Issues in Informing Science + Information Technology

An Official Publication of the Informing Science Institute InformingScience.org

IISIT.org

Volume 22, 2025

Systematic Literature Review of the Development of Data Science Applications in Healthcare

Anshuman Rangaraj	University of Cincinnati, Cincinnati, OH, United States	rangarar@mail.uc.edu
Theodore Langdon*	University of Cincinnati, Cincinnati, OH, United States	theodore.langdon@uc.edu
Hazem Said	University of Cincinnati, Cincinnati, OH, United States	saidhm@uc.edu

* Corresponding author

Abstract

Aim/Purpose	This study aims to identify the influential factors in the application of Applied Data Science in healthcare.
Background	The research examines the historical adoption and use of Applied Data Science in the healthcare industry, addressing the specific question: What are the factors that influence the use of Applied Data Science in healthcare throughout the history of information technology?
Methodology	The study was conducted using a systematic literature review methodology following Kitchenham and Charters' (2007) protocol, which includes planning, conducting, and reporting the review.
Contribution	The study contributes to understanding key factors that shape the adoption and application of data science technologies in healthcare.
Findings	The study used thematic analysis and identified six factors that influenced the application of Applied Data Science in healthcare: deep learning, data organization, medical or healthcare applications, techniques used, data and computing infrastructure, and language systems.
Recommendation for Practitioners	The study recommends that the application of data science should adapt in response to advancements in technologies and computing infrastructure. It

Accepting Editor: Eli Cohen │ Received: January 20, 2025 │ Revised: March 16, 2025 │ Accepted: April 24, 2025

Cite as: Rangaraj, A., Langdon, T, & Said, H. (2025). Systematic literature review of the development of data science applications in healthcare. *Issues in Informing Science and Information Technology, 22,* Article 1. https://doi.org/10.28945/5503

	further suggests that recent developments in these areas may drive a significant transformation in the field.
Recommendation for Researchers	The study recommends further examination of the evolution of applied data science in healthcare, especially from 1975 to 1990, when limited primary studies were available.
Impact on Society	The results highlight the importance of technological infrastructure, advanced learning models, organized datasets, and ethical considerations for effective data science integration in healthcare.
Future Research	Future studies should explore the timeline more, particularly in the time frame of 1975 to 1990, when a gap in the literature was found. Studies could also investigate regional differences, especially focusing on how differences in incomes and GDP affect healthcare data. Additionally, further research should expand beyond the ACM Digital Library to create a more comprehensive view of the field.

Keywords	applied computing, life and medical sciences, health informatics

INTRODUCTION

The use of the term "data science" dates back to 1962, as stated by Tukey (1992) and Peter Naur (1992). Naur (1992, p. 21) defined it as the "science of dealing with data, once they have been established, while the relation of the data to what they represent is delegated to other fields and sciences." Over the years, the definition of data science has evolved. Kotu and Deshpande (2018, p. 1) define data science as the "collection of techniques used to extract value from data … finding useful patterns, connections, and relationships within data." Data Science is increasingly recognized as a critical component of healthcare, enabling insights from large-scale data sources to inform decision-making and improve patient outcomes, as conveyed by Copenhaver et al. (2019).

Applied data science is a subset of data science that falls under the umbrella of information technology (IT). Spruit and Lytras (2018) mention applied data science, stating that its primary aim is to improve the world around us by reaching maximum social impact. This aligns with the framework of the IT discipline, as mentioned in "A Framework for the Discipline of Information Technology" by Said et al. (2021). As compared to general data science, applied data science research is primarily focused on applying machine learning techniques for solving challenging problems in the day-to-day practices of domain experts by developing advanced software systems that can provide an effective and efficient end-user interface to incorporate machine learning or statistical techniques into. In applied data science, it is not necessarily a primary objective to create new statistical and machine learning techniques for doing data science more effectively (Spruit & Lytras, 2018).

Rapid digitization and initiatives such as the Health Information Technology for Economic and Clinical Health Act have driven unprecedented data growth from electronic medical records, imaging, and other sources. A greater wealth of data not only serves to improve patient care but furthers medical research. Copenhaver et al. (2019) noted that the potential of data science in health is developing and optimizing clinical care and therapeutics to improve operations management and resource allocation. Such data can be mined, processed, and analyzed to uncover specific human traits, biological dynamics, and social interactions. However, it was also a driver behind the health sector that adopted data science and analytics to match the increasing costs, quality, and safety challenges.

Many barriers, such as poor quality, scarce resources, and high patient confidentiality needs, impede the translation of research in data science into tangible patient care. To this end, as expressed in Cohen and Kovacheva (2023), several researchers have discussed various approaches that could make the architecture of such a collaborative healthcare platform both scalable and resource-efficient. Data science applications in medicine and healthcare empower the industry to enhance patient data access, analysis, and treatment outcomes. Cohen and Kovacheva (2023) introduced the MERLIN platform, which is scalable and collaborative to accelerate the development of AI models in health care. The platform automated data acquisition from Electronic Health Records, which can speed up the high-volume processing of multimodal datasets. This system enables the researcher to conduct real-time cohort generation and dataset analysis, accelerating the development of clinical AI models and improving healthcare delivery. Data science is thereby proving crucial for reducing resource-intensive bottlenecks in AI research.

The extraction of insights that may be transformed into actionable information in large sets of healthcare data is emphasized as one of the significant tools in enhancing decision-making, patient care, and operational efficiency. AI-driven data analytics seeks patterns to enhance diagnosis, treatment planning, and resource allocation to drive improved patient outcomes. AI and data science are now related, so considerable contributions to medicine are happening in every aspect, including diagnosis, patient monitoring, and predictive analytics. Kavitha et al. (2023) give a bibliometric analysis regarding large language models, such as ChatGPT, and their growth in importance in healthcare. Applications of LLMs include clinical decision-making, interaction with patients, and analysis of medical literature. Such models have proven especially effective in enhancing healthcare information retrieval, diagnosis assistance, and patient communication. However, bias and ethical considerations remain some of the challenges with using this technology. Data science will continue to leverage the important roles of LLM capabilities toward better healthcare systems by processing extensive data and creating insights from it in real-time.

Wieben et al. (2023) conducted a review of data science applications in nursing practice, focusing on trends in 2021. The review emphasizes how data science-driven tools are used to address nursing-sensitive indicators and improve healthcare outcomes. Predictive modeling and machine learning algorithms enable healthcare professionals to anticipate patient needs, optimize care protocols, and improve operational efficiency. The integration of data science in nursing has shown significant potential to improve patient care and streamline clinical decision-making, thereby enhancing patient outcomes.

Kushwah et al. (2024) analyzed the progress of data science in healthcare, education, and urban planning, underscoring its ability to extract meaningful insights from complex data. In healthcare, this translates to using big data analytics and machine learning to revolutionize medical diagnostics, treatment planning, and patient management. Consequently, healthcare providers can make more informed, personalized treatment decisions, which directly improve patient outcomes and streamline the overall delivery of care.

Katsoulakis et al. (2024) provided an overview of the applications of digital twins (DTs) in healthcare, which represents a cutting-edge use of data science. In this work, they explain that digital twins simulate patient conditions and medical procedures, offering predictive tools that enhance treatment outcomes. Data science plays a pivotal role in their work, refining DT models through the integration of big data and AI, which helps optimize treatment protocols and supports real-time decision-making in healthcare settings. Inam et al. (2024) explore the potential of health data science to address the rising burden of cardiovascular diseases (CVD) in Africa. Data science facilitates the collection, analysis, and interpretation of large datasets to identify health trends and improve healthcare outcomes. By leveraging AI and machine learning, health data science helps policymakers and healthcare providers in Africa make informed decisions about CVD prevention and treatment. The application of data science in this context is crucial for addressing the unique healthcare challenges faced by low-income and middle-income countries (Inam et al., 2024).

Causio et al. (2024) discussed the potential of AI and data science to revolutionize healthcare in Italy, focusing on insights from the Italian Society for Artificial Intelligence in Medicine. AI applications such as remote patient monitoring, machine learning algorithms for diagnostics, and large language models for medical communication are explored as key innovations. Data science is fundamental to the success of these technologies, helping to streamline healthcare processes, improve patient care, and support evidence-based decision-making in the Italian healthcare system.

The paper by Subrahmanya et al. (2021) explored the transformative impact of data science on the healthcare sector. The paper highlights how big data analytics, machine learning, and data mining have revolutionized healthcare by improving decision-making, patient care, and operational efficiency. The authors discuss how Electronic Health Records and data collected from IoT devices can be analyzed to predict disease patterns, optimize treatment plans, and improve patient outcomes. Integrating data science into healthcare enables early detection and personalized treatments, enhancing delivery and reducing costs. Xu et al. (2022) discuss the application of data science in healthcare, with a focus on transparency, explainability, and human-centered artificial intelligence (AI). The emphasis on explainability and ethical AI ensures that healthcare providers can trust and effectively use these technologies to enhance patient care and outcomes. Xu et al. (2022) also highlight interdisciplinary collaboration between academia, industry, and regulatory bodies to foster innovation in healthcare through data science.

The research on applied data science in healthcare so far underlines that making proper use of the huge volume of healthcare data, which is generated through electronic health records, IoT devices, and wearable technologies, is very important, as mentioned by Abedjan et al. (2019). Although experts recognize applied data science as a potential field in healthcare services, there is a possible gap in understanding why its application is necessary for addressing certain critical challenges in healthcare services. The reviewed literature identified the research gaps mentioned above, and based on those, the following research questions have been identified.

RQ1: How has the adoption of applied data science in healthcare changed over time?

RQ2: What key factors have driven the adoption of applied data science in healthcare?

METHODOLOGY

The study uses a systematic literature review methodology following Kitchenham and Charters' (2007) protocol, which is structured into three phases: planning, conducting, and reporting the review. This method was chosen due to its clear stages and use of clearly defined inclusion and exclusion criteria.

SEARCH PROCESS

The study used the ACM Digital Library as the data source due to its focus on technology and the fact that it contains 760,190 full-text articles and 24,445,802 citations in the database as of October 12, 2024, as found on the Association for Computing Machinery (n.d.) digital library's *About ACM DL* page. Papers from the database were searched using the following query:

```
"query": {AllField:(applications of data science in healthcare) AND All-
Field:(history of applied data science) AND Title:(medicine) AND All-
Field:(healthcare)} "filter": {ACM Content: DL}
```

The search was conducted on Saturday, October 12, 2024, by accessing the ACM digital library via the University of Cincinnati Libraries. The initial search returned 507 articles. These articles were saved within the ACM digital library account, and a bibliography was exported. The bibliography was uploaded to Rayyan for sorting and review. Rayyan is a software for organizing systematic literature reviews that was developed by the Qatar Computing Research Institute (QCRI). It is available in free and paid modes, depending on the researcher's needs. This review utilized the paid student tier to organize the articles.

Once the bibliography was loaded into Rayyan, the duplicate article's function was used to identify articles it found that appeared to be duplicated. The system looks at the title and abstract to make this determination and brings any articles showing 90% or greater similarity to the researcher. Rayyan found 13 articles that met these criteria. Upon review, seven of these articles were found to be duplicates, and one copy was deleted. The remaining copy was retained in the system.

The articles were then screened for applicability to the research question. The researchers developed a list of inclusion and exclusion criteria (Table 1). These criteria were chosen to focus on the factors involved in the adoption and history of using applied data science in healthcare. The screening process was completed using Rayyan's "screening" function, which allows researchers to include or exclude articles or mark them as "maybe" for further review. One researcher read the article titles and abstracts in the software and assigned a status of yes, no, or maybe. This researcher could also add comments explaining why the article fell into its respective category. After the first reading of the titles and abstracts, the second researcher performed the same steps using Rayyan. Rayyan software allows each researcher to screen articles in a "blind" mode, where each review is completed without revealing the other screening results. After screening, the two researchers matched 95% of the articles. The 25 articles where there was disagreement were then reviewed and all were found to have been rated as maybe or to be excluded by each of the reviewers. The decision was made to remove these articles because neither reviewer had marked them for inclusion. Using the criteria shown in Table 1, 213 articles met the criteria, and 19 did not. There were 59 articles placed into the "maybe" category, which were then removed after reviewing the abstracts a second time for applicability to the research question.

Table 1. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Articles available in full text.	Articles focused on data science as a theory and not as applied to healthcare.
Articles are written in English.	Articles only incorporate applied data science with no historical information.
Articles discuss the history of applied data science in healthcare.	Documents that contain only opinions.
Articles are taken from peer-reviewed publications.	
Article is a full study, a meta-analysis, or a structured literature review.	

Quality Assessment

The researchers then scored the remaining 213 articles as Slightly Focused, Semi-Focused, and Fully Focused. The researchers read the title and abstract of each article again to ascertain how well the article answered the question shown in Table 2. Completing this review allowed the researchers to measure the relevance and applicability of the topic of exploring the history of applied data science within the healthcare industry. Once the scores of both researchers were compared, a final score was tabulated. The majority (61%) of the articles were found to be fully focused on the topic, while 21.6% were semi-focused, and 17.4% were only slightly focused. No articles were found to be unfocused due to the prior screening of the search results.

Table 2. Quality assessment of the included articles

Question	Slightly focused	Semi-focused	Fully focused
To what degree does the article discuss the use of applied data science in healthcare?	37 (17.4%)	46 (21.6%)	130 (61%)

Thematic Analysis

In preparation for thematic analysis, the 213 remaining articles were downloaded as full-text PDFs from the ACM Digital Library. The ACM uses policies that forbid automated or scripted downloading, so the articles were downloaded individually by using the article's DOI included in the bibliography downloaded previously. A list of these articles are included in Appendix A1. These full-text PDFs were then loaded into a Python environment within Google Colab (https://colab.research.google.com/). Google Colab offers a cloud-based solution, providing a Jupyter notebook environment with access to graphics processing units and tensor processing units that exceed the speeds of either researcher's computers, allowing for faster processing of the documents.

Once the files were loaded, a script was written to search the PDF text for words having three or more letters per word. This script is shown in Appendix A2. These criteria were chosen to capture as many words as possible while eliminating two-letter words that often do not contain useful information. This list contained 82,491 words. After this list was created, the text preprocessing techniques explained by Dudhabaware and Madankar (2014) were applied to the list of repeated phrases. Some steps from their process were omitted, and the steps did not apply to the research of this article. Sentiment Analysis was the first step omitted because there was no need to measure the sentiment to construct a timeline.

Named Entity Recognition (NER) was then applied to the list of repeated phrases in accordance with Dudhabaware and Madankar's (2014) protocol. This was completed using the spaCy.io model via the script shown in Appendix A3. The NER model within spaCy looks for numbers without text, named organizations or people, and numeric dates. This created a column in the data to be used after other steps took place.

The next technique applied to the spreadsheet was lemmatization, shown in Appendix A4. Lemmatization is a method to reduce words to their root form to identify similar terms. This process occurs using a dictionary and morphological analysis to determine different inflections or tenses of words (Manning et al., 2009). This technique was used instead of stemming, as Dudhabaware and Madankar (2014) called for because it better identifies the root word. Stemming is more likely to delete letters at the end of the word, with no context to the root word, and can misidentify tenses. After lemmatizing, the terms that were duplicated in their new form were combined into one row in the spreadsheet using the script in Appendix A5. After downloading this combined document, the first researcher looked through any words that only appeared once and deleted any unrelated to applied data science. The words that were tagged with Named Entity Recognition were also checked, and all were deleted due to none being related terms.

The researchers used the list of lemmatized words to find words and phrases that were repeated to find similar terms. This list contained 147 terms which can be found in Appendix A13. The researchers defined these terms using context or definitions found within the articles. After compiling this list, terms that were about the same subject were combined along with their count, leaving 48 terms. These terms were then grouped by the researchers' themes. These themes are shown in Table 3.

Table 3. Themes in the articles found

Theme	Percentage of articles that included the theme
Deep learning	31.91%
Data organization	27.44%
Medical or healthcare applications	12.31%
Techniques used	11.76%
Data and computing infrastructure	11.21%
Language systems	5.38%

RESULTS

The terms were then grouped based on their relationship, resulting in the six themes. The list in Appendix A6 shows which terms were shown in which primary sources. Within each theme, the word frequency was measured by theme and how prevalent the terms were in the 213 articles. These terms and the percentage of articles that included these terms are shown in Appendices A7-A12.

After the themes were extracted, a Python script was created in Google Colab to find which primary studies contained which words. It was chosen due to the features it offers to integrate with Google Drive cloud storage, where all the full-text articles were stored. This created a CSV that was then analyzed to see which year each word was first seen in the documents and when it was last seen.

The data in Figure 1 describes the themes generated by papers from the 1970s to 2025. The colors represent each theme, and the size of the bubble represents the frequency of the terms mentioned under those themes. The interactive chart was created with Flourish (2016). This is an interactive chart, and an active link is included in Appendix A14.

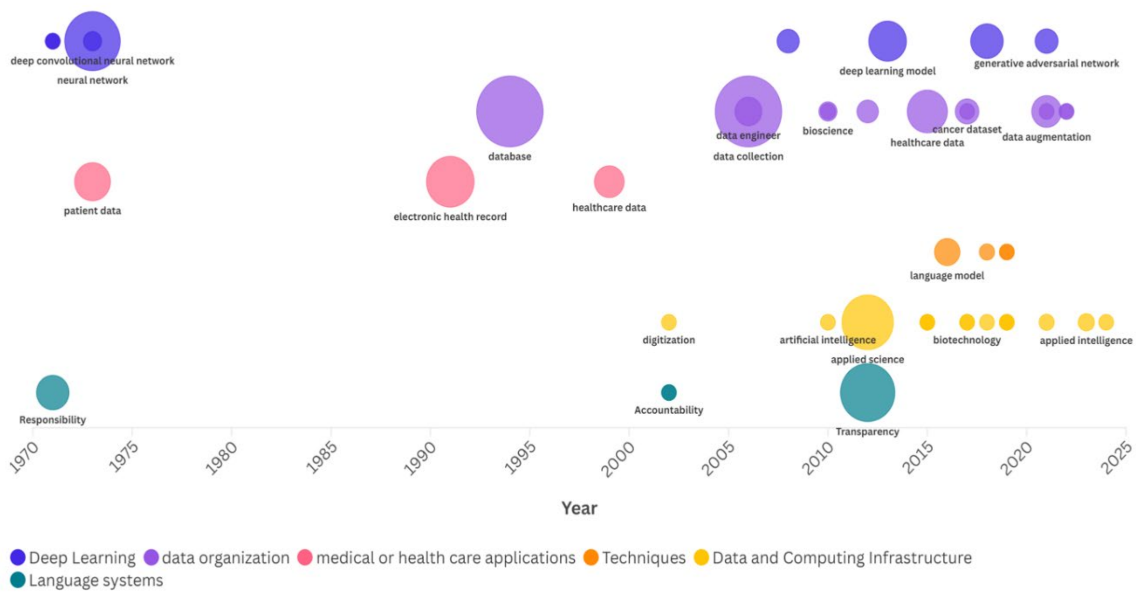

Figure 1. Timeline of terms and the first mention in the literature found

DISCUSSION

This study aimed to identify factors in applying Applied Data Science to healthcare, as shown in the research questions. The results found six factors that impacted the development of data science and its applications in healthcare. The six major factors are data and computing infrastructure, data organization, deep learning, language systems, techniques used, and medical or healthcare applications cut across the results and indicate key enablers for Applied Data Science in healthcare. By categorizing these themes, the study makes it clear that for Applied Data Science applications to meet the operational demands of healthcare, the infrastructure, advanced models of learning, and organized datasets are paramount.

The themes that emerged from the 213 primary articles underscore the diverse factors influencing data science applications in healthcare. Deep Learning was the most prominent theme, appearing in 31.91% of the articles, underpinning how core data science is in developing better operational efficiency and patient outcomes. Terms like 'data organization,' 'medical or clinical applications,' and 'techniques used' frequently appeared, demonstrating the increasing use of data-driven techniques to improve diagnostics and resource allocation in clinical settings. This agrees with previous literature stating that data science is helpful in optimizing care pathways and predictive analytics, as shown by Copenhaver et al. (2019). Each theme's prevalence across the studied articles demonstrates the breadth and complexity of Applied Data Science applications in healthcare.

The emerging theme of Deep Learning was found in 31.91% of the articles and underlined a great interest in powerful learning models such as neural networks or deep convolutional networks, opening new horizons toward healthcare predictive analytics. The terms 'DeepMind' and 'neural model' have appeared frequently since the 1970s, indicating a long-standing interest in deep learning for diagnostics and personalized care. This is probably due to the increasing demand for handling large volumes of data and extracting useful conclusions from them, which suggests that deep learning-based approaches have played a foundational role in promoting Applied Data Science applications in the health sector.

Data Organization, as covered by 27.44% of the total papers, came second. High-frequency terms like 'data collection,' 'database,' and 'healthcare data' highlight the importance of structuring and curating data for efficient access and availability. Since most healthcare data is housed in piecemeal systems, it is of prime importance that such data be organized so that Applied Data Science can apply itself most effectively. The use of data augmentation and labeling for specific applications – for example, cancer datasets – very well depicts how organized data serves to enhance the ability of Applied Data Science to improve outcomes by way of model training and analytics.

The theme of Medical or Healthcare Applications, which accounted for 12.31% of the articles, emphasizes the role of Applied Data Science in supporting clinical workflows. Terms such as 'electronic health record' and 'patient data' illustrate how applications of data science impinge directly on patients. This theme supports the research question even more by showing exactly how Applied Data Science is threaded through various healthcare operations in support of broader adoption.

The Techniques Used theme has been identified in 11.76% of the articles, referring to various methods enabling Applied Data Science. These techniques will convert raw data into actionable insights through methods such as 'neural machine translation' and 'virtual digital assistant,' thus reinforcing the application of Applied Data Science in real-time within a clinical setting. The focus on different techniques underlines the flexibility of Applied Data Science methodologies in meeting a variety of needs in health care, further confirming its flexibility and value for use in industry.

The Data and Computing Infrastructure theme, which appeared in 11.21% of the articles, highlights the importance of technological advancements such as 'applied intelligence,' 'federated machine learning,' and 'biotechnology' in driving the adoption of Applied Data Science. The reason is that healthcare environments need strong systems for data storage, security, and processing. It is such an infrastructure that aids in real-time data analysis, which plays a vital role in the decision-making and operational aspects pertaining to healthcare environments. Therefore, it acts as a critical determinant that affects the level and efficiency of Applied Data Science utilization in the healthcare domain.

Finally, Language Systems are featured in 5.38% of the articles. This is perhaps symptomatic of an interest in using the tools of Applied Data Science in a transparent and nondiscriminatory way in health. The terms "accountability" and "transparency" point to the ethical dimensions of the use of Applied Data Science. Ethical applications of Applied Data Science approaches are expected to be obvious and thus adopted with more integration into patient care. However, ethical issues, as identified in the Language Systems theme, highlight the importance of transparency and accountability with a view to addressing concerns on bias and patient confidentiality.

The healthcare tale of Applied Data Science has gradually progressed over five decades. The tale started in the 1970s with rudimentary neural networks and basic patient data gathering while the discipline made huge leaps in progress. In the 1990s, when electronic health records and healthcare databases began to be established, digital foundations began to be built. The 2000s were more concerned with engineering and methodologies involved in data gathering, while responsibility, accountability, and later transparency found their way into ethics. The 2010s gave the world a revolutionary century with deep learning models, bioscience use cases, and artificial intelligence being used for healthcare issues. By the 2020s, we will be standing in a dense ecosystem of technologies that are interwoven, such as generative adversarial networks, data augmentation, and applied intelligence. Each of these is designed to coax greater value out of healthcare data. Along the way, we have always kept sophisticated ethics and synthesizing different technologies in mind to promote improved healthcare outcomes.

A look at the history of applied data science and healthcare yields several very deep insights. What started as primitive neural networks and simple patient data were captured in the 1970s and laid the groundwork for a revolution that spanned five decades. The graph readily illustrates the construction of data infrastructure. Databases and electronic health records came before and facilitated higher-order analytic capabilities. At the same time, ethics matured from simple accountability to accountability and transparency, developing in tandem with technical advancements. Post 2010, the speed was quite rapid, with deep learning, artificial intelligence, and specialty fields such as bioscience and data engineering quickly converging into healthcare applications. Most insightful, perhaps, is how space has moved from discrete technological innovations to a densely integrated system wherein biotechnology, language models, and advanced analytics come together – this pictorial history. The graphic in Figure 1 shows technological innovation and an underlying movement towards multidisciplinary solutions that utilize various computational techniques to solve increasingly challenging healthcare issues.

As Ellenberg (1994) mentioned, selection bias may have impacted both internal and external validity. External validity was impacted by the fact that the articles were selected only from the ACM Digital Library; this excludes studies from other databases. This could explain the lack of articles from 1975 to 1990. Internal validity is achieved because of the usage of explicit inclusion and exclusion criteria. Furthermore, geographical and contextual issues of interest are restricted to the selected articles, many of which relate to high-resource settings, raising concerns about generalizations to low-income and middle-income countries with underdeveloped data science infrastructures. Reliability is enhanced by consistent reliance on systematic tools such as Rayyan for screening. However, any thematic analysis will differ across different studies and thus may influence the reproducibility of the results. Due to time constraints, the primary studies were processed with software, with only a full reading of the titles and abstracts. This could lead to valid articles being skipped if the abstracts did not include the subjects the researchers were looking for within the inclusion and exclusion criteria.

Conclusion and Future Work

The study revealed that Applied Data Science in healthcare emerged in the 1970s and remains an active area of research today. There was a greater focus on the application and methods used to conduct the work in healthcare, but the research also showed that challenges and concerns would need addressing. This discussion consolidates the major findings and provides a response to the research question. It highlights the importance of infrastructure, methodological adaptability, and ethical integrity for effective Applied Data Science in healthcare. Although limitations exist, this study gives much-needed insight into both historical and current drivers of Applied Data Science and lays a foundation for future studies to develop these drivers in a range of healthcare settings.

Further research could delve deeper into the timeline of each topic area, providing a more nuanced understanding. Additionally, the overall timeline could be studied in sections because five decades of articles were found. The earlier eras were more related to theory, whereas the later eras were more related to applications. Splitting further study by this could create a deeper understanding of the timeline. The study of the ethics of Applied Data Science in healthcare was present from the start of the concept but has become more prevalent in the last two decades. This area needs to be looked at more closely as the techniques developed advance quickly. The gap in research from 1975 to 1990 should be further studied to determine if there was any development in that time frame.

While the research questions have largely been answered, there are areas where added depth could be used to strengthen the findings. Moving beyond the ACM Digital Library, a larger scope in the database can cover seminal studies over other databases related to health and, hence, enhance that vision of data science applied across a variety of contexts. Additionally, seeking out other literature reviews that focus on the six factors could add additional insight to the topic.

Another area of future work would be the adoption of papers with regional variation, which could include research papers from different countries. This could introduce limitations like language barriers, but papers that have been published could bring insights regarding applied data science throughout history and diverse backgrounds in the healthcare sector. More discussion on the challenges of low-income and middle-income countries in adopting applied data science might provide a fuller view of adoption trends at the global level. Additionally, finding case studies of specific healthcare applications would add in-depth examples, showing how data science impacted healthcare outcomes or systems over time would strongly reinforce the historical perspective.

REFERENCES

Abedjan, Z., Boujemaa, N., Campbell, S., Casla, P., Chatterjea, S., Consoli, S., Costa-Soria, C., Czech, P., Despenic, M., Garattini, C., Hamelinck, D., Heinrich, A., Kraaij, W., Kustra, J., Lojo, A., Sanchez, M. M., Mayer, M. A., Melideo, M., Menasalvas, E., … Wuyts, R. (2019). Data Science in Healthcare: Benefits, Challenges and Opportunities. In S. Consoli, D. Reforgiato Recupero, & M. Petković (Eds.), *Data Science for Healthcare: Methodologies and Applications* (pp. 3–38). Springer International Publishing. https://doi.org/10.1007/978-3-030-05249-2_1

Association for Computing Machinery. (n.d.). *About ACM digital library.* https://dl.acm.org/about

Causio, F. A., De Angelis, L., & Diedenhofen, G., Talio, A., & Baglivo, F. (2024). Perspectives on AI use in medicine: Views of the Italian Society of Artificial Intelligence in Medicine. *Journal of Preventive Medicine and Hygiene*, *65*(2), E285-E289. https://doi.org/10.15167/2421-4248/jpmh2024.65.2.3261

Cohen, R. Y., & Kovacheva, V. P. (2023). A methodology for a scalable, collaborative, and resource-efficient platform, MERLIN, to facilitate healthcare AI research. *IEEE Journal of Biomedical and Health Informatics*, *27*(6), 3014-3025. https://doi.org/10.1109/JBHI.2023.3259395

Copenhaver, M. S., Hu, M., Levi, R., Safavi, K., & Langle, A. C. Z. (2019). Health system innovation: Analytics in action. *INFORMS Tutorials in Operations Research*, 238-266. https://doi.org/10.1287/educ.2019.0202

Dudhabaware, R. S., & Madankar, M. S. (2014, December). Review on natural language processing tasks for text documents. IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India, 1-5. https://doi.org/10.1109/ICCIC.2014.7238427

Ellenberg, J. H. (1994). Selection bias in observational and experimental studies. *Statistics in Medicine*, *13*(5-7), 557-567. https://doi.org/10.1002/sim.4780130518

Flourish. (2016). *About us.* https://flourish.studio/company/

Inam, M., Sheikh, S., Khoja, A., Abubakar, A., Shah, R., Samad, Z., Ngugi, A., Alarakhiya, F., Waljee, A., & Virani, S. S. (2024). Health data sciences and cardiovascular disease in Africa: Needs and the way forward. *Current Atherosclerosis Reports*, *26*, 659-671. https://doi.org/10.1007/s11883-024-01235-1

Katsoulakis, E., Wang, Q., & Wu, H., Shahriyari, L., Fletcher, R., Liu, J., Achenie, L., Liu, H., Jackson, P., Xiao, Y., Syeda-Mahmood, T., Tuli, R., & Deng, J. (2024). Digital twins for health: A scoping review. *npj Digital Medicine*, *7*, Article 77. https://doi.org/10.1038/s41746-024-01073-0

Kavitha, R. K., Jaisingh, W., & Kaarthiekheyan, V. (2023). Application of artificial intelligence and data science across domains: A perspective study. In T. Murugan & N. E. (Eds.), *Handbook of research on data science and* *cybersecurity innovations in Industry 4.0 technologies* (pp. 1-29). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-6684-8145-5.ch001

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering. Version 2.3*. Keele University and University of Durham. https://www.researchgate.net/publication/302924724_Guidelines_for_performing_Systematic_Literature_Reviews_in_Software_Engineering

Kotu, V., & Deshpande, B. (2018). Data science process. In V. Kotu, & B. Deshpande (Eds.), *Data science* (2nd ed., 19-37). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-814761-0.00001-0

Kushwah, V. S., Parashar, J., Dabas, P., Meena, L., & Sharma, V. (2024). Data science advancements in healthcare, education, and cities: An overview. In T. Murugan, J. W., & V. P. (Eds.), *Technologies for sustainable healthcare development* (pp. 17-36). IGI Global. https://doi.org/10.4018/979-8-3693-2901-6.ch002

Manning, C., Raghavan, P., & Schütze, H. (2009). *An introduction to information retrieval.* Cambridge University Press.

Naur, P. (1992). *Computing: A human activity.* ACM Press.

Said, H., Zidar, M., Varlioglu, S., & Itodo, C. (2021). A framework for the discipline of information technology. *Proceedings of the 22nd Annual Conference on Information Technology Education* (pp. 53–54). Association for Computing Machinery. https://doi.org/10.1145/3450329.3478313

Spruit, M., & Lytras, M. (2018). Applied data science in patient-centric healthcare: Adaptive analytic systems for empowering physicians and patients. *Telematics and Informatics*, *35*(4), 643-653. https://doi.org/10.1016/j.tele.2018.04.002.

Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., Naik, N., & Somani, B. K. (2021). The role of data science in healthcare advancements: applications, benefits, and future prospects. *Irish Journal of Medical Science*, *191*(4), 1473–1483. https://doi.org/10.1007/s11845-021-02730-z

Tukey, J. W. (1992). The future of data analysis. In S. Kotz, & N. L. Johnson (Eds.), Breakthroughs in statistics: Methodology and distribution (pp. 408-452). Springer. https://doi.org/10.1007/978-1-4612-4380-9_31

Wieben, A. M., Walden, R. L., Alreshidi, B. G., Brown, S. F., Cato, K., Coviak, C. P., Cruz, C., D'Agostino, F., Douthit, B. J., Forbes, T. H., Gao, G., Johnson, S. G., Lee, M. A., Mullen-Fortino, M., Park, J. I., Park, S., Pruinelli, L., Reger, A., Role, J., … Jeffery, A. D. (2023). Data science implementation trends in nursing practice: A review of the 2021 literature. *Applied Clinical Information*, *14*(3), 585-593. https://doi.org/10.1055/a-2088-2893

Xu, T., Wang, F., Chakraborty, P., Hsueh, P. Y. S., Stiglic, G., Bian, J., Yao, L., Gossmann, A., & Buettner, F. (2022). Workshop on applied data science for healthcare: Transparent and human-centered AI. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 4908-4909). Association for Computing Machinery. https://doi.org/10.1145/3534678.3542901

APPENDIX

A1 - PRIMARY STUDIES

[PS1] V. E. Lange, "1999: The benefits of simulation modeling in medical planning and medical design".

[PS2] A. J. Burns, M. E. Johnson, and P. Honeyman, "A brief chronology of medical device security," *Commun. ACM*, vol. 59, no. 10, pp. 66–72, Sep. 2016, https://doi.org/10.1145/2890488.

[PS3] N. Zhu, J. Cao, K. Shen, X. Chen, and S. Zhu, "A Decision Support System with Intelligent Recommendation for Multi-disciplinary Medical Treatment," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 1s, pp. 1–23, Jan. 2020, https://doi.org/10.1145/3352573.

[PS4] R. Yesha, A. Gangopadhyay, and E. Siegel, "A Graph-Based Method for Analyzing Electronic Medical Records," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, Paris France: ACM, Aug. 2015, pp. 1036–1041. https://doi.org/10.1145/2808797.2808806.

[PS5] M. De La Villa, F. Aparicio, M. J. Maña, and M. De Buenaga, “A learning support tool with clinical cases based on concept maps and medical entity recognition,” in *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, Lisbon Portugal: ACM, Feb. 2012, pp. 61–70. https://doi.org/10.1145/2166966.2166978.

[PS6] V. Koufi, F. Malamateniou, and G. Vassilacopoulos, “A medical diagnostic and treatment advice system for the provision of home care,” in *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*, Athens Greece: ACM, Jul. 2008, pp. 1–7. https://doi.org/10.1145/1389586.1389641.

[PS7] J. Guan, R. Li, S. Yu, and X. Zhang, “A Method for Generating Synthetic Electronic Medical Record Text,” *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, vol. 18, no. 1, pp. 173–182, Jan. 2021, https://doi.org/10.1109/TCBB.2019.2948985.

[PS8] C.-F. Lee, C.-W. Hsu, and I.-C. Lin, “A Novel Reversible Data Hiding Method with Contrast Enhancement for Color Medical Images,” in *2023 the 7th International Conference on Medical and Health Informatics (ICMHI)*, Kyoto Japan: ACM, May 2023, pp. 287–292. https://doi.org/10.1145/3608298.3608351.

[PS9] D. Power, E. Politou, M. Slaymaker, S. Harris, and A. Simpson, “A relational approach to the capture of DICOM files for Grid-enabled medical imaging databases,” in *Proceedings of the 2004 ACM symposium on Applied computing*, Nicosia Cyprus: ACM, Mar. 2004, pp. 272–279. https://doi.org/10.1145/967900.967959.

[PS10] J. J. Nalluri *et al.*, “A smart healthcare portal for clinical decision making and precision medicine,” in *Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking*, Varanasi India: ACM, Jan. 2018, pp. 1–6. https://doi.org/10.1145/3170521.3170530.

[PS11] Y. Xu, Q. Fu, Y. Wu, M. Gao, and X. Jiang, “A Study of the Multi-Agent Benefit Distribution Model in the Transformation of Medical Science and Technology Achievements,” in *Proceedings of the third International Conference on Medical and Health Informatics 2019 - ICMHI 2019*, Xiamen, China: ACM Press, 2019, pp. 54–59. https://doi.org/10.1145/3340037.3340038.

[PS12] E. Lamiae, E. Fatiha, B. Mohammed, and G. T. Hicham, “A study on smart home for medical surveillance: contribution to smart healthcare paradigm,” in *Proceedings of the 4th International Conference on Smart City Applications*, Casablanca Morocco: ACM, Oct. 2019, pp. 1–6. https://doi.org/10.1145/3368756.3368994.

[PS13] J. Zhu, Y. Liu, S. Yang, S. Zhai, Y. Zhang, and C. Wen, “A Supervised Learning Framework for Prediction of Incompatible Herb Pair in Traditional Chinese Medicine,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Torino Italy: ACM, Oct. 2018, pp. 1799–1802. https://doi.org/10.1145/3269206.3269241.

[PS14] P. Messina *et al.*, “A Survey on Deep Learning and Explainability for Automatic Report Generation from Medical Images,” *ACM Comput. Surv.*, vol. 54, no. 10s, pp. 1–40, Jan. 2022, https://doi.org/10.1145/3522747.

[PS15] M. Wang, M. Blount, J. Davis, A. Misra, and D. Sow, “A time-and-value centric provenance model and architecture for medical event streams,” in *Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments*, San Juan Puerto Rico: ACM, Jun. 2007, pp. 95–100. https://doi.org/10.1145/1248054.1248082.

[PS16] E. Domazet, M. Gusev, and L. Antovski, “A Time-Critical Mobile Application based on ECG Medical Monitoring,” in *Proceedings of the 8th Balkan Conference in Informatics*, Skopje Macedonia: ACM, Sep. 2017, pp. 1–4. https://doi.org/10.1145/3136273.3136281.

[PS17] S. Ji *et al.*, “A Unified Review of Deep Learning for Automated Medical Coding,” *ACM Comput. Surv.*, vol. 56, no. 12, pp. 1–41, Dec. 2024, https://doi.org/10.1145/3664615.

[PS18] Q. Zhang, C. Bai, L. T. Yang, Z. Chen, P. Li, and H. Yu, “A Unified Smart Chinese Medicine Framework for Healthcare and Medical Services,” *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, vol. 18, no. 3, pp. 882–890, May 2021, https://doi.org/10.1109/TCBB.2019.2914447.

[PS19] J. Cherian, S. Ray, and T. Hammond, "An Activity Recognition System for Taking Medicine Using In-The-Wild Data to Promote Medication Adherence," in *26th International Conference on Intelligent User Interfaces*, College Station TX USA: ACM, Apr. 2021, pp. 575–584. https://doi.org/10.1145/3397481.3450673.

[PS20] S. Jagannath, A. Sarcevic, and I. Marsic, "An Analysis of Speech as a Modality for Activity Recognition during Complex Medical Teamwork," in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, New York NY USA: ACM, May 2018, pp. 88–97. https://doi.org/10.1145/3240925.3240941.

[PS21] J. Gardner, L. Xiong, F. Wang, A. Post, J. Saltz, and T. Grandison, "An evaluation of feature sets and sampling techniques for de-identification of medical records," in *Proceedings of the 1st ACM International Health Informatics Symposium*, Arlington Virginia USA: ACM, Nov. 2010, pp. 183–190. https://doi.org/10.1145/1882992.1883019.

[PS22] M. C. Martinis, C. Zucco, and M. Cannataro, "An Italian lexicon-based sentiment analysis approach for medical applications," in *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, Northbrook Illinois: ACM, Aug. 2022, pp. 1–4. https://doi.org/10.1145/3535508.3545594.

[PS23] A. Shomirov and J. Zhang, "An Overview of Deep Learning in MRI and CT Medical Image Processing," in *2021 3rd International Symposium on Signal Processing Systems (SSPS)*, Beijing China: ACM, Mar. 2021, pp. 72–78. https://doi.org/10.1145/3481113.3481125.

[PS24] S. Zhang, W. Fu, Y. Hu, and H. Chen, "Application of A Question Recognition Strategy for Small Data Sets in Medical Insurance QA Systems," in *Proceedings of the 2024 3rd International Conference on Frontiers of Artificial Intelligence and Machine Learning*, Yichang China: ACM, Apr. 2024, pp. 318–324. https://doi.org/10.1145/3653644.3665180.

[PS25] L. Wang, K. Deng, and Y. Li, "Application progress of artificial intelligence in traditional Chinese medicine nursing," in *Proceedings of the 2nd International Symposium on Artificial Intelligence for Medicine Sciences*, Beijing China: ACM, Oct. 2021, pp. 69–74. https://doi.org/10.1145/3500931.3500945.

[PS26] R. Ellerweg, K. Hofer, A. Khromov, P. Voigt, and S. Fischer, "Applying trajectory mining in medical image data," in *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, Barcelona Spain: ACM, May 2017, pp. 101–109. https://doi.org/10.1145/3154862.3154889.

[PS27] S. Sinha, T. Agarwal, and P. Pandey, "Artificial intelligence in healthcare: medical named entity recognition based audio prescription generator," in *Proceedings of the 5th International Conference on Information Management & Machine Intelligence*, Jaipur India: ACM, Nov. 2023, pp. 1–5. https://doi.org/10.1145/3647444.3647841.

[PS28] J. Chen, Y. Lu, T. Zhang, and Z. Ouyang, "Artificial Intelligence in Medicine in the United States, China and India," in *Proceedings of the 1st International Symposium on Artificial Intelligence in Medical Sciences*, Beijing China: ACM, Sep. 2020, pp. 257–264. https://doi.org/10.1145/3429889.3429938.

[PS29] F. Cabitza, A. Campagner, and V. Cavosi, "Assessing the impact of medical AI: a survey of physicians' perceptions," in *2021 5th International Conference on Medical and Health Informatics*, Kyoto Japan: ACM, May 2021, pp. 225–231. https://doi.org/10.1145/3472813.3473195.

[PS30] S. Klakegg *et al.*, "Assisted Medication Management in Elderly Care Using Miniaturised Near-Infrared Spectroscopy," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 2, pp. 1–24, Jul. 2018, https://doi.org/10.1145/3214272.

[PS31] H. Ding and X. Luo, "Attention-based Unsupervised Keyphrase Extraction and Phrase Graph for COVID-19 Medical Literature Retrieval," *ACM Trans. Comput. Healthcare*, vol. 3, no. 1, pp. 1–16, Jan. 2022, https://doi.org/10.1145/3473939.

[PS32] V. Pereira, S. Matos, and J. L. Oliveira, "Automated ICD-9-CM medical coding of diabetic patient's clinical reports," in *Proceedings of the First International Conference on Data Science, E-learning and Information Systems*, Madrid Spain: ACM, Oct. 2018, pp. 1–6. https://doi.org/10.1145/3279996.3280019.

[PS33] A. Hliaoutakis, K. Zervanou, E. G. M. Petrakis, and E. E. Milios, "Automatic document indexing in large medical collections," in *Proceedings of the international workshop on Healthcare information and knowledge management*, Arlington Virginia USA: ACM, Nov. 2006, pp. 1–8. https://doi.org/10.1145/1183568.1183570.

[PS34] R. Kumari, D. K. Sah, K. Cengiz, N. Ivković, and P. Balaji, "Automatic graph construction and Exploring different types of LSTMs for Asian Hindi languages for Medical review Sentiment Analysis," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, p. 3626196, Oct. 2023, https://doi.org/10.1145/3626196.

[PS35] R. A. Ebardo and N. J. Celis, "Barriers to the Adoption of Electronic Medical Records in Select Philippine Hospitals: A Case Study Approach," in *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence*, Bali Indonesia: ACM, Apr. 2019, pp. 123–128. https://doi.org/10.1145/3330482.3330503.

[PS36] M. Alizadeh, S. Peters, S. Etalle, and N. Zannone, "Behavior analysis in the medical sector: theory and practice," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, Pau France: ACM, Apr. 2018, pp. 1637–1646. https://doi.org/10.1145/3167132.3167307.

[PS37] N. Savage, "Better medicine through machine learning," *Commun. ACM*, vol. 55, no. 1, pp. 17–19, Jan. 2012, https://doi.org/10.1145/2063176.2063182.

[PS38] B. A. Abdelhakim, B. A. Mohamed, and S. Fellaji, "Big data architecture for decision making in protocols and medications assignment," in *Proceedings of the Mediterranean Symposium on Smart City Application*, Tangier Morocco: ACM, Oct. 2017, pp. 1–4. https://doi.org/10.1145/3175628.3175653.

[PS39] O. El Andaloussi, M. A. Kbir, and B. D. R. Hassani, "Biological data annotation using complementary and alternative Medicine with a collection tracking system (Hirbalink)," in *Proceedings of the 3rd International Conference on Smart City Applications*, Tetouan Morocco: ACM, Oct. 2018, pp. 1–5. https://doi.org/10.1145/3286606.3286838.

[PS40] S. Zhang, Z. Chen, D. Liu, and Q. Lv, "Building Structured Patient Follow-up Records from Chinese Medical Records via Deep Learning," in *2022 2nd International Conference on Bioinformatics and Intelligent Computing*, Harbin China: ACM, Jan. 2022, pp. 65–71. https://doi.org/10.1145/3523286.3524517.

[PS41] F. T. Jaigirdar, C. Rudolph, and C. Bain, "Can I Trust the Data I See?: A Physician's Concern on Medical Data in IoT Health Architectures," in *Proceedings of the Australasian Computer Science Week Multiconference*, Sydney NSW Australia: ACM, Jan. 2019, pp. 1–10. https://doi.org/10.1145/3290688.3290731.

[PS42] A. Kaponis, A. A. Kaponis, and M. Maragoudakis, "Case study analysis of medical and pharmaceutical chatbots in digital marketing and proposal to create a reliable chatbot with summary extraction based on users' keywords.," in *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, Corfu Greece: ACM, Jul. 2023, pp. 357–363. https://doi.org/10.1145/3594806.3604765.

[PS43] S. Zhao, M. Jiang, M. Liu, B. Qin, and T. Liu, "CausalTriad: Toward Pseudo Causal Relation Discovery and Hypotheses Generation from Medical Text Data," in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Washington DC USA: ACM, Aug. 2018, pp. 184–193. https://doi.org/10.1145/3233547.3233555.

[PS44] R. A. Jeewantha, M. N. Halgamuge, A. Mohammad, and G. Ekici, "Classification Performance Analysis in Medical Science: Using Kidney Disease Data," in *Proceedings of the 2017 International Conference on Big Data Research - ICBDR 2017*, Osaka, Japan: ACM Press, 2017, pp. 1–6. https://doi.org/10.1145/3152723.3152724.

[PS45] J. Yan, C. Zong, and J. Xu, "Combination of Loss-based Active Learning and Semi-supervised Learning for Recognizing Entities in Chinese Electronic Medical Records," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 22, no. 5, pp. 1–19, May 2023, https://doi.org/10.1145/3588314.

[PS46] D.-N. Nguyen, L.-A. Le-Pham, H.-D. Nguyen, and M.-T. Tran, "Combining Deep Learning And Medical Knowledge to Detect Cadiomegaly and Pleural Effusion in Chest X-rays Diagnosis," in *Proceedings of the 12th International Symposium on Information and Communication Technology*, Ho Chi Minh Vietnam: ACM, Dec. 2023, pp. 562–569. https://doi.org/10.1145/3628797.3628981.

[PS47] A. Charisi and V. Megalooikonomou, “Content-based medical image retrieval in peer-to-peer systems,” in *Proceedings of the 1st ACM International Health Informatics Symposium*, Arlington Virginia USA: ACM, Nov. 2010, pp. 724–733. https://doi.org/10.1145/1882992.1883103.

[PS48] R. Ding, F. Rong, X. Han, and L. Wang, “Cross-center Early Sepsis Recognition by Medical Knowledge Guided Collaborative Learning for Data-scarce Hospitals,” in *Proceedings of the ACM Web Conference 2023*, Austin TX USA: ACM, Apr. 2023, pp. 3987–3993. https://doi.org/10.1145/3543507.3583989.

[PS49] R. W. White and E. Horvitz, “Cyberchondria: Studies of the escalation of medical concerns in Web search,” *ACM Trans. Inf. Syst.*, vol. 27, no. 4, pp. 1–37, Nov. 2009, https://doi.org/10.1145/1629096.1629101.

[PS50] H. Grisstte and El. Nfaoui, “Daily life patients sentiment analysis model based on well-encoded embedding vocabulary for related-medication text,” in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Vancouver British Columbia Canada: ACM, Aug. 2019, pp. 921–928. https://doi.org/10.1145/3341161.3343854.

[PS51] Z. Zhang, K. Joy, P. Upadhyayula, M. Ozkaynak, R. Harris, and K. Adelgais, “Data Work and Decision Making in Emergency Medical Services: A Distributed Cognition Perspective,” *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW2, pp. 1–32, Oct. 2021, https://doi.org/10.1145/3479500.

[PS52] A. Choudhury, S. Kaushik, and V. Dutt, “Data-mining medications for pain: analyzing adoption behavior of physicians and switching behavior of patients,” in *Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing*, Cambridge United Kingdom: ACM, Mar. 2017, pp. 1–8. https://doi.org/10.1145/3018896.3025164.

[PS53] Z. Zheng *et al.*, “DDR: Dialogue Based Doctor Recommendation for Online Medical Service,” in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Washington DC USA: ACM, Aug. 2022, pp. 4592–4600. https://doi.org/10.1145/3534678.3539201.

[PS54] M. S. Malik, S. Jawad, S. A. Moqurrab, and G. Srivastava, “DeepMedFeature: An Accurate Feature Extraction and Drug-Drug Interaction Model for Clinical Text in Medical Informatics,” *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, p. 3651159, Mar. 2024, https://doi.org/10.1145/3651159.

[PS55] A. Abbasi, F. “Mariam” Zahedi, and S. Kaza, “Detecting Fake Medical Web Sites Using Recursive Trust Labeling,” *ACM Trans. Inf. Syst.*, vol. 30, no. 4, pp. 1–36, Nov. 2012, https://doi.org/10.1145/2382438.2382441.

[PS56] N. J. Apao, L. S. Feliscuzo, and C. L. C. Sta. Romana, “Developing a Patient Information and Descriptive Analytics System for Data Actors of University of Bohol Medical and Rehabilitation Center: Towards Policy Making,” in *Proceedings of the 2019 2nd International Conference on Data Science and Information Technology*, Seoul Republic of Korea: ACM, Jul. 2019, pp. 42–48. https://doi.org/10.1145/3352411.3352419.

[PS57] M. Mannino, J. Fredrickson, F. Banaei-Kashani, I. Linck, and R. A. Raghda, “Development and Evaluation of a Similarity Measure for Medical Event Sequences,” *ACM Trans. Manage. Inf. Syst.*, vol. 8, no. 2–3, pp. 1–26, Sep. 2017, https://doi.org/10.1145/3070684.

[PS58] M. M. Al Aziz, T. Ahmed, T. Faequa, X. Jiang, Y. Yao, and N. Mohammed, “Differentially Private Medical Texts Generation Using Generative Neural Networks,” *ACM Trans. Comput. Healthcare*, vol. 3, no. 1, pp. 1–27, Jan. 2022, https://doi.org/10.1145/3469035.

[PS59] G. Malykhina, V. Salnikov, V. Semenyutin, and D. Tarkhov, “Digitalization of medical services for detecting violations of cerebrovascular regulation based on a neural network signal analysis algorithm,” in *Proceedings of the 2nd International Scientific Conference on Innovations in Digital Economy: SPBPU IDE-2020*, Saint Petersburg Russian Federation: ACM, Oct. 2020, pp. 1–7. https://doi.org/10.1145/3444465.3444526.

[PS60] M. H. Hodge, “Direct use by physicians of the TDS medical information system,” in *A history of medical informatics*, B. I. Blum and K. Duncan, Eds., New York, NY, USA: ACM, 1990, pp. 345–354. https://doi.org/10.1145/89482.89512.

[PS61] S. Eftekhari, N. Yaraghi, R. Singh, R. D. Gopal, and R. Ramesh, “Do Health Information Exchanges Deter Repetition of Medical Services?,” *ACM Trans. Manage. Inf. Syst.*, vol. 8, no. 1, pp. 1–27, Mar. 2017, https://doi.org/10.1145/3057272.

[PS62] C.-F. Lee, T.-Y. Kang, and I.-C. Lin, “Dual Medical-Image Reversible Data Embedding Method with Vector Directions,” in *Proceedings of the 2024 8th International Conference on Medical and Health Informatics*, Yokohama Japan: ACM, May 2024, pp. 244–250. https://doi.org/10.1145/3673971.3674017.

[PS63] Y. Cheng *et al.*, “Efficient Low-Resource Medical Information Processing Based on Semantic Analysis and Granular Computing,” *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, p. 3626319, Oct. 2023, https://doi.org/10.1145/3626319.

[PS64] G. Marfia, M. Roccetti, C. E. Palazzi, and A. Amoroso, “Efficient vehicle-to-pedestrian exchange of medical data: an empirical model with preliminary results”.

[PS65] M. A. Rahman, S. M. Preum, R. D. Williams, H. Alemzadeh, and J. Stankovic, “EMS-BERT: A Pre-Trained Language Representation Model for the Emergency Medical Services (EMS) Domain,” in *Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies*, Orlando FL USA: ACM, Jun. 2023, pp. 34–43. https://doi.org/10.1145/3580252.3586978.

[PS66] J. Oduro-Afriyie and H. M. Jamil, “Enabling the Informed Patient Paradigm with Secure and Personalized Medical Question Answering,” in *Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Houston TX USA: ACM, Sep. 2023, pp. 1–6. https://doi.org/10.1145/3584371.3613016.

[PS67] K. Bugrara, “Enhancing Inter-Operability and Data Sharing In Medical Information Systems”.

[PS68] R. Kumar, D. R. K. Gattani, and K. Singh, “Enhancing Medical History Collection using LLMs,” in *Proceedings of the 2024 Australasian Computer Science Week*, Sydney NSW Australia: ACM, Jan. 2024, pp. 140–143. https://doi.org/10.1145/3641142.3641174.

[PS69] C. Solomou, “Enhancing Medical Specialty Assignment to Patients using NLP Techniques,” in *Proceedings of the 2023 7th International Conference on Natural Language Processing and Information Retrieval*, Seoul Republic of Korea: ACM, Dec. 2023, pp. 203–209. https://doi.org/10.1145/3639233.3639251.

[PS70] I. A. Knight, M. L. Wilson, D. F. Brailsford, and N. Milic-Frayling, “Enslaved to the Trapped Data: A Cognitive Work Analysis of Medical Systematic Reviews,” in *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, Glasgow Scotland UK: ACM, Mar. 2019, pp. 203–212. https://doi.org/10.1145/3295750.3298937.

[PS71] R. Miikkulainen *et al.*, “Evaluating medical aesthetics treatments through evolved age-estimation models,” in *Proceedings of the Genetic and Evolutionary Computation Conference*, Lille France: ACM, Jun. 2021, pp. 1009–1017. https://doi.org/10.1145/3449639.3459378.

[PS72] A. M. Elghriani, A. M. Maatuk, E. K. Elberkawi, and T. El-Turki, “Evaluation of the Knowledge of Medical and Health Informatics Students for Bioinformatics and Biomedical Research,” in *The 7th International Conference on Engineering & MIS 2021*, Almaty Kazakhstan: ACM, Oct. 2021, pp. 1–7. https://doi.org/10.1145/3492547.3492588.

[PS73] A. M. Cohen *et al.*, “Evidence-based medicine, the essential role of systematic reviews, and the need for automated text mining tools,” in *Proceedings of the 1st ACM International Health Informatics Symposium*, Arlington Virginia USA: ACM, Nov. 2010, pp. 376–380. https://doi.org/10.1145/1882992.1883046.

[PS74] G. S. Avrunin *et al.*, “Experience modeling and analyzing medical processes: UMass/baystate medical safety project overview”.

[PS75] D. Patnaik, P. Butler, N. Ramakrishnan, L. Parida, B. J. Keller, and D. A. Hanauer, “Experiences with mining temporal event sequences from electronic medical records: initial successes and some challenges”, 2019

[PS76] A. M. Storås, I. Strümke, M. A. Riegler, and P. Halvorsen, “Explainability methods for machine learning systems for multimodal medical datasets: research proposal,” in *Proceedings of the 13th ACM Multimedia*

Systems Conference, Athlone Ireland: ACM, Jun. 2022, pp. 347–351. https://doi.org/10.1145/3524273.3533925.

[PS77] C. Patrício, J. C. Neves, and L. F. Teixeira, "Explainable Deep Learning Methods in Medical Image Classification: A Survey," *ACM Comput. Surv.*, vol. 56, no. 4, pp. 1–41, Apr. 2024, https://doi.org/10.1145/3625287.

[PS78] H. Kambhamettu, D. Metaxa, K. Johnson, and A. Head, "Explainable Notes: Examining How to Unlock Meaning in Medical Notes with Interactivity and Artificial Intelligence," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, May 2024, pp. 1–19. https://doi.org/10.1145/3613904.3642573.

[PS79] M. Tian *et al.*, "Exploiting collaborative learning for concept extraction in the medical field," in *Proceedings of the 2nd International Conference on Communication and Information Processing*, Singapore Singapore: ACM, Nov. 2016, pp. 1–5. https://doi.org/10.1145/3018009.3018054.

[PS80] O. Tamburis, E. Masciari, and G. Fatone, "Exploratory analysis of methods for automated classification of clinical diagnoses in Veterinary Medicine," in *25th International Database Engineering & Applications Symposium*, Montreal QC Canada: ACM, Jul. 2021, pp. 141–148. https://doi.org/10.1145/3472163.3472165.

[PS81] A. K. Baughman *et al.*, "Exploring an evolutionary medical analytic wallet," in *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*, Montreal Québec Canada: ACM, Jul. 2009, pp. 1987–1992. https://doi.org/10.1145/1570256.1570264.

[PS82] S. Scepanovic, E. Martin-Lopez, D. Quercia, and K. Baykaner, "Extracting medical entities from social media," in *Proceedings of the ACM Conference on Health, Inference, and Learning*, Toronto Ontario Canada: ACM, Apr. 2020, pp. 170–181. https://doi.org/10.1145/3368555.3384467.

[PS83] F. Ren, S. Yuan, and F. Gao, "Extraction of Transitional Relations in Healthcare Processes from Chinese Medical Text based on Deep Learning," in *Proceedings of the 2019 4th International Conference on Mathematics and Artificial Intelligence*, Chegndu China: ACM, Apr. 2019, pp. 56–60. https://doi.org/10.1145/3325730.3325771.

[PS84] N. M. T. Phan, L. Chen, C.-H. Chen, and W.-C. Peng, "FastRx: Exploring Fastformer and Memory-Augmented Graph Neural Networks for Personalized Medication Recommendations," *ACM Trans. Intell. Syst. Technol.*, p. 3696111, Sep. 2024, https://doi.org/10.1145/3696111.

[PS85] G. Canino, Q. Suo, P. H. Guzzi, G. Tradigo, A. Zhang, and P. Veltri, "Feature Selection Model for Diagnosis, Electronic Medical Records and Geographical Data Correlation," in *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Seattle WA USA: ACM, Oct. 2016, pp. 616–621. https://doi.org/10.1145/2975167.2985847.

[PS86] B. Pfitzner, N. Steckhan, and B. Arnrich, "Federated Learning in a Medical Context: A Systematic Literature Review," *ACM Trans. Internet Technol.*, vol. 21, no. 2, pp. 1–31, Jun. 2021, https://doi.org/10.1145/3412357.

[PS87] S. Greengard, "Finding a healthier approach to managing medical data," *Commun. ACM*, vol. 61, no. 5, pp. 31–33, Apr. 2018, https://doi.org/10.1145/3193759.

[PS88] A. Bellot and M. V. D. Schaar, "Flexible Modelling of Longitudinal Medical Data: A Bayesian Nonparametric Approach," *ACM Trans. Comput. Healthcare*, vol. 1, no. 1, pp. 1–15, Jan. 2020, https://doi.org/10.1145/3377164.

[PS89] M. Fischer and M. Lam, "From Books to Bots: Using Medical Literature to Create a Chat Bot," in *Proceedings of the First Workshop on IoT-enabled Healthcare and Wellness Technologies and Systems*, Singapore Singapore: ACM, Jun. 2016, pp. 23–28. https://doi.org/10.1145/2933566.2933573.

[PS90] C. Ng, A. Gruzd, C. Cheng, B. Crocker, D. Doiron, and K. Stevens, "From data to knowledge: discovery of medical laboratory demand patterns through visualisation techniques," in *Proceedings of the 2012 iConference*, Toronto Ontario Canada: ACM, Feb. 2012, pp. 585–586. https://doi.org/10.1145/2132176.2132298.

[PS91] F. Nunes, "From medicalized to mundane self-care technologies," *interactions*, vol. 26, no. 3, pp. 67–69, Apr. 2019, https://doi.org/10.1145/3319374.

[PS92] P. Kumari and S. Sharma, "Fuzzy based medicine recommendation system: an example of thyroid medicine," in *Proceedings of the Third International Conference on Advanced Informatics for Computing Research*, Shimla India: ACM, Jun. 2019, pp. 1–7. https://doi.org/10.1145/3339311.3339346.

[PS93] R. Gatta *et al.*, "Generating and Comparing Knowledge Graphs of Medical Processes Using pMineR," in *Proceedings of the Knowledge Capture Conference*, Austin TX USA: ACM, Dec. 2017, pp. 1–4. https://doi.org/10.1145/3148011.3154464.

[PS94] S. F. Bokhare, W. M. N. Wan Zainon, and A. Z. Talib, "Genogram Visualization Using Social Network Approach for Medical Family Tree," in *Proceedings of the 11th International Conference on Computer Modeling and Simulation*, North Rockhampton QLD Australia: ACM, Jan. 2019, pp. 214–218. https://doi.org/10.1145/3307363.3307367.

[PS95] S. Cipiere *et al.*, "Global Initiative for Sentinel e-Health Network on Grid (GINSENG): Medical Data Integration and Semantic Developments for Epidemiology," in *2014 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, Chicago, IL, USA: IEEE, May 2014, pp. 755–763. https://doi.org/10.1109/CCGrid.2014.45.

[PS96] G. Wu, S. S. Krishna, S. Sakai, R. Jhangiani, and S. Kurniawan, "Google Assistant and ChatGPT: Is it Useful for Non Medical Professionals Looking for Information about Stroke or Glaucoma?," in *Proceedings of the International Conference on the Internet of Things*, Nagoya Japan: ACM, Nov. 2023, pp. 240–244. https://doi.org/10.1145/3627050.3631579.

[PS97] H. D. Zając, N. R. Avlona, F. Kensing, T. O. Andersen, and I. Shklovski, "Ground Truth Or Dare: Factors Affecting The Creation Of Medical Datasets For Training AI," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, Montr\'{e}al QC Canada: ACM, Aug. 2023, pp. 351–362. https://doi.org/10.1145/3600211.3604766.

[PS98] I. Lee and M. Pajic, Eds., "Guest Editorial: Special Issue on Medical Cyber-Physical Systems," *ACM Trans. Cyber-Phys. Syst.*, vol. 2, no. 4, pp. 1–2, Oct. 2018, https://doi.org/10.1145/3243122.

[PS99] S. Aswal, N. J. Ahuja, and R. MElectronic Health Recordsa, "Handling Imbalance and Limited Data in Thyroid Ultrasound and Diabetic Retinopathy Datasets Using Discrete Levy Flights Grey Wolf Optimizer Based Random Forest for Robust Medical Data Classification," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, p. 3648363, Feb. 2024, https://doi.org/10.1145/3648363.

[PS100] A. M. Elghriani, A. A. Alabbar, A. M. Maatuk, and E. A. Elfallah, "Health Care Workers' Use of Electronic Medical Information Systems: Benefits and Challenges," in *International Conference on Data Science, E-learning and Information Systems 2021*, Ma'an Jordan: ACM, Apr. 2021, pp. 35–39. https://doi.org/10.1145/3460620.3460627.

[PS101] E. W. Huang *et al.*, "HEMnet: Integration of Electronic Medical Records with Molecular Interaction Networks and Domain Knowledge for Survival Analysis," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Boston Massachusetts USA: ACM, Aug. 2017, pp. 378–387. https://doi.org/10.1145/3107411.3107422.

[PS102] M. H. Hodge, "History of the TDS medical information system," in *A history of medical informatics*, B. I. Blum and K. Duncan, Eds., New York, NY, USA: ACM, 1990, pp. 328–344. https://doi.org/10.1145/89482.89511.

[PS103] W. Yang, Z. Wu, N. Y. Mok, and X. Ma, "How to Save Lives with Microblogs? Lessons From the Usage of Weibo for Requests for Medical Assistance During COVID-19," in *CHI Conference on Human Factors in Computing Systems*, New Orleans LA USA: ACM, Apr. 2022, pp. 1–18. https://doi.org/10.1145/3491102.3517591.

[PS104] F. Luo, Y. Zhang, and X. Wang, "IMAS++: An Intelligent Medical Analysis System Enhanced with Deep Graph Neural Networks," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Virtual Event Queensland Australia: ACM, Oct. 2021, pp. 4754–4758. https://doi.org/10.1145/3459637.3481966.

[PS105] X. Li, J. You, and W. Zhao, "Impact Analysis of Mobile Medical APPs Based on AHP-TOPSIS: A Case Study of China," in *Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence*, Chengdu China: ACM, Apr. 2020, pp. 214–218. https://doi.org/10.1145/3395260.3395272.

[PS106] P. Karmalkar, H. Gurulingappa, J. Muhith, S. Singhal, G. Megaro, and F. Buchholz, "Improving Consumer Experience for Medical Information Using Text Analytics," in *2021 International Symposium on Electrical, Electronics and Information Engineering*, Seoul Republic of Korea: ACM, Feb. 2021, pp. 471–476. https://doi.org/10.1145/3459104.3459182.

[PS107] X. Ji, H. Choi, O. Sokolsky, and I. Lee, "Incremental Anomaly Detection with Guarantee in the Internet of Medical Things," in *Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation*, San Antonio TX USA: ACM, May 2023, pp. 327–339. https://doi.org/10.1145/3576842.3582374.

[PS108] A. Yera *et al.*, "Inferring Visual Behaviour from User Interaction Data on a Medical Dashboard," in *Proceedings of the 2018 International Conference on Digital Health*, Lyon France: ACM, Apr. 2018, pp. 55–59. https://doi.org/10.1145/3194658.3194676.

[PS109] R. Gazzotti, C. Faron-Zucker, F. Gandon, V. Lacroix-Hugues, and D. Darmon, "Injection of automatically selected DBpedia subjects in electronic medical records to boost hospitalization prediction," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, Brno Czech Republic: ACM, Mar. 2020, pp. 2013–2020. https://doi.org/10.1145/3341105.3373932.

[PS110] "Inside Chronic Autoimmune Disease Communities: A Social Networks Perspective to Crohn's Patient Behavior and Medical Information," in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Istanbul, Turkey: IEEE, 2012.

[PS111] X. Shao, "Integration of Intelligent Drug Systems and Personalized Medicine: Applications and Prospects," in *Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science*, Chengdu China: ACM, Oct. 2023, pp. 1338–1342. https://doi.org/10.1145/3644116.3644344.

[PS112] H. Singal and S. Kohli, "Intellectualizing TRUST for Medical Websites," in *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, Udaipur India: ACM, Mar. 2016, pp. 1–4. https://doi.org/10.1145/2905055.2905293.

[PS113] Z. Zheng, L. Wang, S. Qiao, Z. Zhou, and J. Chen, "Intelligent medical consultation system based on the GPT model," in *Proceedings of the 2023 7th International Conference on Electronic Information Technology and Computer Engineering*, Xiamen China: ACM, Oct. 2023, pp. 1192–1197. https://doi.org/10.1145/3650400.3650601.

[PS114] A. Saxena and S. Mittal, "Internet of Medical Things (IoMT) Security and Privacy: A Survey of Recent Advances and Enabling Technologies," in *Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing*, Noida India: ACM, Aug. 2022, pp. 550–559. https://doi.org/10.1145/3549206.3549301.

[PS115] A. Karunakaran, Y. Hee-Nam, and M. Reddy, "Investigating barriers to electronic medical record use during collaborative information seeking activities," in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, Miami Florida USA: ACM, Jan. 2012, pp. 743–748. https://doi.org/10.1145/2110363.2110451.

[PS116] D. Chen, M. Huang, and W. Li, "Knowledge-Powered Deep Breast Tumor Classification With Multiple Medical Reports," *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, vol. 18, no. 3, pp. 891–901, May 2021, https://doi.org/10.1109/TCBB.2019.2955484.

[PS117] S. Oumaima, K. Soulaimane, and B. Omar, "Latest Trends in Recommender Systems applied in the medical domain: A Systematic Review," in *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, Marrakech Morocco: ACM, Mar. 2020, pp. 1–12. https://doi.org/10.1145/3386723.3387860.

[PS118] N. Limsopatham, C. Macdonald, and I. Ounis, "Learning to selectively rank patients' medical history," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management - CIKM '13*, San Francisco, California, USA: ACM Press, 2013, pp. 1833–1836. https://doi.org/10.1145/2505515.2507874.

[PS119] Z. Zhao, Y. Jing, F. Feng, J. Wu, C. Gao, and X. He, "Leave No Patient Behind: Enhancing Medication Recommendation for Rare Disease Patients," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Washington DC USA: ACM, Jul. 2024, pp. 533–542. https://doi.org/10.1145/3626772.3657785.

[PS120] S. Luz, "Locating case discussion segments in recorded medical team meetings," in *Proceedings of the third workshop on Searching spontaneous conversational speech*, Beijing China: ACM, Oct. 2009, pp. 21–30. https://doi.org/10.1145/1631127.1631131.

[PS121] Q. Li *et al.*, "MaraVis: Representation and Coordinated Intervention of Medical Encounters in Urban Marathon," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–12. https://doi.org/10.1145/3313831.3376281.

[PS122] J. I. L. Palisoc and M. I. Norona, "Mathematical Modeling and Analysis of Factors Affecting Medical Errors in Emergency Department of Government Hospitals in Province X," in *Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering*, Osaka Japan: ACM, Apr. 2020, pp. 39–43. https://doi.org/10.1145/3396743.3396795.

[PS123] Z. S. A. Albayood, E. A. O. Elfallah, and A. Elghriani, "Measuring the awareness of health care providers at Benghazi Medical Center for Health Informatics," in *Proceedings of the 6th International Conference on Engineering & MIS 2020*, Almaty Kazakhstan: ACM, Sep. 2020, pp. 1–6. https://doi.org/10.1145/3410352.3410733.

[PS124] S. L. Smith, "Medical applications of evolutionary computation," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, Kyoto Japan: ACM, Jul. 2018, pp. 1141–1169. https://doi.org/10.1145/3205651.3207873.

[PS125] D. Mondal, A. Gangopadhyay, and W. Russell, "Medical decision making using vector space model," in *Proceedings of the 1st ACM International Health Informatics Symposium*, Arlington Virginia USA: ACM, Nov. 2010, pp. 386–390. https://doi.org/10.1145/1882992.1883048.

[PS126] S. Saxena, N. Parveen, and D. Agarwal, "Medical Image Analysis Using Deep Learning Technnique," in *Proceedings of the 5th International Conference on Information Management & Machine Intelligence*, Jaipur India: ACM, Nov. 2023, pp. 1–8. https://doi.org/10.1145/3647444.3647938.

[PS127] D. Zhao, F. Zhang, W. Wang, Z. Xiao, and L. Geng, "Medical Images Super Resolution Reconstruction based on Residual Network," in *2021 7th International Conference on Computing and Artificial Intelligence*, Tianjin China: ACM, Apr. 2021, pp. 119–126. https://doi.org/10.1145/3467707.3467724.

[PS128] J. C. G. Ramirez, L. A. Smith, and L. L. Peterson, "Medical information systems: characterization and challenges," *SIGMOD Rec.*, vol. 23, no. 3, pp. 44–53, Sep. 1994, https://doi.org/10.1145/187436.187458.

[PS129] W. Lu, S. Wei, X. Peng, Y.-F. Wang, U. Naseem, and S. Wang, "Medical Question Summarization with Entity-driven Contrastive Learning," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 23, no. 4, pp. 1–19, Apr. 2024, https://doi.org/10.1145/3652160.

[PS130] A. Z. Qashri, O. Karnalim, and H. Toba, "Medical Specialists Retrieval System Using Unified Medical Language System," in *Proceedings of the 1st International Conference on Medical and Health Informatics 2017*, Taichung City Taiwan: ACM, May 2017, pp. 28–34. https://doi.org/10.1145/3107514.3107521.

[PS131] Z. Liu, X. Li, Z. You, T. Yang, W. Fan, and P. Yu, "Medical Triage Chatbot Diagnosis Improvement via Multi-relational Hyperbolic Graph Neural Network," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event Canada: ACM, Jul. 2021, pp. 1965–1969. https://doi.org/10.1145/3404835.3463095.

[PS132] T. O. Andersen, "Medication management in the making: on ethnography-design relations," in *Proceedings of the 2013 conference on Computer supported cooperative work*, San Antonio Texas USA: ACM, Feb. 2013, pp. 1103–1112. https://doi.org/10.1145/2441776.2441901.

[PS133] S. A. Carro, J. Scharcanski, and J. V. de Lima, "MedISeek: A Web Based Diffusion System for Medical Visual Information". 2003

[PS134] Y. Zhuang, J. Zhang, R. Lu, K. He, and X. Li, “MedNER: Enhanced Named Entity Recognition in Medical Corpus via Optimized Balanced and Deep Active Learning,” *ACM Trans. Intell. Syst. Technol.*, p. 3678178, Jul. 2024, https://doi.org/10.1145/3678178.

[PS135] M. Ye, S. Cui, Y. Wang, J. Luo, C. Xiao, and F. Ma, “MedRetriever: Target-Driven Interpretable Health Risk Prediction via Retrieving Unstructured Medical Text,” in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Virtual Event Queensland Australia: ACM, Oct. 2021, pp. 2414–2423. https://doi.org/10.1145/3459637.3482273.

[PS136] Y. Deng *et al.*, “MedTruth: A Semi-supervised Approach to Discovering Knowledge Condition Information from Multi-Source Medical Data,” in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing China: ACM, Nov. 2019, pp. 719–728. https://doi.org/10.1145/3357384.3357934.

[PS137] T.-C. Tran, T. B. Dinh, and V. Gascon, “Meta-heuristics to Solve a Districting Problem of a Public Medical Clinic,” in *Proceedings of the Eighth International Symposium on Information and Communication Technology*, Nha Trang City Viet Nam: ACM, Dec. 2017, pp. 127–134. https://doi.org/10.1145/3155133.3155146.

[PS138] H. H. Le *et al.*, “Methods for Analyzing Medical-Order Sequence Variants in Sequential Pattern Mining for Electronic Medical Record Systems,” *ACM Trans. Comput. Healthcare*, vol. 4, no. 1, pp. 1–28, Jan. 2023, https://doi.org/10.1145/3561825.

[PS139] X. S. Zhou, Y. Zhan, V. C. Raykar, G. Hermosillo, L. Bogoni, and Z. Peng, “Mining anatomical, physiological and pathological information from medical images,” *SIGKDD Explor. Newsl.*, vol. 14, no. 1, pp. 25–34, Dec. 2012, https://doi.org/10.1145/2408736.2408741.

[PS140] M. Adamou, G. Antoniou, E. Greasidou, V. Lagani, P. Charonyktakis, and I. Tsamardinos, “Mining Free-Text Medical Notes for Suicide Risk Assessment,” in *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*, Patras Greece: ACM, Jul. 2018, pp. 1–8. https://doi.org/10.1145/3200947.3201020.

[PS141] Y. Wang, W. Chen, D. Pi, L. Yue, M. Xu, and X. Li, “Multi-hop Reading on Memory Neural Network with Selective Coverage for Medication Recommendation,” in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Virtual Event Queensland Australia: ACM, Oct. 2021, pp. 2020–2029. https://doi.org/10.1145/3459637.3482278.

[PS142] E. Choi *et al.*, “Multi-layer Representation Learning for Medical Concepts,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 1495–1504. https://doi.org/10.1145/2939672.2939823.

[PS143] Y. Zhang, S. Qian, Q. Fang, and C. Xu, “Multi-modal Knowledge-aware Hierarchical Attention Network for Explainable Medical Question Answering,” in *Proceedings of the 27th ACM International Conference on Multimedia*, Nice France: ACM, Oct. 2019, pp. 1089–1097. https://doi.org/10.1145/3343031.3351033.

[PS144] F. Jovan, C. Morgan, R. McConville, E. L. Tonkin, I. Craddock, and A. Whone, “Multimodal Indoor Localisation in Parkinson’s Disease for Detecting Medication Use: Observational Pilot Study in a Free-Living Setting,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Long Beach CA USA: ACM, Aug. 2023, pp. 4273–4283. https://doi.org/10.1145/3580305.3599872.

[PS145] J. Kalpathy-Cramer and W. Hersh, “Multimodal medical image retrieval: image categorization to improve search precision”.

[PS146] D. Zheng, H. Zhang, and F. Yu, “Named entity recognition of Chinese electronic medical rec-ords based on adversarial training and feature fusion,” in *Proceedings of the 2023 International Joint Conference on Robotics and Artificial Intelligence*, Shanghai China: ACM, Jul. 2023, pp. 175–179. https://doi.org/10.1145/3632971.3632983.

[PS147] X. Xing *et al.*, “Non-imaging Medical Data Synthesis for Trustworthy AI: A Comprehensive Survey,” *ACM Comput. Surv.*, vol. 56, no. 7, pp. 1–35, Jul. 2024, https://doi.org/10.1145/3614425.

[PS148] B. Seref and E. Bostanci, “Opportunities, Threats and Future Directions in Big Data for Medical Wearables,” in *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*, Blagoevgrad Bulgaria: ACM, Nov. 2016, pp. 1–5. https://doi.org/10.1145/3010089.3010100.

[PS149] A. Tiwari and M. Pant, "Optimized Deep-Neural Network for Content-based Medical Image Retrieval in a Brownfield IoMT Network," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 18, no. 2s, pp. 1–26, Jun. 2022, https://doi.org/10.1145/3546194.

[PS150] S. Wang, P. Ren, Z. Chen, Z. Ren, J. Ma, and M. De Rijke, "Order-free Medicine Combination Prediction with Graph Convolutional Reinforcement Learning," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing China: ACM, Nov. 2019, pp. 1623–1632. https://doi.org/10.1145/3357384.3357965.

[PS151] T. August, L. L. Wang, J. Bragg, M. A. Hearst, A. Head, and K. Lo, "Paper Plain : Making Medical Research Papers Approachable to Healthcare Consumers with Natural Language Processing," *ACM Trans. Comput.-Hum. Interact.*, vol. 30, no. 5, pp. 1–38, Oct. 2023, https://doi.org/10.1145/3589955.

[PS152] D. Sonntag, "Part III: Emergency Trends: Medical and health systems," in *The Handbook of Multimodal-Multisensor Interfaces: Language Processing, Software, Commercialization, and Emerging Directions*, Association for Computing Machinery and Morgan & Claypool, 2019, pp. 423–476. Accessed: Oct. 20, 2024. [Online]. Available: https://doi.org/10.1145/3233795.3233808

[PS153] Q. B. M. Le and C. T. N. Vo, "Patient Information Retrieval Based on BERT Variants and Clinical Texts in Electronic Medical Records," in *Proceedings of the 2023 7th International Conference on Big Data and Internet of Things*, Beijing China: ACM, Aug. 2023, pp. 188–194. https://doi.org/10.1145/3617695.3617702.

[PS154] E. Tavares *et al.*, "Performance evaluation of medical imaging service," in *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, Trento Italy: ACM, Mar. 2012, pp. 1349–1354. https://doi.org/10.1145/2245276.2231990.

[PS155] V. Duvvuri, Q. Guan, S. Daddala, M. Harris, and S. Kaushik, "Predicting Depression Symptoms from Discord Chat Messaging Using AI Medical Chatbots," in *2022 The 6th International Conference on Machine Learning and Soft Computing*, Haikou China: ACM, Jan. 2022, pp. 111–119. https://doi.org/10.1145/3523150.3523168.

[PS156] D. A. Davis, N. V. Chawla, N. Blumm, N. Christakis, and A.-L. Barabasi, "Predicting individual disease risk based on medical history," in *Proceedings of the 17th ACM conference on Information and knowledge management*, Napa Valley California USA: ACM, Oct. 2008, pp. 769–778. https://doi.org/10.1145/1458082.1458185.

[PS157] Q. Lu, T. H. Nguyen, and D. Dou, "Predicting Patient Readmission Risk from Medical Text via Knowledge Graph Enhanced Multiview Graph Convolution," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event Canada: ACM, Jul. 2021, pp. 1990–1994. https://doi.org/10.1145/3404835.3463062.

[PS158] Y. Baba *et al.*, "Predictive Approaches for Low-Cost Preventive Medicine Program in Developing Countries," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney NSW Australia: ACM, Aug. 2015, pp. 1681–1690. https://doi.org/10.1145/2783258.2788587.

[PS159] W. Mukono and T. O. Tokosi, "Premier Service Medical Investments: Challenges and perceptions of healthcare practitioners in the adoption and implementation of healthcare information technology (HIT)," in *Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019*, Skukuza South Africa: ACM, Sep. 2019, pp. 1–10. https://doi.org/10.1145/3351108.3351114.

[PS160] Z. Yue *et al.*, "Privacy-preserving Time-series Medical Images Analysis Using a Hybrid Deep Learning Framework," *ACM Trans. Internet Technol.*, vol. 21, no. 3, pp. 1–21, Aug. 2021, https://doi.org/10.1145/3383779.

[PS161] G. Gefen, O. Ben-Porat, M. Tennenholtz, and E. Yom-Tov, "Privacy, Altruism, and Experience: Estimating the Perceived Value of Internet Data for Medical Uses," in *Companion Proceedings of the Web Conference 2020*, Taipei Taiwan: ACM, Apr. 2020, pp. 552–556. https://doi.org/10.1145/3366424.3383414.

[PS162] H. M. Conboy, G. S. Avrunin, and L. A. Clarke, "Process-based derivation of requirements for medical devices," in *Proceedings of the 1st ACM International Health Informatics Symposium*, Arlington Virginia USA: ACM, Nov. 2010, pp. 656–665. https://doi.org/10.1145/1882992.1883095.

[PS163] A. Ema, K. Nagakura, and T. Fujita, “Proposal for Type Classification for Building Trust in Medical Artificial Intelligence Systems,” in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, New York NY USA: ACM, Feb. 2020, pp. 251–257. https://doi.org/10.1145/3375627.3375846.

[PS164] H. Kawamura *et al.*, “pSCOUTER: Time-Series Emotion Classification Using Contactless Measured Multimodal Biosignals in Medical Diagnosis,” in *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Melbourne VIC Australia: ACM, Oct. 2024, pp. 46–50. https://doi.org/10.1145/3675094.3677582.

[PS165] A. Tommasel, R. Pablos Sarabia, and I. Assent, “Re2Dan: Retrieval of Medical Documents for e-Health in Danish,” in *Proceedings of the 17th ACM Conference on Recommender Systems*, Singapore Singapore: ACM, Sep. 2023, pp. 1208–1211. https://doi.org/10.1145/3604915.3610655.

[PS166] J. Xie, B. Zhang, J. Ma, D. Zeng, and J. Lo-Ciganic, “Readmission Prediction for Patients with Heterogeneous Medical History: A Trajectory-Based Deep Learning Approach,” *ACM Trans. Manage. Inf. Syst.*, vol. 13, no. 2, pp. 1–27, Jun. 2022, https://doi.org/10.1145/3468780.

[PS167] J.-F. Rajotte *et al.*, “Reducing bias and increasing utility by federated generative modeling of medical images using a centralized adversary,” in *Proceedings of the Conference on Information Technology for Social Good*, Roma Italy: ACM, Sep. 2021, pp. 79–84. https://doi.org/10.1145/3462203.3475875.

[PS168] P. Whitley, E. Ennis, N. Taaca, S. Sneha, H. Shahriar, and C. Zhang, “Reduction of Medical Errors in Emergency Medical Care,” in *Proceedings of the 19th Annual SIG Conference on Information Technology Education*, Fort Lauderdale Florida USA: ACM, Sep. 2018, pp. 160–160. https://doi.org/10.1145/3241815.3241884.

[PS169] P. Li, R. Williams, S. Gilbert, and S. Anderson, “Regulating AI/ML-enabled Medical Devices in the UK,” in *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, Edinburgh United Kingdom: ACM, Jul. 2023, pp. 1–10. https://doi.org/10.1145/3597512.3599704.

[PS170] Y. Li *et al.*, “Reliable Medical Diagnosis from Crowdsourcing: Discover Trustworthy Answers from Non-Experts,” in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, Cambridge United Kingdom: ACM, Feb. 2017, pp. 253–261. https://doi.org/10.1145/3018661.3018688.

[PS171] N. Savage, “Remaking American medicine,” *Commun. ACM*, vol. 54, no. 9, pp. 19–21, Sep. 2011, https://doi.org/10.1145/1995376.1995384.

[PS172] M. Porumb, I. Barbantan, C. Lemnaru, and R. Potolea, “REMed: automatic relation extraction from medical documents,” in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, Brussels Belgium: ACM, Dec. 2015, pp. 1–10. https://doi.org/10.1145/2837185.2837239.

[PS173] G. Yan *et al.*, “ReMeDi: Resources for Multi-domain, Multi-service, Medical Dialogues,” in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Madrid Spain: ACM, Jul. 2022, pp. 3013–3024. https://doi.org/10.1145/3477495.3531809.

[PS174] Z. Liu, X. Zhou, W. Xiong, J. Chen, T. Xu, and Z. Xu, “Research and Design of a Database for Intelligent Medication Taking in Traditional Chinese Medicine,” in *Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science*, Chengdu China: ACM, Oct. 2023, pp. 898–903. https://doi.org/10.1145/3644116.3644270.

[PS175] A. Cao, G. Xie, and M. Li, “Research on the application of Bot Chat Technology in the field of medical assistance,” in *Proceedings of the 2023 International Conference on Computer, Vision and Intelligent Technology*, Chenzhou China: ACM, Aug. 2023, pp. 1–7. https://doi.org/10.1145/3627341.3630398.

[PS176] K. Zheng, J. Gao, K. Y. Ngiam, B. C. Ooi, and W. L. J. Yip, “Resolving the Bias in Electronic Medical Records,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax NS Canada: ACM, Aug. 2017, pp. 2171–2180. https://doi.org/10.1145/3097983.3098149.

[PS177] S. Zhang *et al.*, “Rethinking Human-AI Collaboration in Complex Medical Decision Making: A Case Study in Sepsis Diagnosis,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, May 2024, pp. 1–18. https://doi.org/10.1145/3613904.3642343.

[PS178] D. Wallace and M.-T. Kechadi, "Retrieval and Clustering of Medicines Within Healthcare Data Records," in *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*, Blagoevgrad Bulgaria: ACM, Nov. 2016, pp. 1–6. https://doi.org/10.1145/3010089.3010101.

[PS179] A. R. Kelly, A. McDougall, and N. Abbott, "Rhetorical models for computational systems: an interdisciplinary approach to reusable, tailorable medical information," in *Proceedings of the 27th ACM international conference on Design of communication*, Bloomington Indiana USA: ACM, Oct. 2009, pp. 155–162. https://doi.org/10.1145/1621995.1622025.

[PS180] F. Ma, J. Gao, Q. Suo, Q. You, J. Zhou, and A. Zhang, "Risk Prediction on Electronic Health Records with Prior Medical Knowledge," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London United Kingdom: ACM, Jul. 2018, pp. 1910–1919. https://doi.org/10.1145/3219819.3220020.

[PS181] J. M. Fernández, M. J. Sánchez-Ledesma, J. M. López-Millan, J. A. J. Méndez, and M. B. García-Cenador, "Scope of the uses of information and communication technologies by occupational medicals involved in pain treatment," in *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, Salamanca Spain: ACM, Nov. 2016, pp. 495–499. https://doi.org/10.1145/3012430.3012563.

[PS182] V. Janjic, J. Bowles, M. Belk, and A. Pitsillides, "Security And Privacy Of Medical Data: Challenges For Next-Generation Patient-Centric Healthcare Systems," in *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, Larnaca Cyprus: ACM, Jun. 2019, pp. 213–214. https://doi.org/10.1145/3314183.3326364.

[PS183] L. Tamine and L. Goeuriot, "Semantic Information Retrieval on Medical Texts: Research Challenges, Survey, and Open Issues," *ACM Comput. Surv.*, vol. 54, no. 7, pp. 1–38, Sep. 2022, https://doi.org/10.1145/3462476.

[PS184] X. S. Zhou *et al.*, "Semantics and CBIR: a medical imaging perspective," in *Proceedings of the 2008 international conference on Content-based image and video retrieval*, Niagara Falls Canada: ACM, Jul. 2008, pp. 571–580. https://doi.org/10.1145/1386352.1386436.

[PS185] Y. Xu *et al.*, "SeqCare: Sequential Training with External Medical Knowledge Graph for Diagnosis Prediction in Healthcare Data," in *Proceedings of the ACM Web Conference 2023*, Austin TX USA: ACM, Apr. 2023, pp. 2819–2830. https://doi.org/10.1145/3543507.3583543.

[PS186] R. P. Jetley, P. L. Jones, and P. Anderson, "Static analysis of medical device software using CodeSonar," in *Proceedings of the 2008 workshop on Static analysis*, Tucson Arizona: ACM, Jun. 2008, pp. 22–29. https://doi.org/10.1145/1394504.1394507.

[PS187] A. Paproki, O. Salvado, and C. Fookes, "Synthetic Data for Deep Learning in Computer Vision & Medical Imaging: A Means to Reduce Data Bias," *ACM Comput. Surv.*, vol. 56, no. 11, pp. 1–37, Nov. 2024, https://doi.org/10.1145/3663759.

[PS188] Y. Chen, J. Zhao, Z. Wen, Z. Li, and Y. Xiao, "TemporalMed: Advancing Medical Dialogues with Time-Aware Responses in Large Language Models," in *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, Merida Mexico: ACM, Mar. 2024, pp. 116–124. https://doi.org/10.1145/3616855.3635860.

[PS189] H. T. Tavani and M. Bottis, "The consent process in medical research involving DNA databanks: some ethical implications and challenges," *SIGCAS Comput. Soc.*, vol. 40, no. 2, pp. 11–21, Jun. 2010, https://doi.org/10.1145/1839994.1839996.

[PS190] Y. You, Y. Kou, X. Ding, and X. Gui, "The Medical Authority of AI: A Study of AI-enabled Consumer-Facing Health Technology," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 2021, pp. 1–16. https://doi.org/10.1145/3411764.3445657.

[PS191] S. Luz, "The nonverbal structure of patient case discussions in multidisciplinary medical team meetings," *ACM Trans. Inf. Syst.*, vol. 30, no. 3, pp. 1–24, Aug. 2012, https://doi.org/10.1145/2328967.2328970.

[PS192] M. Alamir and M. Alghamdi, "The Role of Generative Adversarial Network in Medical Image Analysis: An In-depth Survey," *ACM Comput. Surv.*, vol. 55, no. 5, pp. 1–36, May 2023, https://doi.org/10.1145/3527849.

[PS193] O. Nov *et al.*, "The transformation of patient-clinician relationships with AI-based medical advice," *Commun. ACM*, vol. 64, no. 3, pp. 46–48, Mar. 2021, https://doi.org/10.1145/3417518.

[PS194] V. Heilala, S. Lehesvuori, R. Hämäläinen, and T. Kärkkäinen, "Toward Scalable and Transparent Multimodal Analytics to Study Standard Medical Procedures: Linking Hand Movement, Proximity, and Gaze Data," in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, Avila Spain: ACM, Apr. 2024, pp. 3–10. https://doi.org/10.1145/3605098.3635929.

[PS195] R. Hou, V. Perez-Rosas, S. Loeb, and R. Mihalcea, "Towards Automatic Detection of Misinformation in Online Medical Videos," in *2019 International Conference on Multimodal Interaction*, Suzhou China: ACM, Oct. 2019, pp. 235–243. https://doi.org/10.1145/3340555.3353763.

[PS196] S. Tsumoto, M. Otake, H. Nakajima, T. Fujita, and Y. Chow, "Towards data: a human/machine-oriented approach of medical data collection," in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, Miami Florida USA: ACM, Jan. 2012, pp. 887–888. https://doi.org/10.1145/2110363.2110484.

[PS197] A. P. Selvaraj and S. Annamalai, "Towards Reliable Medical Transactions for Blockchain based Healthcare Systems using Hybrid Consensus Method," in *Proceedings of the 4th International Conference on Information Management & Machine Intelligence*, Jaipur India: ACM, Dec. 2022, pp. 1–8. https://doi.org/10.1145/3590837.3590867.

[PS198] X. Zeng, S. Lin, and C. Liu, "Transformer-based unsupervised patient representation learning based on medical claims for risk stratification and analysis," in *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, Gainesville Florida: ACM, Aug. 2021, pp. 1–9. https://doi.org/10.1145/3459930.3469519.

[PS199] J. Carmichael, "Translating Human-Centred Artificial Intelligence for Clinical Decision Support Systems into Practice: A Medical Retina Case Study.," in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, May 2024, pp. 1–5. https://doi.org/10.1145/3613905.3638182.

[PS200] A. Dubovitskaya, T. Buclin, M. Schumacher, K. Aberer, and Y. Thoma, "TUCUXI: An Intelligent System for Personalized Medicine from Individualization of Treatments to Research Databases and Back," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Boston Massachusetts USA: ACM, Aug. 2017, pp. 223–232. https://doi.org/10.1145/3107411.3107439.

[PS201] D. A. R. Villamor, C. E. Pulmano, and M. R. J. E. Estuar, "Understanding Adoption of Electronic Medical Records: Application of Process Mining for Health Worker Behavior Analysis," in *Proceedings of the 4th International Conference on Medical and Health Informatics*, Kamakura City Japan: ACM, Aug. 2020, pp. 98–104. https://doi.org/10.1145/3418094.3418109.

[PS202] T.-Y. Chen, J. A. Aquino, Y.-W. Chiu, W.-C. Yeh, and Y.-C. Chang, "Understanding the Clinical Context of Medication Change Events in Clinical Narratives using Pre-trained Clinical Language Models," in *2023 the 7th International Conference on Medical and Health Informatics (ICMHI)*, Kyoto Japan: ACM, May 2023, pp. 98–103. https://doi.org/10.1145/3608298.3608318.

[PS203] D. S. Kusunoki, A. Sarcevic, Z. Zhang, and R. S. Burd, "Understanding visual attention of teams in dynamic medical settings through vital signs monitor use," in *Proceedings of the 2013 conference on Computer supported cooperative work*, San Antonio Texas USA: ACM, Feb. 2013, pp. 527–540. https://doi.org/10.1145/2441776.2441836.

[PS204] S. Ayanouz, A. A. Boudhir, and M. Ben Ahmed, "Unleashing the Power of NLP and Transformers: A Game-Changer in Medical Research and Clinical Practice and a revolution of Medical Text Analysis.: Case Study: Cancer report classification by priority," in *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security*, Larache Morocco: ACM, May 2023, pp. 1–6. https://doi.org/10.1145/3607720.3607787.

[PS205] L.-S. Chen and J.-C. Chen, "Using Data Mining Methods to Detect Medical Fraud," in *Proceedings of the 7th International Conference on Management of e-Commerce and e-Government*, Jeju Island Republic of Korea: ACM, Jul. 2020, pp. 89–93. https://doi.org/10.1145/3409891.3409902.

[PS206] J. Bowen and A. Hinze, "Using ontologies to reason about the usability of interactive medical devices in multiple situations of use," in *Proceedings of the 4th ACM SIGCHI symposium on Engineering interactive computing systems*, Copenhagen Denmark: ACM, Jun. 2012, pp. 247–256. https://doi.org/10.1145/2305484.2305525.

[PS207] R. Feldman, O. Netzer, A. Peretz, and B. Rosenfeld, "Utilizing Text Mining on Online Medical Forums to Predict Label Change due to Adverse Drug Reactions," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney NSW Australia: ACM, Aug. 2015, pp. 1779–1788. https://doi.org/10.1145/2783258.2788608.

[PS208] A. Khan, J. A. Doucette, and R. Cohen, "Validation of an ontological medical decision support system for patient treatment using a repository of patient data: Insights into the value of machine learning," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, pp. 1–31, Sep. 2013, https://doi.org/10.1145/2508037.2508049.

[PS209] C. Plaisant, D. Heller, J. Li, B. Shneiderman, R. Mushlin, and J. Karat, "Visualizing medical records with LifeLines," in *CHI 98 Conference Summary on Human Factors in Computing Systems*, Los Angeles California USA: ACM, Apr. 1998, pp. 28–29. https://doi.org/10.1145/286498.286513.

[PS210] F. Li, Y. Wang, O. Beyan, M. Schöneck, and L. L. Caldeira, "Voxel-Wise Medical Image Generalization for Eliminating Distribution Shift," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 7, pp. 1–16, Aug. 2024, https://doi.org/10.1145/3643034.

[PS211] N. Suo, "Watson for the Cloud: How IBM is Leading the Way in Medical AI Research and DevelopmentAI-Powered Mental Health Monitoring: Transforming Healthcare," in *Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science*, Chengdu China: ACM, Oct. 2023, pp. 811–817. https://doi.org/10.1145/3644116.3644253.

[PS212] J. Xie, B. Zhang, S. Brown, and D. Zeng, "Write Like a Pro or an Amateur? Effect of Medical Language Formality," *ACM Trans. Manage. Inf. Syst.*, vol. 12, no. 3, pp. 1–25, Sep. 2021, https://doi.org/10.1145/3458752.

[PS213] K. Armstrong, L. Zhang, P. Lee, and X. Ye, "Zero-dimensional biomarker based medical action recognition: towards more explainable AI in healthcare," in *Proceedings of the 2023 10th International Conference on Bioinformatics Research and Applications*, Barcelona Spain: ACM, Sep. 2023, pp. 143–148. https://doi.org/10.1145/3632047.3632069.

A2 - SCRIPT USED TO FIND REPEATED WORDS

```
import os
import re
from collections import Counter
import PyPDF2
from openpyxl import Workbook

def extract_text_from_pdf(pdf_path):
    text = ““
    try:
        with open(pdf_path, ‘rb’) as pdf_file:
            pdf_reader = PyPDF2.PdfReader(pdf_file)
            for page_num in range(len(pdf_reader.pages)):
                page = pdf_reader.pages[page_num]
                text += page.extract_text()
    except Exception as e:
        print(f”Error processing {pdf_path}: {e}”)
    return text

def count_word_frequency(text):
    words = re.findall(r’\b\w+\b’, text.lower())
    word_counts = Counter(words)
```

```
    return word_counts

pdf_directory = "/content/drive/MyDrive/pdfs"
all_word_counts = Counter()

for filename in os.listdir(pdf_directory):
    if filename.endswith(".pdf"):
        pdf_path = os.path.join(pdf_directory, filename)
        text = extract_text_from_pdf(pdf_path)
        word_counts = count_word_frequency(text)
        all_word_counts.update(word_counts)

workbook = Workbook()
sheet = workbook.active
sheet.title = "Word Counts"

sheet.append(["Word", "Count"])
for word, count in all_word_counts.most_common():
    sheet.append([word, count])

workbook.save("counts.xlsx")
```

A3 - Script Performing Named Entity Recognition on the Word Count File

```
import spacy
from openpyxl import load_workbook


# Load the spaCy model
nlp = spacy.load("en_core_web_sm")


# Load the existing Excel file
workbook = load_workbook("counts.xlsx")
sheet = workbook.active


# Create a new sheet for NER results
ner_sheet = workbook.create_sheet("NER")
ner_sheet.append(["Word", "Entity Type"])


# Process each word from the "Word Counts" sheet
for row in sheet.iter_rows(min_row=2):  # Start from row 2 to skip the header
    word = row[0].value
    if word:  # Check if the cell is not empty
        doc = nlp(word)
        for ent in doc.ents:
            ner_sheet.append([word, ent.label_])


workbook.save("NER.xlsx")
```

A4 - Script to Lemmatize the Text

```
import spacy
from openpyxl import load_workbook

# Load the spaCy model
nlp = spacy.load(“en_core_web_sm”)

# Load the Excel file
workbook = load_workbook(“NER.xlsx”)
sheet = workbook.active

# Add a new column for lemmatized words
sheet.insert_cols(4)  # Insert a column at index 4 (column E)
sheet.cell(row=1, column=4).value = “Lemmatized Words”

# Process each row (starting from the second row, as the first is a header)
for row in range(2, sheet.max_row + 1):
    word = sheet.cell(row=row, column=1).value
    if word:  # Check for empty cells
        doc = nlp(word)
        lemmatized_words = [token.lemma_ for token in doc]
        sheet.cell(row=row, column=4).value = “, “.join(lemmatized_words)

# Save the updated Excel file
workbook.save(“lemmatized_ner.xlsx”)
```

A5 - Script to Combine the Lemmatized Words and Categories from Named Entity Recognition

```
from openpyxl import load_workbook
from collections import defaultdict

# Load the workbook
workbook = load_workbook(“/content/lemmatized_ner2.xlsx”)
sheet = workbook.active

# Create a dictionary to store combined data
combined_data = defaultdict(lambda: {“Combined Count”: 0, “NER Combined”: []})

# Iterate through rows, skipping the header
for row in sheet.iter_rows(min_row=2):
    word = row[0].value
    count = row[1].value
```

```
    ner_categories = row[2].value


    if word:
        combined_data[word][“Combined Count”] += count
        if ner_categories:
            combined_data[word][“NER Combined”].extend([cat.strip() for cat in ner_catego-
ries.split(“,”) if cat.strip()])


# Create a new sheet for combined data
new_sheet = workbook.create_sheet(“Combined Data”)
new_sheet.append([“Word”, “Combined Count”, “NER Combined”])


# Populate the new sheet with combined data
for word, data in combined_data.items():
    new_sheet.append([word, data[“Combined Count”],  “, “.join(set(data[“NER Combined”]))]) # Use
set to remove duplicate NERs


# Save the updated workbook
workbook.save(“/content/combined_lemmatized_ner2.xlsx”)
```

A6 – Topics and the Articles Where They Were Found

Term	Article
Algorithm	[PS3], [PS10], [PS15], [PS9], [PS7], [PS14], [PS06], [PS08], [PS18], [PS13], [PS16], [PS04], ,[PS12], [PS37], [PS28], [PS32], [PS22], [PS27], [PS29], [PS36], [PS24], [PS30], [PS19], [PS21], [PS31], [PS23], [PS25], [PS34], [PS33], [PS40], [PS26], [PS45], [PS54], [PS49], [PS47], [PS58], [PS44], [PS43], [PS48], [PS50], [PS46], [PS57], [PS55], [PS56], [PS59], [PS73], [PS63], [PS75], [PS64], [PS66], [PS79], [PS74], [PS77], [PS80], [PS69], [PS82], [PS71], [PS94], [PS93], [PS97], [PS83], [PS88], [PS96], [PS89], [PS84], [PS92], [PS86], [PS85], [PS108], [PS101], i[PS111], [PS99], [PS106], [PS114], [PS109], [PS107], [PS113], [PS103], [PS104], [PS116], [PS124], [PS126], [PS121], [PS128], [PS125], [PS133], [PS117], [PS120], [PS119], [PS129], [PS127], [PS123], [PS137], [PS135], [PS138], [PS134], [PS139], [PS145], [PS154], [PS149], [PS150], [PS143], [PS140], [PS147], [PS142], [PS148], [PS144], [PS152], [PS153], [PS146], [PS151], [PS167], [PS156], [PS166], [PS162], [PS170], [PS161], [PS160], [PS157], [PS163], [PS169], [PS158], [PS176], [PS186], [PS175], [PS76], [PS174], [PS188], [PS185], [PS180], [PS177], [PS184], [PS187], [PS183], [PS173], [PS171], [PS178], [PS172], [PS200], [PS189], [PS190], [PS199], [PS191], [PS197], [PS193], [PS192], [PS194], [PS201], [PS198], [PS153], [PS210], [PS213], [PS204], [PS208], [PS208], [PS211], [PS205], [PS207], [PS206], [PS81], [PS52]
Analysis	[PS3], [PS10], [PS15], [PS9], [PS7], [PS05], [PS14], [PS11], [PS06], [PS08], [PS13], [PS16], [PS04], [PS17], [PS12], [PS20], [PS28], [PS32], [PS22], [PS29], [PS39], [PS36], [PS24], [PS30], [PS19], [PS35], [PS31], [PS23], [PS25], [PS34], [PS33], [PS40], [PS26], [PS45], [PS54], [PS49], [PS47], [PS58], [PS42], [PS51], [PS44], [PS43], [PS48], [PS50], [PS46], [PS53], [PS57], [PS55], [PS56], [PS59], [PS73], [PS63], [PS72], [PS75], [PS61], [PS66], [PS79], [PS74], [PS77] ,[PS65], [PS80], [PS69], [PS78], [PS82], [PS71], [PS94], [PS93], [PS97], [PS83], [PS87], [PS88], [PS95], [PS96], [PS90], [PS84], [PS92], [PS86], [PS85], [PS98], [PS108], [PS101], [PS100], i[PS111], [PS99], [PS106], [PS105], [PS114], [PS109], [PS107], [PS115] ,[PS60], [PS110], [PS113], [PS103], [PS104], [PS112], [PS124], [PS126], [PS121], [PS128], [PS122], [PS125], [PS117], [PS120], [PS119], [PS132], [PS129], [PS127], [PS123], [PS137], [PS135], [PS138], [PS134], [PS139], [PS136], [PS145], [PS154], [PS149], [PS150], [PS143], [PS140], [PS147], [PS142], [PS141], [PS148], [PS144], [PS152], [PS153], [PS146], [PS151], [PS155], [PS167], [PS156], [PS166], [PS162], [PS170], [PS161], [PS165], [PS160], [PS164], [PS159], [PS157], [PS163], [PS169], [PS176], [PS186], [PS182], [PS175], [PS76], [PS174], [PS188], [PS185], [PS180], [PS177], [PS184], [PS181], [PS187], [PS183], [PS173], [PS179], [PS178], [PS172], [PS200], [PS189], [PS190], [PS199], [PS191], [PS196], [PS197], [PS192], [PS194], [PS201], [PS198], [PS203], [PS210], [PS213], [PS204], [PS208], [PS208], [PS211], [PS205], [PS207], [PS206], [PS81], [PS38], [PS52], [PS70]
Bayesian network	[PS17], [PS31], [PS55], [PS117], [PS120], [PS147], [PS141], [PS191]
Bert model	[PS17], [PS24], [PS31], [PS40], [PS58], [PS65], [PS69], [PS82], [PS113], [PS103], [PS153], [PS146], [PS183], [PS173], [PS153]
DeepMind	[PS3]
FL algorithm	[PS86]

Term	Article
FL model	[PS86]
Mining	[PS3], [PS10], [PS15], [PS9], [PS05], [PS14], [PS06], [PS02], [PS13], [PS04], [PS17], [PS37], [PS28], [PS32], [PS22], [PS36], [PS30], [PS19], [PS35], [PS21], [PS31], [PS23], [PS25], [PS34], [PS33], [PS26], [PS45], [PS54], [PS49], [PS47], [PS58], [PS51], [PS44], [PS43], [PS48], [PS50], [PS46], [PS53], [PS57], [PS55], [PS56], [PS59], [PS73], [PS63], [PS75], [PS61], [PS66], [PS74], [PS77] ,[PS65], [PS80], [PS69], [PS78], [PS82], [PS94], [PS93], [PS97], [PS83], [PS87], [PS95], [PS84], [PS92], [PS86], [PS85], [PS108], [PS101], [PS100], [PS106], [PS105], [PS114], [PS109], [PS107], [PS110], [PS103], [PS116], [PS112], [PS130], [PS124], [PS126], [PS121], [PS128], [PS131], [PS117], [PS119], [PS123], [PS135], [PS138], [PS134], [PS139], [PS145], [PS154], [PS149], [PS150], [PS143], [PS140], [PS147], [PS141], [PS148], [PS144], [PS152], [PS153], [PS151], [PS156], [PS166], [PS170], [PS161], [PS165], [PS160], [PS157], [PS163], [PS169], [PS158], [PS176], [PS175], [PS188], [PS185], [PS180], [PS177], [PS184], [PS181], [PS183], [PS173], [PS171], [PS179], [PS178], [PS172], [PS189], [PS01], [PS190], [PS199], [PS196], [PS197], [PS194], [PS201], [PS198], [PS203], [PS153], [PS210], [PS213], [PS204], [PS208], [PS212], [PS211], [PS205], [PS207], [PS81], [PS38], [PS52], [PS70]
Modeling	[PS3], [PS10], [PS05], [PS14], [PS06], [PS04], [PS17], [PS20], [PS37], [PS22], [PS23], [PS40], [PS54], [PS49], [PS47], [PS58], [PS42], [PS43], [PS48], [PS50], [PS46], [PS53], [PS57], [PS56], [PS59], [PS72], [PS61], [PS79], [PS74], [PS65], [PS80], [PS78], [PS82], [PS87], [PS88], [PS84], [PS86], [PS98], i[PS111], [PS106], [PS109], [PS110], [PS103], [PS104], [PS116], [PS121], [PS128], [PS131], [PS122], [PS133], [PS120], [PS119], [PS132], [PS137], [PS135], [PS138], [PS134], [PS139], [PS154], [PS150], [PS143], [PS140], [PS147], [PS142], [PS141], [PS152], [PS151], [PS155], [PS167], [PS156], [PS166], [PS162], [PS161], [PS164], [PS157], [PS168], [PS158], [PS176], [PS182], [PS175], [PS76], [PS188], [PS185], [PS180], [PS177], [PS184], [PS183], [PS173], [PS179], [PS200], [PS01], [PS190], [PS191], [PS192], [PS198], [PS153], [PS210], [PS204], [PS212], [PS207], [PS206]
adversarial network	[PS7], [PS14], [PS17], [PS58], [PS77] ,[PS71], [PS99], [PS126], [PS127], [PS147], [PS167], [PS175], [PS180], [PS187], [PS192], [PS210]
application programming interface	[PS42], [PS152]
applied intelligence	[PS63], [PS77]
applied science	[PS37], [PS63], [PS75], [PS77] ,[PS99], [PS144], [PS200], [PS204]
artificial intelligence	[PS3], [PS14], [PS17], [PS12], [PS20], [PS28], [PS22], [PS27], [PS29], [PS24], [PS30], [PS31], [PS23], [PS25], [PS40], [PS47], [PS58], [PS42], [PS59], [PS73], [PS66], [PS77] ,[PS69], [PS78], [PS71], [PS97], [PS83], [PS87], [PS96], [PS84], [PS86], i[PS111], [PS99], [PS106], [PS105], [PS114], [PS107], [PS113], [PS126], [PS131], [PS117], [PS119], [PS129], [PS127], [PS135], [PS138], [PS134], [PS149], [PS150], [PS143], [PS140], [PS147], [PS141], [PS148], [PS153], [PS146], [PS151], [PS155], [PS167], [PS166], [PS164], [PS157], [PS163], [PS169], [PS175], [PS76], [PS174], [PS188], [PS187], [PS183], [PS173], [PS199], [PS193], [PS192], [PS194], [PS210], [PS213], [PS204], [PS208], [PS211]
artificial intelligence in medicine	[PS14], [PS17], [PS28], [PS23], [PS77] ,[PS113], [PS147], [PS153], [PS157], [PS163], [PS175], [PS183], [PS213], [PS211]
artificial neural network	[PS10], [PS17], [PS37], [PS25], [PS77] ,[PS126], [PS149], [PS150], [PS144], [PS81]
auxiliary classifier	[PS147], [PS187], [PS192]
based algorithm	[PS10], [PS22], [PS31], [PS34], [PS50], [PS55], [PS93], [PS99], [PS121], [PS138], [PS147]
based answering system	[PS152]
based deep learn	[PS99], [PS150], [PS166], [PS160]
based dialogue system	[PS53], [PS152]
based lstm architecture	[PS34]
benchmark database	[PS152]
biodata	[PS190]
bioscience	[PS29], [PS89], [PS153], [PS169], [PS76], [PS180], [PS189]
biotechnology	[PS39], [PS72], [PS101], [PS153]
cancer dataset	[PS101], [PS116], [PS149], [PS147]
classification algorithm	[PS19], [PS44], [PS43], [PS56], [PS92], [PS140], [PS157], [PS187], [PS213]
classification dataset	[PS14], [PS149]
clinical medical and health system	[PS152]
coding model	[PS17]
computerization	[PS92], [PS147], [PS193]
contact information	[PS17], [PS58], [PS187]
conversational system	[PS152]
convolutional neural network	[PS7], [PS14], [PS18], [PS17], [PS32], [PS23], [PS34], [PS45], [PS54], [PS43], [PS50], [PS46], [PS63], [PS77] ,[PS83], [PS84], [PS86], [PS99], [PS126], [PS131], [PS127], [PS134], [PS149], [PS150], [PS143], [PS147],

Term	Article
	[PS142], [PS152], [PS153], [PS160], [PS164], [PS163], [PS175], [PS76], [PS185], [PS180], [PS187], [PS183], [PS192], [PS210]
cryptosystem	[PS86], [PS160]
cybersecurity	[PS02], [PS36], [PS114], [PS169], [PS208]
data augmentation	[PS7], [PS17], [PS23], [PS46], [PS77] ,[PS65], [PS71], [PS88], [PS84], [PS103], [PS126], [PS129], [PS149], [PS147], [PS160], [PS175], [PS185], [PS177],[PS187], [PS173], [PS192], [PS210]
data collection	[PS3], [PS10], [PS02], [PS12], [PS20], [PS22], [PS30], [PS19], [PS25], [PS33], [PS49], [PS58], [PS42], [PS51], [PS44], [PS73], [PS63], [PS77] ,[PS82], [PS97], [PS86], i[PS111], [PS105], [PS114], [PS115] ,[PS103], [PS120], [PS123], [PS149], [PS147], [PS152], [PS155], [PS170], [PS161], [PS159], [PS158], [PS176], [PS187], [PS183], [PS200], [PS190], [PS191], [PS196], [PS192], [PS194], [PS213], [PS208], [PS211], [PS70]
data engineer	[PS27], [PS19], [PS33], [PS26], [PS47], [PS48], [PS109], [PS104], [PS119], [PS138], [PS150], [PS143], [PS147], [PS185], [PS187], [PS178], [PS210]
data generation	[PS58], [PS77] ,[PS147], [PS167], [PS187], [PS192]
databank	[PS189]
database	[PS3], [PS10], [PS15], [PS9], [PS14], [PS06], [PS13], [PS16], [PS04], [PS17], [PS12], [PS37], [PS28], [PS32], [PS22], [PS27], [PS39], [PS36], [PS24], [PS30], [PS21], [PS23], [PS25], [PS33], [PS40], [PS26], [PS54], [PS49], [PS47], [PS58], [PS42], [PS51], [PS44], [PS43], [PS48], [PS46], [PS57], [PS55], [PS56], [PS73], [PS75], [PS61], [PS66], [PS77] ,[PS65], [PS80], [PS69], [PS67], [PS78], [PS82], [PS97], [PS87], [PS95], [PS89], [PS84], [PS86], [PS85], [PS108], [PS101], [PS100], i[PS111], [PS99], [PS106], [PS114], [PS109], [PS104], [PS116], [PS130], [PS124], [PS126], [PS121], [PS128], [PS131], [PS133], [PS117], [PS120], [PS119], [PS129], [PS127], [PS123], [PS135], [PS138], [PS134], [PS139], [PS136], [PS145], [PS149], [PS150], [PS140], [PS147], [PS141], [PS148], [PS152], [PS153], [PS151], [PS167], [PS156], [PS166], [PS165], [PS160], [PS159], [PS157], [PS163], [PS169], [PS158], [PS176], [PS182], [PS175], [PS174], [PS188], [PS185], [PS180], [PS177], [PS184], [PS187], [PS183], [PS171], [PS172], [PS200], [PS189], [PS190], [PS191], [PS196], [PS192], [PS153], [PS213], [PS204], [PS208], [PS208], [PS207], [PS209], [PS81], [PS38], [PS70]
databiox	[PS77]
deep convolutional neural network	[PS14], [PS77] ,[PS126], [PS149], [PS163], [PS187], [PS210]
deep ehr	[PS65]
deep learning model	[PS7], [PS18], [PS17], [PS31], [PS23], [PS34], [PS45], [PS54], [PS58], [PS50], [PS46], [PS63], [PS77] ,[PS65], [PS69], [PS82], [PS71], [PS97], [PS84], [PS86], [PS116], [PS126], [PS121], [PS135], [PS134], [PS149], [PS150], [PS147], [PS166], [PS160], [PS164], [PS157], [PS176], [PS175], [PS76], [PS180], [PS187], [PS183], [PS199], [PS192], [PS198], [PS213], [PS208]
deep neural network	[PS7], [PS14], [PS18], [PS17], [PS45], [PS42], [PS50], [PS53], [PS77] ,[PS71], [PS83], [PS86], [PS99], [PS107], [PS116], [PS126], [PS127], [PS135], [PS149], [PS150], [PS147], [PS142], [PS144], [PS152], [PS146], [PS155], [PS167], [PS170], [PS160], [PS76], [PS180], [PS183], [PS192]
deepr	[PS143], [PS183]
dialogue act model	[PS152]
diffusion model	[PS61], [PS77] ,[PS147], [PS187], [PS210], [PS52]
digitalization	[PS35], [PS59], [PS63], [PS128], [PS152], [PS38]
disambiguation	[PS32], [PS33], [PS45], [PS73], [PS79], [PS145], [PS152], [PS151], [PS169], [PS184]
document object model	[PS152]
electronic health record	[PS3], [PS10], [PS7], [PS14], [PS04], [PS17], [PS32], [PS35], [PS21], [PS40], [PS45], [PS58], [PS42], [PS51], [PS43], [PS48], [PS57], [PS56], [PS72], [PS75], [PS61], [PS66], [PS74], [PS65], [PS69], [PS78], [PS82], [PS93], [PS95], [PS84], [PS86], [PS85], [PS100], [PS114], [PS109], [PS115] ,[PS104], [PS131], [PS119], [PS123], [PS135], [PS138], [PS134], [PS136], [PS143], [PS147], [PS142], [PS152], [PS153], [PS166], [PS165], [PS159], [PS157], [PS168], [PS158], [PS176], [PS174], [PS188], [PS185], [PS180], [PS177], [PS183], [PS173], [PS178], [PS191], [PS197], [PS198], [PS204], [PS81]
enriched deep learning	[PS166], [PS208]
federated machine learning	[PS86]
field digital mammographic database	[PS14], [PS77]
full model	[PS144], [PS166], [PS191]
general data protection regulation	[PS36], [PS77] ,[PS86], [PS147], [PS152], [PS169], [PS200]
generative adversarial network	[PS7], [PS14], [PS17], [PS58], [PS77] ,[PS71], [PS99], [PS126], [PS127], [PS147], [PS167], [PS175], [PS180], [PS187], [PS192], [PS210]
generative model	[PS7], [PS17], [PS45], [PS58], [PS77] ,[PS86], [PS126], [PS147], [PS151], [PS167], [PS187], [PS192], [PS210]
graph neural network	[PS14], [PS17], [PS24], [PS84], [PS104], [PS131], [PS141], [PS144], [PS157]

Term	Article
health information	[PS10], [PS06], [PS02], [PS04], [PS29], [PS35], [PS21], [PS45], [PS49], [PS58], [PS42], [PS51], [PS53], [PS55], [PS56], [PS68], [PS61], [PS66], [PS79], [PS80], [PS78], [PS82], [PS87], [PS90], [PS85], [PS100], [PS106], [PS103], [PS112], [PS130], [PS123], [PS135], [PS139], [PS136], [PS147], [PS142], [PS144], [PS152], [PS151], [PS166], [PS160], [PS159], [PS168], [PS181], [PS183], [PS171], [PS179], [PS190], [PS210], [PS208], [PS212], [PS38], [PS70]
healthcare data	[PS10], [PS04], [PS27], [PS34], [PS42], [PS51], [PS63], [PS75], [PS66], [PS87], [PS86], [PS85], [PS101], [PS100], [PS99], [PS104], [PS135], [PS147], [PS142], [PS167], [PS160], [PS176], [PS182], [PS185], [PS178], [PS200], [PS210]
iid data	[PS86]
iii dataset	[PS17], [PS32], [PS27], [PS58], [PS48], [PS65], [PS84], [PS131], [PS150], [PS157], [PS176], [PS185], [PS177], [PS183]
in computer	[PS14], [PS18], [PS23], [PS26], [PS49], [PS46], [PS53], [PS77] ,[PS67], [PS71], [PS97], [PS85], [PS114], [PS109], [PS126], [PS120], [PS132], [PS137], [PS145], [PS154], [PS147], [PS141], [PS152], [PS162], [PS175], [PS187], [PS183], [PS179], [PS191], [PS192], [PS204], [PS208]
incomplete datasets using generative model	[PS86], [PS187]
initiative dialogue system	[PS152]
intelligence in medicine	[PS10], [PS14], [PS04], [PS17], [PS28], [PS19], [PS23], [PS77] ,[PS101], [PS113], [PS138], [PS147], [PS153], [PS162], [PS157], [PS163], [PS175], [PS183], [PS213], [PS211]
intelligent system	[PS12], [PS28], [PS27], [PS29], [PS31], [PS26], [PS63], [PS72], [PS69], [PS78], [PS97], [PS98], [PS101], [PS100], i[PS111], [PS106], [PS109], [PS143], [PS172], [PS200], [PS204], [PS208]
interactive conversational system	[PS152]
internet technology	[PS86], [PS149], [PS160]
interpretability of machine intelligence	[PS14], [PS77]
interpretable medical image diagnosis network	[PS14], [PS77]
label embedding initialisation	[PS17]
labelled data	[PS45], [PS77] ,[PS97], [PS126], [PS118], [PS149], [PS144], [PS155], [PS187], [PS191]
language model	[PS7], [PS17], [PS20], [PS22], [PS27], [PS31], [PS40], [PS45], [PS58], [PS42], [PS50], [PS53], [PS55], [PS68], [PS66], [PS77] ,[PS65], [PS69], [PS78], [PS82], [PS96], [PS113], [PS118], [PS129], [PS135], [PS134], [PS142], [PS141], [PS152], [PS153], [PS146], [PS151], [PS167], [PS170], [PS165], [PS157], [PS174], [PS188], [PS183], [PS173], [PS153], [PS204]
loop learning	[PS17]
loss prediction model	[PS45]
lstm architecture	[PS14], [PS34], [PS143], [PS166], [PS160]
machine learning	[PS3], [PS10], [PS7], [PS14], [PS18], [PS04], [PS17], [PS12], [PS37], [PS32], [PS22], [PS29], [PS36], [PS24], [PS30], [PS19], [PS21], [PS23], [PS25], [PS34], [PS45], [PS54], [PS49], [PS58], [PS42], [PS44], [PS48], [PS50], [PS46], [PS53], [PS57], [PS55], [PS56], [PS59], [PS73], [PS63], [PS66], [PS79], [PS77] ,[PS65], [PS69], [PS82], [PS71], [PS97], [PS83], [PS87], [PS88], [PS89], [PS84], [PS92], [PS86], i[PS111], [PS99], [PS106], [PS114], [PS109], [PS107], [PS113], [PS103], [PS116], [PS126], [PS121], [PS117], [PS120], [PS119], [PS129], [PS135], [PS138], [PS134], [PS139], [PS149], [PS150], [PS143], [PS140], [PS147], [PS142], [PS141], [PS148], [PS144], [PS152], [PS153], [PS151], [PS155], [PS167], [PS166], [PS165], [PS160], [PS164], [PS157], [PS163], [PS169], [PS158], [PS176], [PS175], [PS76], [PS185], [PS180], [PS177], [PS184], [PS187], [PS183], [PS173], [PS179], [PS172], [PS190], [PS199], [PS191], [PS193], [PS192], [PS194], [PS198], [PS210], [PS213], [PS204], [PS208], [PS211], [PS205], [PS207], [PS70]
machine translation	[PS7], [PS14], [PS17], [PS45], [PS58], [PS50], [PS53], [PS77] ,[PS65], [PS116], [PS117], [PS129], [PS134], [PS147], [PS141], [PS152], [PS146], [PS151], [PS176], [PS175], [PS188], [PS180], [PS183], [PS173], [PS153], [PS204]
medical and health system	[PS152]
medical image analysis	[PS14], [PS28], [PS23], [PS46], [PS77] ,[PS69], [PS97], [PS99], [PS126], [PS127], [PS134], [PS167], [PS160], [PS184], [PS199], [PS192], [PS210], [PS213]
medical image computing	[PS14], [PS77] ,[PS139], [PS187], [PS192]
medical information retrieval	[PS05], [PS31], [PS55], [PS130], [PS165], [PS183]
metadata	[PS15], [PS9], [PS17], [PS26], [PS42], [PS73], [PS77] ,[PS87], [PS95], [PS103], [PS116], [PS133], [PS120], [PS145], [PS149], [PS152], [PS182], [PS184], [PS183], [PS171], [PS191], [PS70]
metamodel	[PS95]
microdata	[PS152]

Term	Article
misinformation	[PS42], [PS55], [PS96], [PS177], [PS208], [PS212]
modal machine learning approach	[PS14], [PS77]
mri dataset	[PS149]
multimodal architecture	[PS152]
multimodal grammar model	[PS152]
multimodal learning	[PS14], [PS18], [PS17], [PS77] ,[PS144], [PS152], [PS194]
nanotechnology	i[PS111]
neural machine translation	[PS7], [PS45], [PS53], [PS65], [PS116], [PS129], [PS134], [PS147], [PS141], [PS152], [PS146], [PS151], [PS180], [PS183], [PS153]
neural model	[PS17], [PS143], [PS183]
neural network	[PS3], [PS10], [PS7], [PS14], [PS18], [PS17], [PS37], [PS32], [PS27], [PS24], [PS30], [PS19], [PS31], [PS23], [PS25], [PS34], [PS40], [PS45], [PS54], [PS58], [PS42], [PS44], [PS43], [PS48], [PS50], [PS46], [PS53], [PS55], [PS59], [PS63], [PS77] ,[PS65], [PS69], [PS82], [PS71], [PS83], [PS88], [PS84], [PS86], [PS85], [PS99], [PS109], [PS107], [PS104], [PS116], [PS126], [PS121], [PS131], [PS117], [PS119], [PS127], [PS135], [PS138], [PS134], [PS139], [PS149], [PS150], [PS143], [PS147], [PS142], [PS141], [PS144], [PS152], [PS153], [PS146], [PS155], [PS167], [PS156], [PS166], [PS170], [PS160], [PS164], [PS157], [PS163], [PS176], [PS175], [PS76], [PS185], [PS180], [PS177], [PS187], [PS183], [PS197], [PS192], [PS198], [PS153], [PS210], [PS204], [PS208], [PS211], [PS205], [PS81], [PS52]
noisy data	[PS7], [PS84], [PS99], [PS152], [PS170], [PS191], [PS208], [PS212], [PS207]
open agent architecture	[PS152]
patient data	[PS3], [PS10], [PS9], [PS06], [PS04], [PS20], [PS37], [PS32], [PS36], [PS51], [PS48], [PS57], [PS61], [PS66], [PS82], [PS87], [PS86], [PS100], i[PS111], [PS114], [PS117], [PS119], [PS135], [PS147], [PS152], [PS153], [PS182], [PS185], [PS177], [PS184], [PS183], [PS171], [PS200], [PS197], [PS210], [PS204], [PS208], [PS211], [PS38]
recurrent neural network	[PS7], [PS14], [PS17], [PS32], [PS27], [PS34], [PS40], [PS54], [PS58], [PS48], [PS50], [PS59], [PS63], [PS77] ,[PS82], [PS88], [PS84], [PS86], [PS99], [PS116], [PS126], [PS131], [PS135], [PS138], [PS134], [PS150], [PS143], [PS147], [PS142], [PS141], [PS144], [PS152], [PS146], [PS166], [PS160], [PS157], [PS176], [PS185], [PS180], [PS177], [PS198], [PS153], [PS204]
robotic application	[PS152]
scale dataset	[PS04], [PS40], [PS77] ,[PS129], [PS138], [PS149], [PS187]
scientific data	[PS14], [PS58], [PS43], [PS46], [PS77] ,[PS84], [PS153], [PS167], [PS157], [PS76], [PS183], [PS153]
senior care information	[PS212]
sensitivity analysis	[PS63], [PS166]
sentiment analysis	[PS22], [PS34], [PS50], [PS63], [PS92], [PS106], [PS153], [PS164], [PS175], [PS204]
sequential database	[PS138]
shot learning	[PS17], [PS68], [PS77],[PS126], [PS134], [PS144], [PS187]
speech mashup architecture	[PS152]
spoken dialogue system	[PS152], [PS188]
statistical language model	[PS152]
support vector machine	[PS3], [PS13], [PS17], [PS32], [PS30], [PS19], [PS21], [PS34], [PS45], [PS54], [PS44], [PS50], [PS55], [PS84], [PS86], [PS99], [PS107], [PS116], [PS126], [PS139], [PS149], [PS147], [PS152], [PS155], [PS164], [PS180], [PS184], [PS172], [PS153], [PS204], [PS208], [PS205]
survival analysis	[PS88], [PS101]
synthesis algorithm	[PS147], [PS162]
synthetic and real data	[PS88], [PS147], [PS187]
synthetic dataset	[PS58], [PS136], [PS147], [PS144], [PS167], [PS187]
thetic data	[PS7], [PS58], [PS43], [PS63], [PS77] ,[PS84], [PS107], [PS136], [PS147], [PS144], [PS167], [PS187], [PS192]
twd model	[PS63]
unified medical language system	[PS17], [PS31], [PS77] ,[PS82], [PS151], [PS157], [PS183], [PS212]
unsupervised representation learning	[PS142], [PS187], [PS192], [PS198]

Term	Article
variate analysis	[PS138], [PS181]
vertically split data	[PS86]
virtual digital assistant	[PS152]
web ontology language	[PS17], [PS31], [PS152]

A7 - Deep Learning Terms - Dates and Percentage Present

Term	First article	Last article	% of articles
neural network	1973	2024	43.66%
deep learning model	2013	2024	20.19%
support vector machine	2018	2018	15.02%
generative adversarial network	2021	2021	7.51%
neural machine translation	2008	2024	7.04%
deep convolutional neural network	1973	2024	4.69%
LSTM architecture	1971	2024	2.35%
neural model	1971	2024	1.41%
DeepMind	1971	2024	0.47%

A8 - Data Organization Terms - Dates and Percentage Present

Term	First article	Last article	% of articles
data collection	2006	2024	62.91%
Database	1994	2024	62.91%
healthcare data	2015	2024	23%
data augmentation	2021	2024	12.68%
data engineer	2006	2024	10.33%
MIMIC-III dataset	2017	2024	7.98%
labelled data	2012	2023	6.57%
Bioscience	2010	2023	4.69%
data generation	2010	2024	3.29%
cancer dataset	2017	2024	2.82%
incomplete datasets using generative model	2021	2024	1.88%
classification dataset	2022	2022	0.94%
field digital mammographic database	2022	2022	0.94

A9 - Medical or Healthcare Applications Terms - Dates and Percentage Present

Term	First article	Last article	% of articles
electronic health record	1991	2024	32.39%
patient data	1973	2024	18.31%
healthcare data	1999	2024	12.68%

A10 - Techniques Used Terms - Dates and Percentage Present

Term	First article	Last article	% of articles
language model	2016	2024	9.26%
neural machine translation	2018	2024	3.31%
virtual digital assistant	2019	2019	0.22%
speech mashup architecture	2019	2019	0.22%

A11 - Data and Computing Infrastructure Terms - Dates and Percentage Present

Term	First article	Last article	% of articles
applied science	2012	2024	37.56%
applied intelligence	2023	2023	3.76
federated machine learning	2021	2021	3.29
computerization	2019	2024	2.82
artificial intelligence	2010	2024	1.88
Biotechnology	2017	2023	1.41%
general data protection regulation	2017	2024	1.41%
robotic application	2019	2019	1.41%
web ontology language	2019	2024	1.41%
application programming interface	2015	2015	0.94%
open agent architecture	2015	2015	0.47%
nano technology	2023	2023	0.47%
internet technology	2018	2022	0.47%
document object model	2015	2015	0.47%
digitization	2002	2024	0.47%

A12 - Language Systems Terms - Dates and Percentage Present

Term	First article	Last article	% of articles
Transparency	2012	2012	42%
Responsibility	1971	2024	15%

Term	First article	Last article	% of articles
Accountability	2002	2024	1%
Fairness	2002	2024	1%

A13 - THE MEANINGS OF THE WORDS WITH RESPECT TO THE APPLIED DATA SCIENCE IN HEALTHCARE

The definition of the words was determined based on their usage in the literature or their relevance to the literature.

Term	Definition
Adversarial Network	A machine learning model where two neural networks compete against each other to improve their performance.
Algorithm	A set of rules or instructions that a computer follows to solve a problem or perform a task.
Analysis	The process of examining data to extract insights and draw conclusions.
Application Programming Interface (API)	A set of rules and protocols for building and integrating software applications.
Applied Intelligence	The practical application of artificial intelligence to solve real-world problems.
Applied Science	The application of scientific knowledge to practical problems.
Artificial Intelligence (AI)	The simulation of human intelligence processes by machines, especially computer systems.
Artificial Intelligence in Medicine	The application of AI techniques to medical diagnosis, treatment, and research.
Artificial Neural Network (ANN)	A computational model inspired by the structure and function of the human brain.
Based Algorithm	An algorithm that is built upon or derived from another algorithm.
Based Answering System	A system that uses AI to provide answers to user queries.
Based Deep Learn	A system that uses deep learning techniques to learn from data.
Based Dialogue System	A system that uses AI to engage in natural language conversations with users.
Based LSTM Architecture	A system that uses Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, to process sequential data.
Bayesian Network	A probabilistic graphical model that represents a set of random variables and their conditional dependencies. 1
BD Cloud	A cloud-based platform for storing and analyzing big data.
Benchmark Database	A database used to evaluate the performance of algorithms and systems.
BERT Model	A pre-trained language model developed by Google AI.
Biodata	Biological data, such as genetic information or medical records.
Bioscience	The science of life and living organisms.
Biotechnology	The application of biological processes to industrial and technological purposes.

Term	Definition
Cancer Dataset	A dataset containing information about cancer patients, including their medical history, treatments, and outcomes.
Classification Algorithm	An algorithm that assigns class labels to data points.
Classification Dataset	A dataset with predefined classes or categories.
Clinical Medical and Health System	A system that provides healthcare services to patients.
Coding Model	A model that automatically assigns codes to medical records.
Computerization	The process of introducing computer technology into a system or organization.
Concepts Dataset	A dataset containing information about concepts and their relationships.
Contact Information	Information used to contact someone, such as a phone number or email address.
Conversational System	A system that can engage in natural language conversations with users.
Convolutional Neural Network (CNN)	A type of neural network that is particularly well-suited for image and video analysis.
CRF Models Sequence Tag	Conditional Random Fields (CRFs) are a type of probabilistic graphical model used for labeling or segmenting sequential data.
Cryptosystem	A system for encrypting and decrypting data.
Cybersecurity	The practice of protecting computer systems and networks from digital attacks.
Data Augmentation	The process of artificially increasing the size of a dataset by creating new data points from existing ones.
Auxiliary Classifier	A classifier that is used to assist in the training of another classifier.
Data Collection	The process of gathering data.
Data Engineer	A professional who specializes in designing, building, and maintaining data systems.
Data Generation	The process of creating new data.
Data Metaprovenance	Information about the origin and history of data.
Databank	A collection of data.
Database	A structured collection of data.
Databiox	A platform for storing and analyzing biological data.
Dataserialization Format	A format for storing data in a structured way.
Deep EHR	A deep learning model for analyzing electronic health records.
Deep FFNN	A deep feedforward neural network. In this network, the information moves in only one direction forward from the input nodes, through the hidden nodes (if any), and to the output nodes
Deep Learning Model	A type of machine learning model that uses multiple layers of neural networks to learn complex patterns.
Deep Neural Network (DNN)	A neural network with multiple hidden layers.
DeepMind	An AI research company.
DeepR	A deep learning framework for recommender systems.
Dialogue Act Model	A model that predicts the intent of a user's utterance in a dialogue system.

Term	Definition
Diffusion Model	A generative model that generates data by iteratively refining noise.
Digitalization	The process of converting information into a digital format.
Disambiguation	The process of resolving ambiguity in language.
Document Object Model (DOM)	A programming interface for HTML and XML documents.
Electronic Health Record (EHR)	A digital version of a patient's medical record.
Enriched Deep Learning	Deep learning models that are enhanced with additional information.
Federated Machine Learning	A machine learning technique that allows multiple organizations to collaborate on training a model without sharing their data.
Field Digital Mammographic Data-base (DDSM)	A public database of mammograms.
FL Algorithm	Federated Learning algorithm.
FL Model	Federated Learning model.
Full Model	A model that includes all relevant features and variables.
General Data Protection Regulation (GDPR)	A European Union law on data protection and privacy.
Generative Adversarial Network (GAN)	A type of neural network that generates new data.
Generative Model	A model that generates new data.
Gig Science	A scientific approach that leverages large-scale data and computing resources.
Graph Neural Network (GNN)	A type of neural network that can process graph-structured data.
Health Information	Information about health and healthcare.
Healthcare Data	Data related to healthcare, such as patient records and medical images.
IID Data	Independent and identically distributed data.
III Dataset	ImageCLEF medical image dataset.
ImageCLEF Dataset	A dataset used in the ImageCLEF medical image analysis challenge.
Deep Convolutional Neural Network	A type of CNN with multiple layers.
ImageNet Classification	A benchmark dataset for image classification.
Image Reversible Data Hide	A technique for hiding data within an image.
Implications Digital Health	The potential impact of digital technologies on healthcare.
In Computer	Within a computer system.
Incomplete Datasets Using Generative Model	The use of generative models to fill in missing data.
Initiative Dialogue System	A dialogue system that is designed to take the initiative in the conversation.
Intelligence in Medicine	The application of intelligence to medical problems.
Intelligent System	A system that exhibits intelligent behavior.
Interactive Conversational System	A conversational system that allows users to interact with it in a natural way.
Internet Technology	Technology related to the internet.

Term	Definition
Interpretability of Machine Intelligence	The ability to understand how a machine learning model makes decisions.
Medical Image Computing	The analysis and interpretation of medical images.
Interpretable Medical Image Diagnosis Network	A neural network that can explain its decisions.
Multimodal Learning	The ability of a machine learning model to process multiple types of data, such as text, images, and audio.
Label Embedding Initialization	The process of initializing the embeddings for class labels in a machine learning model.
Labeled Data	Data that has been labeled with the correct answer.
Language Model	A model that can generate or understand human language.
Loop Learning	A type of machine learning where the model learns from its own mistakes.
Loss Prediction Model	A model that predicts the loss function of a machine learning model.
LSTM Architecture	Long Short-Term Memory architecture, a type of recurrent neural network.
Machine Learning	A type of artificial intelligence that allows computers to learn from data without being explicitly programmed. 2
Machine Translation	Translation of text from one language to another using machine learning.
Medical and Health System	A system that provides healthcare services.
Medical Image Analysis	The analysis and interpretation of medical images.
Medical Information Retrieval	The retrieval of medical information from databases.
Metadata	Data about data.
Metamodel	A model of a model.
Microdata	Detailed data about individuals or households.
Mining	The process of extracting valuable information from data.
Misinformation	False or misleading information.
Modal Machine Learning Approach	A machine learning approach that uses multiple models.
Modeling	The process of creating a model.
MRI Dataset	A dataset of magnetic resonance imaging (MRI) scans.
Multi Database	A database that stores data from multiple sources.
Multimodal Architecture	An architecture that can process multiple types of data.
Multimodal Grammar Model	A grammar model that can represent multiple modalities of language.
Nano Bioscience	The application of nanotechnology to biological systems.
Nanotechnology	The manipulation of matter on an atomic and molecular scale.
Neural Machine Translation	A machine translation technique that uses neural networks.
Neural Model	A model based on artificial neural networks.
Neural Network	A computational model inspired by the structure and function of the human brain.
Noisy Data	Data that contains errors or inaccuracies.

Term	Definition
Open Agent Architecture	Architecture for building intelligent agents that can interact with the environment.
Parenetic	Giving advice or instruction.
Patient Data	Data related to patients, such as medical records and demographics.
Pre-trained Model	A model that has been trained on a large dataset and can be fine-tuned for specific tasks.
Public Classification Dataset	A publicly available dataset that has been labeled with class labels.
Ray Dataset	A dataset that is optimized for use with the Ray distributed computing framework.
Recurrent Neural Network (RNN)	A type of neural network that is well-suited for processing sequential data.
Reversible Data Hiding	A technique for hiding data within a cover medium in a reversible way.
Robotic Application	The application of robotics to a specific task or problem.
Scale Dataset	A large dataset.
Scientific Data	Data collected through scientific research.
Senior Care Information	Information related to the care of elderly people.
Sensitivity Analysis	A technique for assessing the impact of uncertainty in model inputs on the model's outputs.
Sentiment Analysis	The process of determining the sentiment of text, such as positive, negative, or neutral.
Sequential Database	A database that stores data in a sequence.
Shot Learning	A machine learning technique that can learn from a small number of examples.
Speech Mashup Architecture	An architecture for combining multiple speech processing components.
Spoken Dialogue System	A system that allows users to interact with a computer using spoken language.
Statistical Language Model	A probabilistic model of language.
Support Vector Machine (SVM)	A supervised machine learning model that can be used for classification and regression.
Survival Analysis	A statistical method for analyzing time-to-event data.
Synthesis Algorithm	An algorithm that generates new data.
Synthetic and Real Data	A combination of synthetic and real data.
Synthetic Dataset	A dataset that is generated artificially.
Thetic Data	Data that is used to train a machine learning model.
TWD Model	A time-weighted decay model.
Unified Medical Language System (UMLS)	A knowledge source that integrates and indexes biomedical terminologies.
Unsupervised Representation Learning	A machine learning technique that learns representations of data without using labeled data.
Variate Analysis	A statistical technique for analyzing the variability of data.
Vertically Split Data	Data that is split into columns.

Term	Definition
Virtual Digital Assistant	A software agent that can perform tasks on behalf of a user.
Web Ontology Language (OWL)	A language for representing knowledge in a structured way.

A14 - VISUALIZATION TOOL USED FOR THE THEMATIC BUBBLE CHART

Flourish is an online interactive data visualization tool. It helps transform data and tables into beautiful visualized graphical stories without programming knowledge. Once the data is uploaded, a graph or a chart can be edited in many ways. For the current paper, the thematically analyzed data was uploaded and transformed into a bubble chart. This link leads to the interactive bubble chart used and mentioned in the results section: https://public.flourish.studio/visualisation/20499227/

AUTHORS

Anshuman Rangaraj is pursuing his PhD in Information Technology at the University of Cincinnati. Holding master's degrees in both information technology and geographic information systems, his research currently focuses on adapting AI literacy frameworks in informal settings. He has experience in various organizations, such as gaming, AI, and data visualization.

Theodore Langdon is pursuing his PhD in Information Technology at the University of Cincinnati. Holding master's degrees in both information technology and criminal justice, his research focuses on the use of natural language processing to improve responses to emergencies by telecommunicators. These methods are bolstered by over 25 years of experience in the field of 9-1-1 operations.

Dr Hazem Said is a Professor of Information Technology and the Director of the School of Information Technology (SoIT) at the University of Cincinnati (UC). He is a certified Project Management Professional (PMP). Dr Said founded the UC Information Technology Solutions Center (ITSC) in 2012, where he consults with government, public, and private organizations and leads teams of professionals, as well as graduate, undergraduate, and high school students, to investigate, develop, and support a variety of information technology solutions. In addition, Dr Said is a co-founder and co-director of the Ohio Cyber Range Institute and the Justice, Law, and Information Technology Institute. Dr Said is the recipient of over 200 grants and contracts totaling over $30 million and has authored over 30 articles on topics related to information technology education.