithm then applies this knowledge to perform diagnostic tasks in unknown datasets" (Wagner al., 2021).

"…algorithm is applied to a data set (in our example, tumor images) and to some knowledge about the data set (in our example, benign or malignant tumors), then the algorithm system can learn from the training data and apply what it has learned to make a prediction. If the algorithm optimizes its parameters such that its performance improves—that is, more test cases are diagnosed correctly—then it is considered to be learning that task" (Erickson et al., 2017).

Additional Definitions for Terms Used in the Article

Training dataset: "The training dataset is used to optimize the machine learning model through minimizing loss (the error of predicted results compared to ground truth)." (Wagner al., 2021).

Test dataset: "Once the model is trained, a test dataset is used to test the performance of the model." (Wagner al., 2021).

"Eventually, several performance metrics assess the model's performance using the prediction on the test dataset. These metrics commonly include sensitivity, specificity, accuracy, precision, recall, and F1 score." (Wagner al., 2021).

Validation dataset: "The validation dataset is a separate cohort that is used to fine tune the model." (Wagner al., 2021).

"The validation dataset is a separate cohort that is used to fine tune the model. In conventional machine learning algorithms such as SVM and RFs, the hyperparameters (e.g., number of trees in RF) must be fine-tuned using a validation dataset to maximize the performance on the test set. In deep learning, the validation cohort is used to determine the stoppage point for the training (optimization) of the model." (Wagner al., 2021).

External validation datasets: "…external validation and testing datasets, i.e., datasets acquired at a different institution, are preferable as they allow for a more robust evaluation of the developed algorithm." (Wagner al., 2021).