Shahid Sadoughi University of Medical Sciences School of Public Health

Journal of Social Behavior and Community Health

Journal of Social Behavior & Community Health | Social Determinants of Health Research Center

The Critical Role of Programming Languages among Healthcare Data Scientists: A Systematic Review of Trends, Applications, and Future Directions

Seyedeh Nahid Seyedhasani [a], **Afrooz Arzehgar** [a], **Mohamad Amin Bakhshali** [a], **Mostafa Sohrabifar** [a], **Seyyed Mohammad Tabatabaei** [a,b,*]

[a] Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[b] Applied biomedical Research Center, Basic Sciences Research Institute, Mashhad University of Medical Sciences, Mashhad, Iran

A R T I C L E I N F O

SYSTEMATIC REVIEW

Article History:
Received: 26 July 2025
Revised: 31 August 2025
Accepted: 1 October 2025

***Corresponding Author:**
Seyyed Mohammad Tabatabaei

Email:
Tabatabaeimh@mums.ac.ir

Tel: +989153082261

Citation:
Seyedhasani SN, Arzehgar A, Bakhshali MA, Sohrabifar M, Tabatabaei SM. The Critical Role of Programming Languages among Healthcare Data Scientists: A Systematic Review of Trends, Applications, and Future Directions. Journal of Social Behavior and Community Health (JSBCH). 2025; 9(2): 1735-1747.

A B S T R A C T

Background: Artificial intelligence (AI) and data science have transformed healthcare by enabling advanced analytical techniques. AI-driven solutions rely on sophisticated algorithms that require specialized programming languages. Understanding the most commonly used programming languages is essential for healthcare data scientists in order to navigate this domain effectively. This study explores the trends and applications of programming languages in healthcare data science, highlighting their roles in machine learning (ML) and related methodologies.

Methods: A systematic search was conducted in PubMed/MEDLINE, Scopus, and Web of Science (WoS) covering the period 2010–2023. Keyword combinations included artificial intelligence, machine learning, programming languages, healthcare, and medical informatics. After screening, 174 studies that explicitly mentioned programming languages in their abstracts were included for analysis.

Results: Public health accounted for 50.6% (n=88/174) of the reviewed studies, followed by medicine at 25.9% (n=45/174) and genomics at 14.4% (n=25/174). Python emerged as the most widely used programming language, appearing in 37.47% (n=65) of the articles, followed by R at 29.6% (n=51) and MATLAB at 17.8% (n=31). Machine learning methods were predominant in genomics and epidemiology. The temporal trend showed an increasing preference for Python, while MATLAB use declined in recent years.

Conclusion: The selection of programming languages in healthcare data science is influenced by technical needs, application-specific requirements, and collaboration dynamics. Python's versatility has made it a dominant choice, while R's statistical focus and MATLAB's specialized toolkits remain significant in specific domains. The findings provide a framework for educational strategies, guiding data scientists in making informed decisions about language proficiency. Future research should evaluate the long-term implications of programming language adoption on healthcare analytics.

Keywords: Artificial Intelligence, Machine Learning, Programming languages, Data Science, Health Informatics.

[DOI: 10.18502/jsbch.v9i2.20265]

Introduction

Artificial intelligence (AI) has revolutionized numerous industries including healthcare by enhancing patient care, improving quality of life, and streamlining operations (Alowais et al., 2023). AI systems rely on machine learning (ML) and deep learning (DL) methodologies to uncover patterns in complex and unstructured data, offering unprecedented opportunities for clinical decision support, predictive modeling, drug development, and healthcare cost reduction (Kalra et al., 2024). These advancements are fueled by the availability of big data from electronic health records (EHRs), genomics, and wearable devices, alongside improvements in computational and analytical tools(Islam et al., 2024).

Within healthcare, data science has emerged as a critical discipline, bridging the gap between raw data and actionable insights. Data scientists employ AI-driven approaches to manage, analyze, and interpret vast amounts of healthcare data (Subrahmanya et al., 2022). However, the implementation of these approaches depends significantly on the choice of programming languages, which serve as foundational tools for developing algorithms, processing data, and creating models. A well-informed choice of programming language is essential to address specific healthcare challenges effectively (Chambers, 2020). Yet, there has been limited research to date on the adoption and application patterns of programming languages within healthcare data science (Ahsan et al., 2022; Innes et al., 2018).

This study focuses on understanding the programming languages commonly used by healthcare data scientists. By examining their prevalence in various domains such as public health, genomics, and medicine, the authors aim to guide professionals and educators in selecting appropriate tools for research and learning. The findings will enable data scientists to align their technical skills with the demands of the healthcare sector, thereby enhancing their ability to contribute to advancements in the field.

To identify and analyze these trends, this study conducted a systematic review of the scientific literature and performed a content analysis of the abstracts to determine the frequency and co-occurrence of specific programming languages in various healthcare contexts.

Programming languages are not merely technical instruments but also strategic choices influenced by ecosystem maturity, community support, and application-specific needs. Python, R, MATLAB, and others have distinct strengths, making them suitable for specific tasks within healthcare. For instance, Python's extensive libraries and community support make it ideal for ML, while R's statistical prowess suits data mining (DM). MATLAB's capabilities in signal processing and modeling find applications in diagnostic studies. The evolution of these languages reflects their adaptation to the growing complexities of healthcare data science.

This research identifies trends in programming language usage across healthcare domains, emphasizing their role in facilitating data analysis methods like ML, data mining (DM), and DL. By exploring these patterns, the authors seek to establish a framework for selecting programming languages that align with the goals of healthcare projects. Ultimately, this study underscores the importance of programming literacy among healthcare data scientists and highlights the need for targeted educational initiatives to equip professionals with relevant skills for this dynamic and impactful field.

Methods

This study conducted a comprehensive systematic review to identify the application of programming languages in healthcare data analytics, particularly in domains employing ML, DL, and DM methodologies. The search was performed across three primary databases: PubMed/MEDLINE, Scopus, and Web of Science (WOS) on August 3, 2023, with no restrictions on the publication year, allowing for a comprehensive assessment of trends and usage patterns over time.

[DOI: 10.18502/jsbch.v9i2.20265]

To ensure inclusivity and comprehensiveness, a structured and systematic approach was utilized to develop the search strategy.

The keywords were generated using Medical Subject Headings (MeSH) terms in PubMed by examining related articles in the field. The selected keywords and their synonyms were categorized under three main groups: analytical methods, healthcare, and programming languages. Synonyms and alternative terminologies were considered to cover variations in terminologies used across different studies. For example, synonyms like "data analytics," "knowledge discovery," and "data processing" were included alongside "machine learning" and "artificial intelligence." For healthcare-related terms, both "health" and "healthcare" were included.

The final search strategy was as follows:

[Analytics OR "data analytics" OR "data analysis" OR "analysis of data" OR "data mining" OR "knowledge discovery" OR "data processing" OR "data science" OR "machine learning" OR "artificial intelligence" OR " deep learning"]AND [health OR "healthcare"] AND [Python OR R OR SQL OR Java OR Julia OR Scala "C" OR "C++" OR "JavaScript" OR Swift OR Go OR MATLAB OR SAS].

To refine the retrieved articles, specific inclusion and exclusion criteria were established to ensure relevance and quality. The inclusion criteria encompassed original research articles that explicitly mentioned the programming language(s) used in healthcare-related data science. Articles that provided information on the application of programming languages in data analytics with a clear focus on healthcare domains were considered. Only peer-reviewed studies published in English were included, with no restrictions on publication year, geographic region, or specific health domain. Conversely, the exclusion criteria ruled out review articles, conference abstracts, book chapters, theoretical papers, dissertations, and letters to editors. Articles that did not explicitly mention the programming language used in the abstract or full text and the studies which were not written in English were excluded.

Eligible articles underwent a rigorous review process independently conducted by three authors to ensure strict adherence to the inclusion criteria. The review process was divided into two phases: title screening and abstract screening. Titles were initially screened for relevance to programming languages and healthcare data science, and the abstracts of shortlisted articles were subsequently examined to confirm the inclusion of programming languages and their relevance to healthcare applications. Discrepancies during the selection process were resolved through discussions among the reviewers, and in cases where disagreements persisted, a fourth reviewer was consulted for final arbitration.

Several key categories of information were extracted from the eligible articles. Domain categories were identified, including public health, genomics, medicine, pharmaceutical studies, health information systems, and dental studies. Subdomains, such as epidemiology, diagnostics, and social networks, were specified where applicable. Programming languages used in the studies, such as Python, R, MATLAB, Java, Weka, C++, C#, and SAS, were also documented. Data analysis methods, including ML, DM, and DL were noted. Geographic information, specifically the country of the lead author, and the year of publication were recorded to analyze temporal trends. Instances where multiple programming languages were used within the same study were also documented to understand their combined application.

The database search for this study was conducted on August 3, 2023. There were no restrictions on the publication year, allowing for a comprehensive assessment of trends and usage patterns over time. Descriptive and inferential statistical methods were used to analyze the collected data. Frequency distribution was calculated to determine the percentage of articles in each domain category and programming language. Temporal trends were then assessed to observe patterns in the adoption of programming

[DOI: 10.18502/jsbch.v9i2.20265]

languages and data analysis methods over time. Co-occurrence analysis was employed to visualize relationships between programming languages, domain categories, and subdomains through network diagrams. Finally, chi-square tests were conducted to evaluate the relationships between data analysis methods and programming languages, as well as between domain categories and programming languages, with a *p*-value of less than 0.05 which was considered statistically significant.

Despite the rigor of these methods, certain limitations in the methodology should be acknowledged. Only articles explicitly mentioning programming languages in the abstract or full text were included, which may have excluded studies where this information was implicit. Additionally, the exclusion of non-English studies might have limited the geographic and cultural diversity of the findings. The search strategy relied on three databases, which, although comprehensive, may not encompass all relevant literature, particularly in niche or non-indexed journals. By adhering to these rigorous methods, this study offers a robust and detailed exploration of programming language trends in healthcare data science, providing valuable insights for researchers and practitioners in the field.

Results

A total of 174 studies employing DM and ML methods in various health domains were included in this review, with the abstracts explicitly indicating the programming languages used. These studies were conducted across 43 countries, based on the country of the responsible author. As illustrated in **Figure 1**, the U.S. contributed the highest number of studies (*n*=40), followed by India (*n*=32) and China (*n*=20). Iran contributed 8 studies, while Saudi Arabia accounted for 7 studies. Turkey, Germany, and the United Kingdom each contributed 5 studies, and South Korea was responsible for 4 studies(Figure 1)

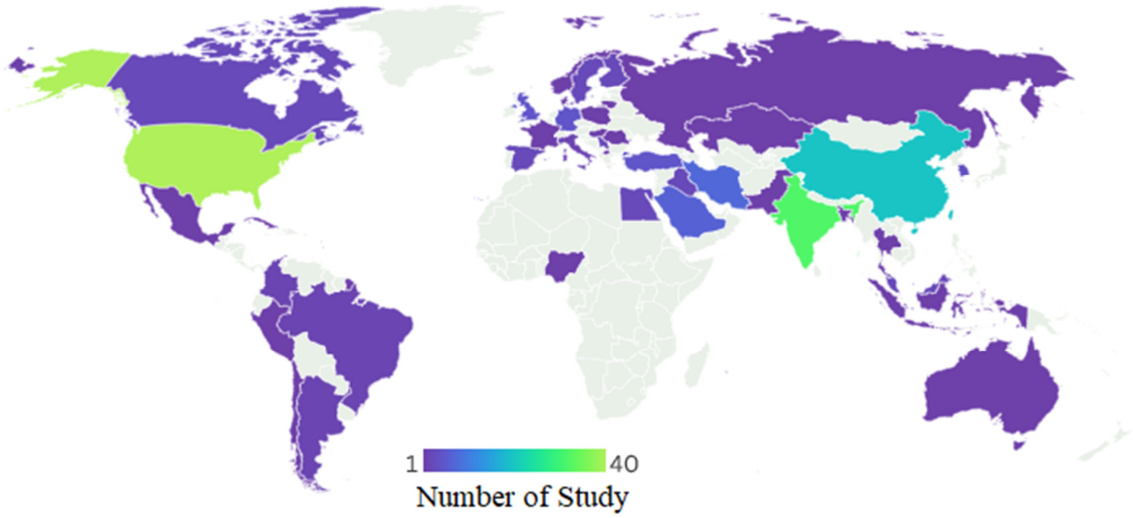

Figure 1. Distribution of Studies by the Country of the Responsible Author.

Overview

The distribution of studies across domain categories is presented in **Figure 2**. Public health accounted for the largest share, representing 51% (*n*=88/174) of the articles, followed by medicine with 26% (*n*=45/174). Genomics represented 14% (*n*=25/174), pharmaceutical studies 4% (*n*=7/174), health information systems 3% (*n*=5/174), and dental-related studies 2% (*n*=4/174)(Figure 2).

[DOI: 10.18502/jsbch.v9i2.20265]

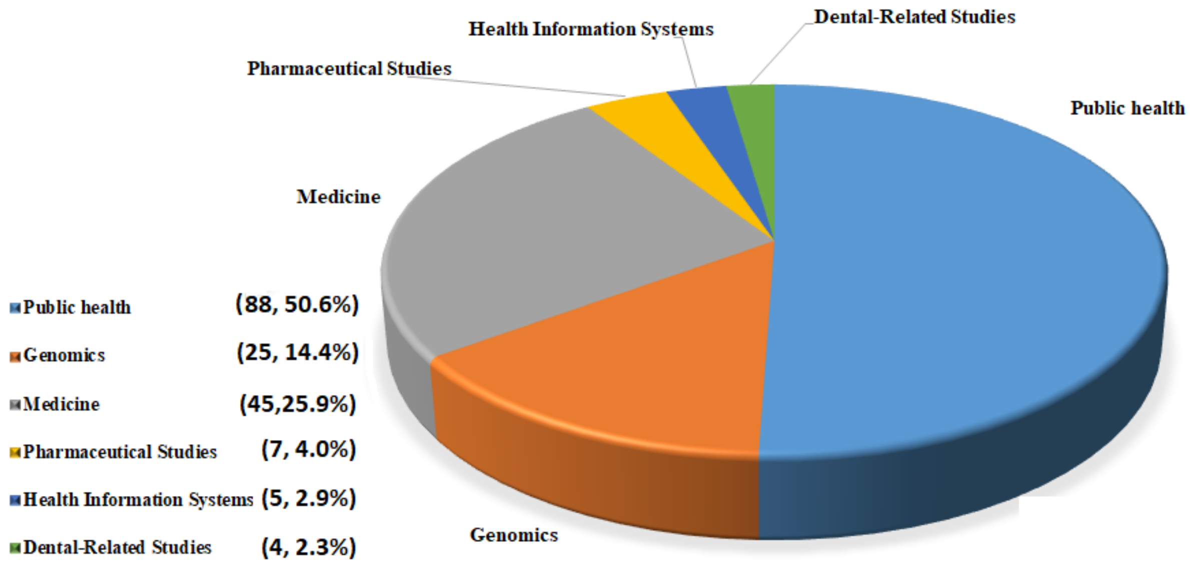

Figure 2. Percentage and number of articles in each domain category (*n*: Number of Articles; %: Percentage).

To avoid inflation of adoption rates, the findings differentiate between (a) the number of unique studies that used a programming language, and (b) the total number of mentions of that language across all studies. Regarding programming languages, Table 1 shows that Python was the most frequently used, appearing in 37% (*n*=65) of the articles. R and MATLAB followed with 30% (*n*=51) and 18% (*n*=31), respectively. Two studies reported the use of multiple programming languages simultaneously: one combined MATLAB and Python, while the other used both R and Python.(Table 1).

Table 1. Percentage and number of articles utilizing each programming language

Method PL	ML (*n*, %)	DM (*n*, %)	DL (*n*, %)	Total (*n*, %)
Python	(48, 28%)	(9, 5%)	(8, 5%)	(65, 37%)
R	(33, 19%)	(14, 8%)	(4, 2%)	(51, 30%)
MATLAB	(18, 10%)	(5, 3%)	(8, 5%)	(31, 18%)
JAVA	(7, 4%)	(7, 4%)	(1, 1%)	(15, 9%)
Weka	(4, 2%)	(0, 0%)	(0, 0%)	(4, 2%)
C++	(2, 1%)	(0, 0%)	(1, 1%)	(3, 2%)
C#	(1, 1%)	(1, 1%)	(0, 0%)	(2, 1%)
SAS	(0, 0%)	(1, 1%)	(0, 0%)	(1, 1%)
Combined	(2, 1%)	(0, 0%)	(0, 0%)	(2, 1%)
Total	(115, 66%)	(37, 21%)	(22, 13%)	(174, 100%)

ML: Machine learning; DM: Data mining; DL: Deep learning; *n*: Number of Articles; %: Percentage; PL: Programming language

[DOI: 10.18502/jsbch.v9i2.20265]

Temporal relationship

Temporal trends in the usage of programming languages are depicted in **Figure 3-a**. The oldest study in this review utilized R, which has shown growth over the years but experienced a decline in 2021. Python was first introduced in studies in 2016, peaking in 2021, and has continued to grow annually. MATLAB usage peaked in 2020, with its introduction dating back to 2004. However, a decline in MATLAB usage is observed post-2020. Other programming languages were employed sporadically and to a lesser extent.

Figure 3-b illustrates the temporal trends in data analysis methods. DM methods have been in use since 1999, while ML methods, introduced in 2004, have exhibited significant growth in recent years, accounting for the majority of studies. DL methods began in 2017 and have an upward trend, although a decline is noted in 2022. Among domain categories**, Figure 3-c** shows that public health and medicine have experienced growth in recent years. **Figure 3-d** indicates a stable upward trend in diagnostic studies, while other categories remain sporadic, irregular, or limited over time. Categories with fewer than three articles were excluded from this visualization(Figure 3)

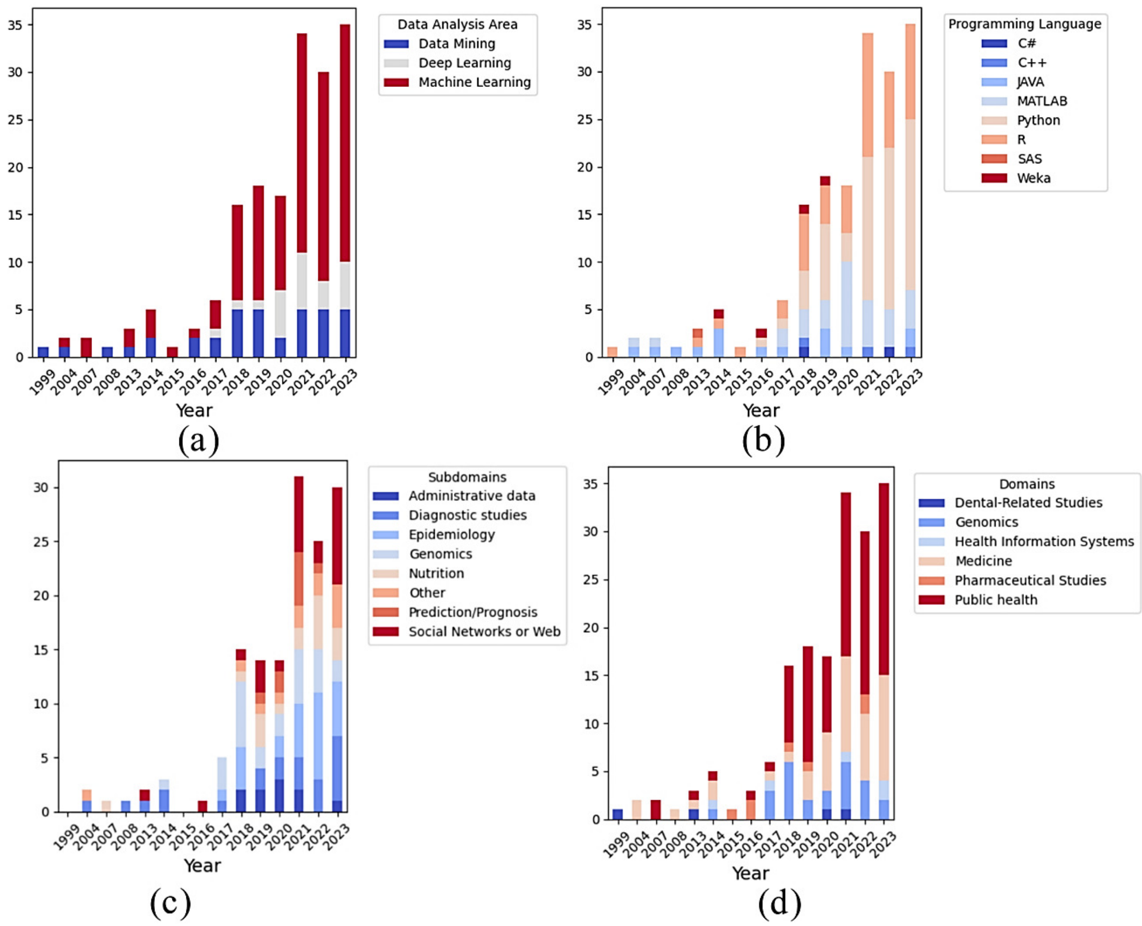

Figure 3: (a) Temporal Analysis of Programming Languages, (b) Data Analysis Methods, (c) Specialized Areas, and (d) Domain Categories.

Programming languages

As shown in Figure 4, Python was the most frequently mentioned language, appearing in 67 instances across the included articles. R and MATLAB were also commonly mentioned, with 52 and 32 instances, respectively. It is important to note that Figure 4 represent the number of times a language was mentioned rather than the number of unique studies, as two studies reported the use of multiple programming languages simultaneously:

JSBCH. Volume 9, Issue 2, Nov 2025; 1135-1147

[DOI: 10.18502/jsbch.v9i2.20265]

one combined MATLAB and Python, while the other used both R and Python. A significant relationship exists between data analysis methods and programming languages (p-value < 0.05) (Figure 4)

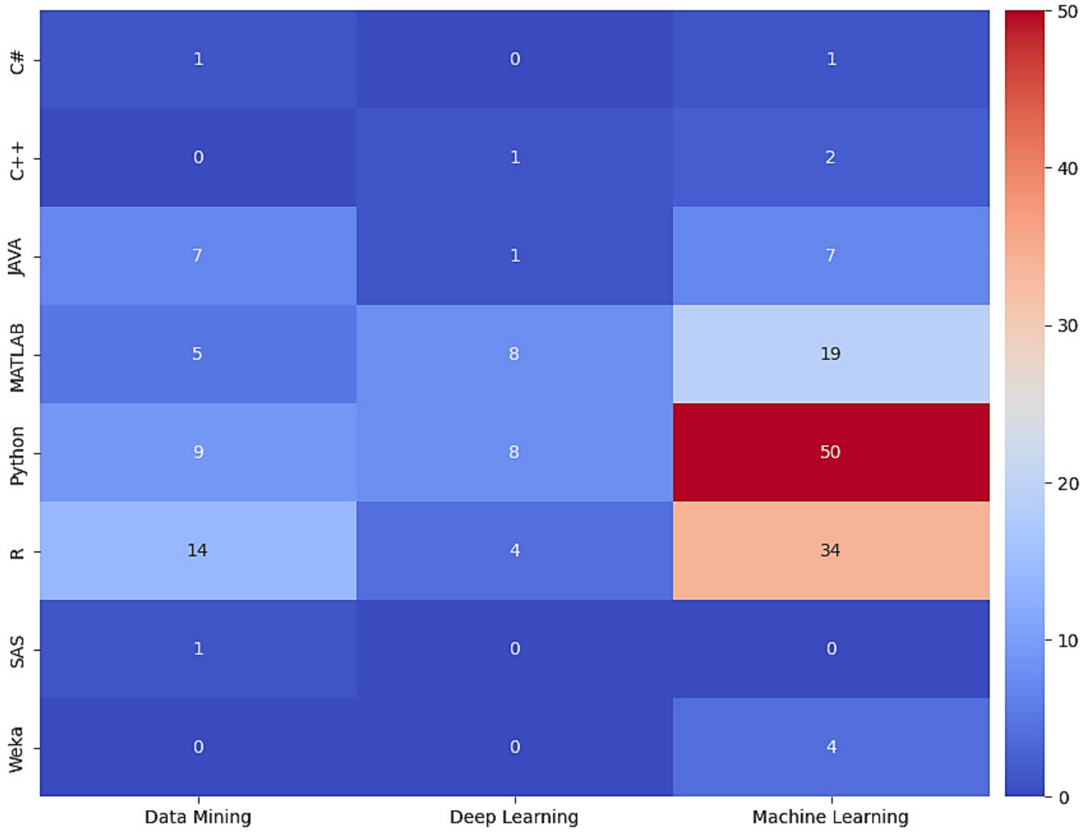

Figure 4. Co-occurrence relationship between data analysis methods and programming languages

Figure 5 demonstrates the co-occurrence relationships between domain categories and programming languages. Python was predominantly used in public health, followed by R and MATLAB. In medicine, Python also dominated in usage, with MATLAB being the second most commonly used. In genomics, R emerged as the most frequently used language. The relationship between domain categories and programming languages was statistically significant (p-value < 0.05).

Regarding subdomains, Python was most frequently applied in social media and web studies (n=13), followed by R (n=8). Python's second most common application was in epidemiology (n=10), with R (n=8) and MATLAB (n=7) also being prominent. The relationship between specialties and programming languages was similarly significant (p-value < 0.05)(Figure 5).

[DOI: 10.18502/jsbch.v9i2.20265]

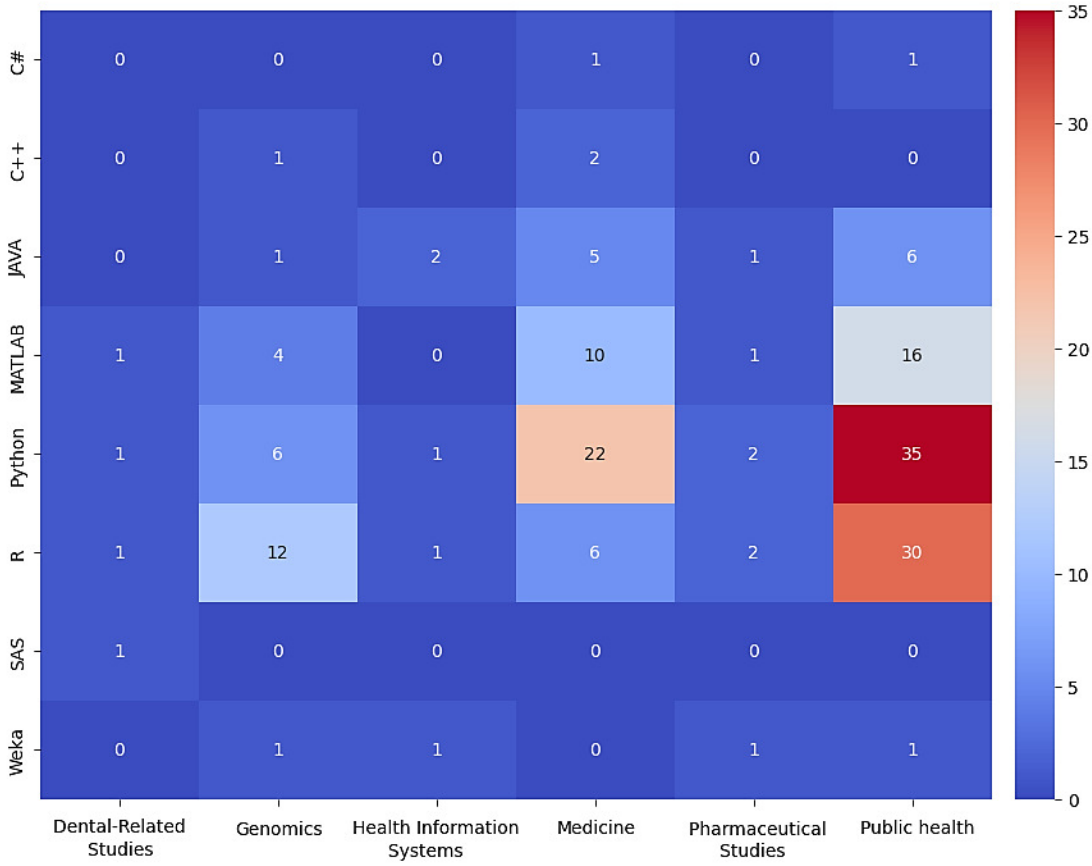

Figure 5. Co-occurrence relationship between domain categories and programming languages

Data analysis methods

Figure 6 highlights the distribution of data analysis methods. ML was the most prevalent, appearing in 66% (*n*=115) of the studies. Its highest application was observed in public health, followed by medicine. DM was also predominantly applied in public health. In terms of subdomains, ML was most commonly used in genomics (*n*=19) and epidemiology (*n*=19), followed by social media and web studies (*n*=16). Diagnostics also showed notable use of ML (*n*=12) (Figure 6).

[DOI: 10.18502/jsbch.v9i2.20265]

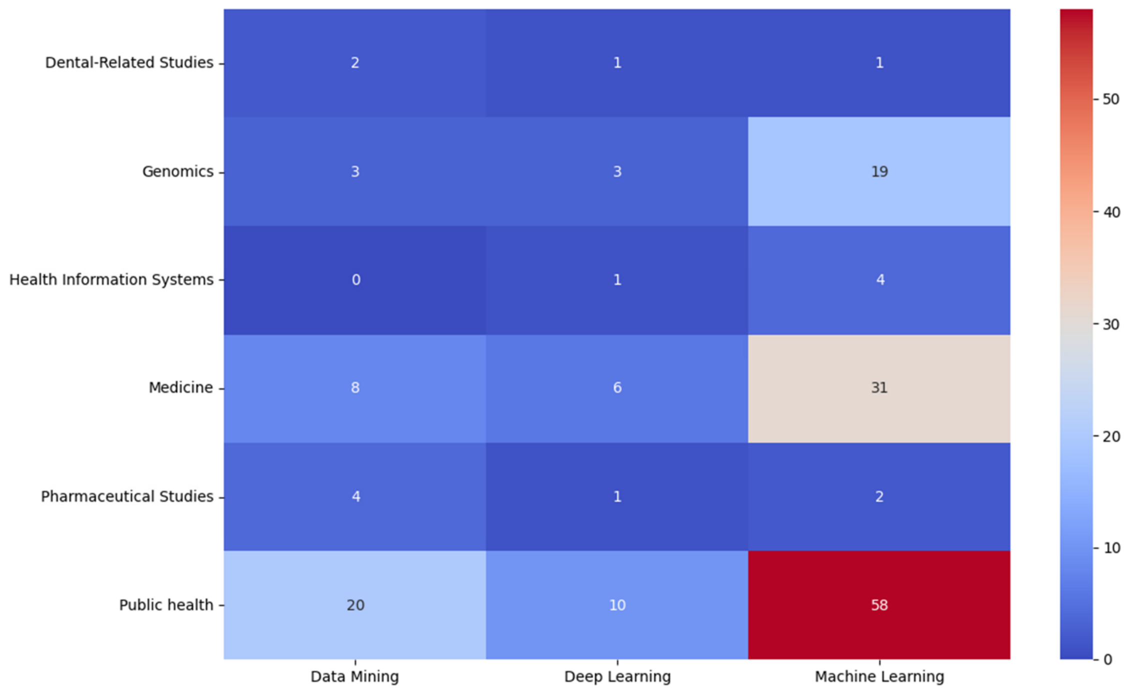

Figure 6. Co-occurrence relationship between domain categories and data analysis methods

Discussion

In recent years, advancements in technology, the expansion of big data, and the increasing need for accurate pattern recognition in decision-making have significantly impacted the fields of health and medicine (Shajii et al., 2021). These developments have not only enhanced health-related processes but also enabled professionals to gain deeper insights into the factors affecting human health through advanced computational techniques (Baytaş et al., 2023). The adoption of specific programming languages to meet these objectives and establish robust infrastructures has become indispensable. In this discussion, the findings are examined across domain categories and specialized areas, considering various data analysis methods and programming languages.

The diverse international participation in this field highlights the growing global interest in health informatics and the potential for cross-border collaboration to address pressing health challenges through data-driven solutions. As computational methods and hardware capabilities improve, studies focusing on advanced analytics in health and medicine have shown notable growth (Wang et al., 2023; Yu et al., 2024). Public health remains the most widely utilized domain, while specialized areas such as social media, web analytics, epidemiology, and genomics have gained prominence. In public health and epidemiology, the analysis of big data has been facilitated by the evolution of analytical methods and advancements in technologies like EHR and genomic sequencing. The increase in these studies underscores the pivotal role of data science in these domains (Aichner et al., 2021).

The rise of social media and web-based-content analysis has further enriched the field, providing valuable sources for data collection and analysis through tools such as recommendation systems and natural language processing. Diagnostic studies, another critical domain, have experienced consistent growth due to improved access to diagnostic data and the increasing sophistication of ML and DL techniques. The upward trend in publications related to medicine reflects the integration of artificial intelligence and data science within the healthcare domain. The observed decline in the number of articles in 2022 may be attributed to shifts in focus toward generative modeling and other emerging AI

[DOI: 10.18502/jsbch.v9i2.20265]

approaches not captured in this study.

Programming languages play a critical role in supporting DM, ML, and DL applications. Usage patterns have evolved over time, with Python emerging as the most widely adopted programming language in recent years (Fourment & Gillings, 2008). Its popularity can be attributed to its versatility, extensive library ecosystem, open-source nature, and a robust community that provides abundant resources for training and development. Conversely, MATLAB, historically favored for numerical computations and specialized toolboxes, has seen a decline in usage not merely as a passing trend but as part of a broader academic and industry realignment toward open-source alternatives. This strategic shift likely reflects several factors: changing community preferences favoring more flexible and collaborative tools, limitations in MATLAB's scalability for large datasets, and competition from open-source languages such as Python and R, which provide extensive libraries, active communities, and cost-effective solutions.

R has maintained steady adoption across studies, valued for its domain-specific capabilities in statistics and DM. Its active community of statistical experts and diverse libraries for visualization and analysis continue to make it a popular choice.

Languages like Python, with its support for rapid prototyping and iterative development, are particularly advantageous in dynamic healthcare environments. These languages enable the management of large datasets generated by EHR systems and wearable devices, enhancing data integration and decision-making processes. MATLAB, although declining in overall usage, remains a preferred choice for diagnostic studies due to its strong capabilities in signal processing, image analysis, and biological modeling (Ekmekci et al., 2016; Virtanen et al., 2020).

The alignment between data analysis methods and programming languages is evident in the study. Python dominates ML applications, while R is more prevalent in DM. Domain-specific trends reveal that Python is extensively used in public health and medicine, R leads in genomics, and MATLAB is most frequently applied in diagnostic studies (Chambers, 2020). These patterns are consistent with the technical strengths of each language in addressing the unique requirements of these domains(Sepulveda, 2020).

ML has become the predominant data analysis method in healthcare, particularly in domains such as public health, medicine, and genomics. In contrast, pharmaceutical studies primarily rely on R, reflecting the statistical nature of analyses in this field. The growing reliance on ML across domains highlights its capacity to extract actionable insights and inform decision-making in healthcare.

A notable limitation of this study is the restriction to articles explicitly mentioning programming languages in their abstracts. Additionally, the focus on English-language studies and the use of only three major databases (PubMed/MEDLINE, Scopus, and WoS) may slightly limit the comprehensiveness and generalizability of the findings. While these choices helped maintain focus and feasibility, future research could expand inclusion criteria and consider additional databases to capture a broader perspective on programming language usage in healthcare data science (Akan & Chaparro, 2024; Kimiafar et al., 2023)

The findings of this study have important implications for education and research in health sciences.

Educational institutions should prioritize Python as the core programming language for healthcare data science curricula, while offering specialized modules on R for advanced statistical modeling and on MATLAB for domains such as signal and image processing. This layered approach ensures that graduates are both versatile and capable of addressing targeted challenges.

Looking ahead, emerging languages such as Julia hold the potential to reshape the field by bridging the gap between high-performance computation and user-friendly syntax. Incorporating awareness of such developments into training and research agendas can keep healthcare data science aligned with the cutting edge of technological progress. Developing curricula that align with the practical applications and trends

[DOI: 10.18502/jsbch.v9i2.20265]

identified in this study can better prepare future professionals. Tailoring educational programs to reflect the prevalent use of Python, R, and MATLAB in healthcare analytics will equip students with the skills needed to engage with current technologies and methodologies. This alignment between education and industry needs can drive innovation, improve clinical outcomes, and advance healthcare research (Ganeshan & Rajendran, 2022; Rasoulian Kasrineh et al., 2023).

By incorporating programming language education into health data science training, institutions can empower researchers and practitioners to contribute effectively to the field. Such efforts will not only foster individual career growth but also enhance the collective impact of healthcare analytics in addressing global health challenges (Asamoah et al., 2017; Brunner & Kim, 2016).

Conclusion

This review explored the applications of programming languages in healthcare data analytics. The findings demonstrate that selecting a programming language is not merely a technical decision, but a strategic one with far-reaching implications for healthcare education, research, and practice. The results highlight that the choice of programming language is closely tied to the specific needs and challenges of healthcare applications, ranging from public health and genomics to diagnostics and pharmaceutical studies.

Python, with its flexibility, extensive libraries, and open-source nature, emerged as the most widely used language across various domains, particularly in ML applications. R, favored for its statistical capabilities, continues to dominate data mining and genomics. MATLAB, while facing a decline in general usage, remains a key tool in diagnostic studies due to its advanced capabilities in signal and image processing. These patterns reflect the unique strengths and limitations of each language, as well as their alignment with domain-specific requirements.

Moving forward, healthcare data analyst should consider not only technical capabilities but also factors such as scalability, ease of learning, cost, and the availability of resources and community support when selecting a programming language. For instance, Python's rapid adoption in healthcare analytics highlights the importance of accessibility and versatility in language selection, particularly in dynamic and evolving environments.

The findings of this review have significant implications for education, research, and practice in healthcare. Educational programs in data science and health informatics should align their curricula with the prevalent programming languages and methodologies used in the field. By doing so, institutions can equip students and researchers with the necessary skills to engage effectively with cutting-edge technologies and data-driven approaches in healthcare.

Future research should focus on assessing the long-term impact of programming language selection on the scalability, maintainability, and performance of ML systems in healthcare and compare them Furthermore, exploring the integration of emerging languages and technologies into healthcare data analytics may help address existing limitations and enhance the efficiency and precision of healthcare systems.

In conclusion, the strategic selection of programming languages plays a pivotal role in advancing healthcare data analytics. By aligning technical capabilities with the practical and collaborative needs of healthcare, stakeholders can better harness the power of data science to address complex challenges and improve healthcare delivery on a global scale.

Acknowledgement

Not Applicable.

Conflicts of Interest

The authors declared no competing interests.

Funding

This study was approved by the Research Council of Mashhad University of Medical Sciences (code: 4012097).

Ethical considerations

This study is a systematic review and analysis of

[DOI: 10.18502/jsbch.v9i2.20265]

previously published articles and does not involve direct interaction with human participants or the use of individual-level data.

Code of Ethics

IR.MUMS.MEDICAL.REC.1402.091

Authors' Contributions

S.M.T., S.N.S., A.A., M.A.B. contributed to the conception and design of the study. The search procedure (paper screening and article extraction) was performed by A.A., S.N.S.. M.A.B., A.A., S.N.S. drafted the manuscript, and S.M.T. revised the manuscript. All authors have read and approved the manuscript.

Open Access Policy

Users are allowed to read, download, copy, distribute, print, search, crawl for indexing or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the author.

References

Ahsan, M. M., Luna, S. A., & Siddique, Z. (2022, March). Machine-learning-based disease diagnosis: A comprehensive review. In *Healthcare*, (10(3), 541.

Aichner, T., Grünfelder, M., Maurer, O., & Jegeni, D. (2021). Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019. *Cyberpsychology, behavior, and social networking*, *24*(4), 215-222.

Akan, A., & Chaparro, L. F. (2024). *Signals and systems using MATLAB®* (4th edition ed.). Elsevier.

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., & Badreldin, H. A. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*, *23*(1), 689.

Asamoah, D. A., Sharda, R., Hassan Zadeh, A., & Kalgotra, P. (2017). Preparing a data scientist: A pedagogic experience in designing a big data analytics course. *Decision Sciences Journal of Innovative Education*, *15*(2), 161-190.

Baytaş, İ. M., Peng, Y., & Özgür, A. (2023). Pattern recognition for healthcare analytics. In (Vol. 5, pp. 1186713): Frontiers Media SA.

Brunner, R. J., & Kim, E. J. (2016). Teaching data science. *Procedia Computer Science*, *80*, 1947-1956.

Chambers, J. M. (2020). S, R, and data science. *Proceedings of the ACM on Programming Languages*, *4*(HOPL), 1-17.

Ekmekci, B., McAnany, C. E., & Mura, C. (2016). An introduction to programming for bioscientists: a Python-based primer. *PLoS computational biology*, *12*(6), e1004867.

Fourment, M., & Gillings, M. R. (2008). A comparison of common programming languages used in bioinformatics. *BMC bioinformatics*, *9*(1), 82.

Ganeshan, P., & Rajendran, K. (2022). Dynamic simulation and optimization of anaerobic digestion processes using MATLAB. *Bioresource Technology*, *351*, 126970.

Innes, M., Karpinski, S., Shah, V., Barber, D., Saito Stenetorp, P., Besard, T., Bradbury, J., Churavy, V., Danisch, S., & Edelman, A. (2018). *On machine learning and programming languages*. In Proceedings of SysML conference(SysML 2018). ACM, New York, NY, USA, 3 pages.

Islam, R., Sultana, A., & Islam, M. R. (2024). A comprehensive review for chronic disease prediction using machine learning algorithms. *Journal of Electrical Systems and Information Technology*, *11*(1), 27.

Kalra, N., Verma, P., & Verma, S. (2024). Advancements in AI based healthcare techniques with FOCUS ON diagnostic techniques. *Computers in Biology and Medicine*, *179*, 108917.

Kimiafar, K., Sarbaz, M., Tabatabaei, S. M., Ghaddaripouri, K., Mousavi, A. S., Mehneh, M. R., & Baigi, S. F. M. (2023). Artificial intelligence literacy among healthcare professionals and students: a systematic review. *Frontiers in Health Informatics*, *12*(0), 168. [persian]

[DOI: 10.18502/jsbch.v9i2.20265]

Rasoulian Kasrineh, M., Sharifzadeh, N., & Tabatabaei, S. M. (2023). Smart hospitals worldwide: a systematic review. *Physiology and Pharmacology*, *27*(3), 234-243. [persian]

Sepulveda, J. L. (2020). Using R and bioconductor in clinical genomics and transcriptomics. *The Journal of Molecular Diagnostics*, *22*(1), 3-20.

Shajii, A., Numanagić, I., Leighton, A. T., Greenyer, H., Amarasinghe, S., & Berger, B. (2021). A Python-based programming language for high-performance computational genomics. *Nature biotechnology*, *39*(9), 1062-1064.

Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. Z., Paul, R., Smriti, K., Naik, N., & Somani, B. K. (2022). The role of data science in healthcare advancements: applications, benefits, and future prospects. *Irish Journal of Medical Science (1971-)*, *191*(4), 1473-1483.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), 261-272.

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., & Deac, A. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47-60.

Yu, K.-H., Healey, E., Leong, T.-Y., Kohane, I. S., & Manrai, A. K. (2024). Medical artificial intelligence and human values. *New England Journal of Medicine*, *390*(20), 1895-1904.

[DOI: 10.18502/jsbch.v9i2.20265]