MEDINFO 2017: Precision Healthcare through Informatics
A.V. Gundlapalli et al. (Eds.)
© 2017 International Medical Informatics Association (IMIA) and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/978-1-61499-830-3-968

Content and Trends in Medical Informatics Publications over the Past Two Decades

Liqin Wang[a,b], Maxim Topaz[a], Joseph M. Plasek[c], Li Zhou[a,b,d]

[a] *Division of General Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, USA,*
[b] *Harvard Medical School, Boston, MA, USA*
[c] *Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, USA*
[d] *Clinical Informatics, Partners eCare, Partners HealthCare System, Boston, MA, USA*

Abstract

This study aims to identify subject content and trends in the medical informatics literature in order to shed light on the past, current, and future directions of diverse education and research activities. A list of 36 core medical informatics journals was compiled through expert consensus. We retrieved 60,862 articles from the U.S. National Library of Medicine's MEDLINE database that were published by those journals from 1992 to2015. A series of descriptive analyses were conducted to reveal the historical productivity of the journals, publications trends, and the subject content based on the Medical Subject Headings (MeSH) term frequencies and debut years. We found 73 core topics and 72 new topics of medical informatics within three relevant MeSH categories (informatics, techniques, and healthcare).

Keywords:

Medical Informatics; Publications; Medical Subject Headings

Introduction

Medical informatics (MI) is defined by the United States (U.S.) National Library of Medicine (NLM) as "the interdisciplinary study of the design, development, adoption, and application of IT-based innovations in healthcare services delivery, management, and planning" [1]. It emerged in the 1950s and has been an interdisciplinary domain for over half a century, involving applications of computer science, information science, engineering, social science, management science, among many others, in all fields of health and medicine [2]. Clinical informatics is now a medical subspecialty that allows physicians to practice clinical informatics in collaboration with other healthcare and information technology professionals to promote safe, efficient, and effective patient-centered care [3]. Today, over 80 U.S.-based academic institutions offer onsite or distance MI training programs [4].

Most published review articles in the MI domain have focused on specific subdomains or research topics, such as clinical decision support, natural language processing, and terminologies. Only a few prior studies have attempted to characterize the overall research activities in the MI literature [5]. To summarize the literature trends, these studies used Medical Subject Headings (MeSH), a comprehensive controlled vocabulary for indexing journal articles and books in the life sciences developed by NLM [6]. MeSH terms are assigned manually by professional indexers and the process can be assisted with automated approaches [7]. DeShazo et al. took the first step to use MeSH terms to analyze publication trends in the MI literature published between 1987 and 2006 [5]. Their study included all articles assigned any term in the MeSH "Medical Informatics" hierarchy as a major topic across all types of journals indexed by MEDLINE/PubMed. To our best knowledge, no literature trend analysis across the overall MI field in the recent decade has been published.

The focus of this paper is to describe the evolution of the field of MI in the recent two decades by conducting a trend analysis on research topics published in MI journals using MeSH descriptor terms. The aim of this study is to update the MI community about the most recent research trends from published MI literature. The results of this study will provide important insights for future educational, research and development directions. For example, our findings may be used to aid in the curriculum development for the MI training programs, core content updates for the MI subspecialty exam [8], as well as suggesting potential research funding areas.

Methods

We utilized a multi-step approach to identify informatics publications and research topic trends: (1) identify core MI journals, (2) form a query to retrieve articles from the MEDLINE database, (3) extract MeSH terms from retrieved articles, (4) select MeSH term categories, and (5) perform descriptive analysis to identify trends (see Figure 1).

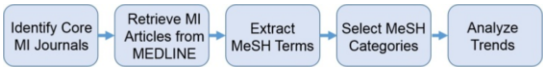

Figure 1– Methods Overview

Step 1: Identifying MI journals

In the MEDLINE database, there are approximately 26,000 journal records [9]. To identify MI publications, DeShazo searched MEDLINE for all documents assigned the term "Medical Informatics" in MeSH or any term in the "Medical Informatics" hierarchy as a Major Topic [5]. However, we found that this approach produces somewhat noisy results with a significant amount of non-relevant articles. For example, "databases, factual" is a descendant term in the MeSH Medical Informatics hierarchy. When one uses this term as a MeSH Major Topic to search MEDLINE, more than 29,000 articles are retrieved. However, many studies such as clinical trials, epidemiological studies, or economic analyses that used some sorts of databases may not necessarily be classified as "true" MI articles. Additionally, we also reviewed the 1200+ articles retrieved from the journal of "BMC Genomics" based on DeShazo's query, the majority of the articles were not about MI studies per se, as this journal mainly focuses on genome-scale analysis, functional genomics, and proteomics. To account for this, we utilized a different approach by identifying MI "core" journals first. We assume that the majority of MI topics are published in the "core" MI journals and topic trends in MI journals can serve as a proxy for the MI topic trends across journals in other fields.

In this study, we specifically include journals and articles related to clinical informatics, public health informatics, and nursing informatics that directly associate with patient care and population health. Although the broad biomedical informatics includes other important branches, such as bioinformatics and imaging informatics, journals focusing on these branches are not included in this analysis. For the purposes of this study, to define MI core journals, we examined whether the majority of the articles in a journal reflect the core content of clinical informatics as defined by Gardner *et al.* [8]. Some journals that cover those topics to a certain extent, but mainly fall into other domains, were excluded. These domains include: (1) biomedical engineering, physics, devices, images; (2) bioinformatics (except journal names containing 'biomedical informatics'); and (3) general medical, nursing, and pharmacy journals.

We defined the inclusion and exclusion criteria and identified the core MI journals in two sub-steps: First, we started with a simple query to the MEDLINE/PubMed in November 2016 for retrieving all documents assigned any MeSH term in the following three MeSH hierarchies: "Medical Informatics", "Public Health Informatics", and "Nursing Informatics". This query retrieved over 379,000 articles that were published in over 9,000 journals indexed by MEDLINE. The publication sources were sorted by the number of retrieved articles. Two MI researchers (LZ, MT) manually reviewed all journals with 50 or more relevant articles. For journals with less than 50 articles, or approximately 90% of publication sources, they used keyword search (e.g., informatics, information technology, and telemedicine) to identify MI journals. The two researchers separately reviewed the journal list and determined whether a source should be included or not. Disagreements were discussed and full consensus on MI journals was achieved. Second, we generated a list of MI journals by searching the NLM Catalog with the query "Medical Informatics [Broad Subject Term]", which returned 91 journals indexed in MEDLINE. We then manually reviewed these 91 journals to identify MI journals. We found that many journals changed their names over the years. We merged different versions of journal names with the preferred name being the most recent. Finally, we compared and reconciled the journal lists generated by these two substeps.

Steps 2: MI Article Query

We formed a MEDLINE query based upon the identified core MI journals to retrieve all articles published in these journals between 1992 and 2015, as follows: *MI_Journals AND "english" [language] AND 1992[PDAT]:2015[PDAT]*. In this query, the MI_Journals refer to the journals from the previous step, where both the current and previous journal names were used. This expanded our query format to include the variations: *("Stud Health Technol Inform" [Journal] OR "AMIA Annu Symp Proc"[Journal] OR "Methods Inf Med" [Journal]...)*. With this query, we were able to plot the number of MI articles over the past 24 years to see the increasing rate of publications in the MI field.

Step 3: MeSH Term and Article Information Extraction

All of the information about the retrieved articles was stored in MEDLINE's Extensible Markup Language (XML) formatted files. We downloaded related files through NLM e-utilities and parsed those XML files using the Java DOM Parser to get each article's identifying number (PMID), title, abstract, MeSH descriptor terms, and publication date. All extracted data was loaded into a local relational database.

Step 4: MeSH Category Selection

Each article is assigned as many MeSH terms as appropriate to cover the topics of the article. MeSH descriptor terms are organized in 16 major categories, (e.g., category L for *Information Science* and category N for *Health Care*) [6]. Each category is further divided into subcategories, and within each subcategory, descriptors are listed in a hierarchical structure. Each descriptor has a number that indicates its tree location, e.g., computing methodologies [L01.224] is in the Information Science [L01] subcategory of the L category. We obtained the tree location numbers for each MeSH term from MeSH RDF Linked Data [10].

In this paper, we restricted our analysis on the topic trends to focus on MeSH terms in the following three categories: *Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]*, *Information Science [L]*, and *Health Care* [N], among which, *L* category was the main focus. We chose these three categories to investigate MI research trends regarding technology and healthcare in general. Although other MeSH categories (such as diseases, humanities, geographicals) are also valuable to describe different aspects and characteristics of MI publications, they were not included in this analysis.

Step 5: Topic Trend Analysis

Detection of Core Topics: We consider a MeSH term as a core topic if it was frequently assigned to MI articles. We count the total number of publications for each individual MeSH term over the past 24 years without aggregating of the counts from its children concepts in the MeSH hierarchy. We sort the MeSH terms by the number of publications in each selected MeSH category. The top ranked MeSH terms were considered as core topics.

Detection of Trending New Topics: Novel topics were defined as those MeSH terms that debut in recent years with increasing interests by the community reflected by an increasing number of indexed articles. The determination of a MeSH term as a trending topic is done through a combined consideration of three factors: debut year, rate of increase, and number of indexed articles before 2016. The debut year is defined as the year of entry of the term into the MeSH terminology. We only included those MeSH terms debuted in the recent 10 years. The rate of increase of a MeSH terms is calculated using the formula: $a/(b+1)$, where a is the number of articles indexed by the term and published in the recent five years (2011-2015), b is the number of indexed articles published in previous five years (2006-2010), and the number "1" is a parameter to handle the situation when some MeSH terms have zero indexed articles in years (2006 – 2010). After we obtained the debut year, rate of increase, and the total number of indexed articles for each MeSH term, we were able to assign three kinds of rankings to the MeSH terms based on how recent they debut, how large the rate of increase is, and how large the total number of indexed articles is. We then obtained a combined ranking number by sum of the three ranking scores. Thereafter, we ranked the MeSH terms with the combined ranking scores and manually picked those terms that were top ranked as novel topics.

Results

Journal Analysis

After combining the journals that were renamed during their lifespan, we selected 36 journals as core MI journals (see Figure 2). Of which, twelve journals have changed their names: "J Biomed Inform", "Int J Med Inform", "Inform Health Soc Care", "Health Manag Technol", "Comput Inform Nurs", "AMIA Annu Symp Proc", "IEEE J Biomed Health Inform", "J Innov Health Inform", "J Healthc Inf Manag", "Top Health Inf Manage", "J (Inst Health Rec Inf Manag)",

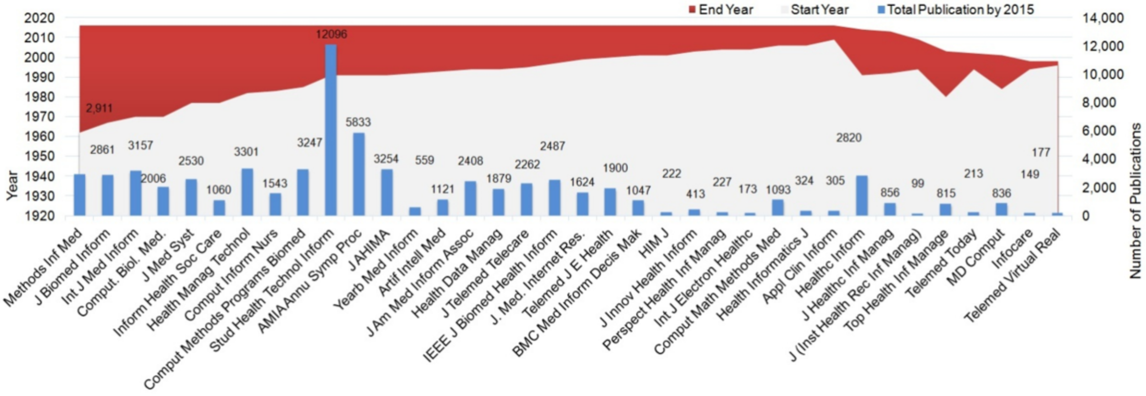

Figure 2 – Core Medical Informatics Journals with Start and End Years and Total Number of Publications.

and "Telemed J E Health". Figure 2 shows the start and end year of each journal, as well as the total number of publications as of 12/31/2015 indexed in MEDLINE/PubMed. The average number of articles per journal is 1,834 [range: 99-12096]. The average age of the journals is 23.7 years [range: 3-55]. In Figure 2, an end year of "2016" indicates that the journal is currently active. For example, "J Biomed Inform" started publishing in 1967 under the original name "Comput. Biomed. Res.". It was renamed using the current title in 2001 and is still active in 2016. One fifth (8 out of 36) of the journals ceased publishing or indexing to MEDLINE during our study period (1992-2015). For example, a journal titled "MD Comput" is still active but the articles are no longer indexed in MEDLINE.

Publication trends

60,862 articles were retrieved from selected MI journals in 1992-2015 (see Figure 3). The increase in the volume of MI publications between 1996 and 2006 almost follows the exponential growth trend. However, a slight decrease was observed between 2006 and 2011. Between 2011 and 2015, the growth of the publication went back to its previous exponential trajectory.

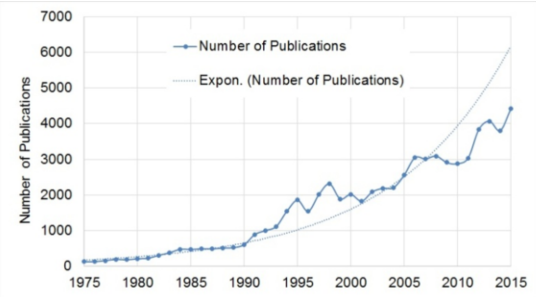

Figure 3– Publication Trends of MI Articles by Year

MeSH Analysis

We extracted a total 8,652 MeSH terms from retrieved articles, with an average of 8.3 MeSH terms per article [range: 1-32]. The MeSH terms were distributed in all 16 categories of MeSH descriptor hierarchy, except the category of Publication Characteristics [V]. Restricting the MeSH terms to the three categories resulted in 2,929 unique MeSH terms, among which, 1,295 were in E, 378 were in L, and 1,260 were in N, while some terms overlapping categories (see Table 1). Within category L, only 68 MeSH terms belong to the branches of "Medical Informatics", "Public Health Informatics", and "Nursing Informatics". Among all of the Information Science MeSH terms, we show the top 50 MeSH terms in Figure 4 sorted by the total number of indexed articles, the top 5 being: computerized medical records systems, algorithms, software, internet, and user-computer interface. The growth along every 6 years is also shown in Figure 4. Figure 5 shows trending new topics that debuted in the last decade with increasing research interests in the MI community, including: social media, mobile application, health information exchange, and cloud computing. Besides the L category, we also detected a list of core and trending new topics in the E and N categories (see Table 1).

Discussion

We identified 36 core informatics journals with over 60,000 MI articles published in the last two decades. The dissemination of MI research is currently occurring at an exponential pace. In addition, new MI journals are becoming increasingly more specialized. For example, some of the most recent MI journals include: "Telemed J E Health" (2000) focusing on publications around provision of telemedicine or telehealth; "Inform Primary Care" (2003) focusing specifically on primary care settings; and "Appl Clin Inform" (2009) providing a venue for publications focused on applied MI projects and studies.

Although we used a different set of MI articles compared to DeShazo's work, the publication trends look similar in terms of exponential growth of publication numbers over the past 40 years. One caveat still remains – no growth was observed between 2006 and 2011, compared to the previous and following periods. One possibility is the effect of the financial crisis around 2007- 2008 and the great recession in world market during the late 2000s and early 2010s.

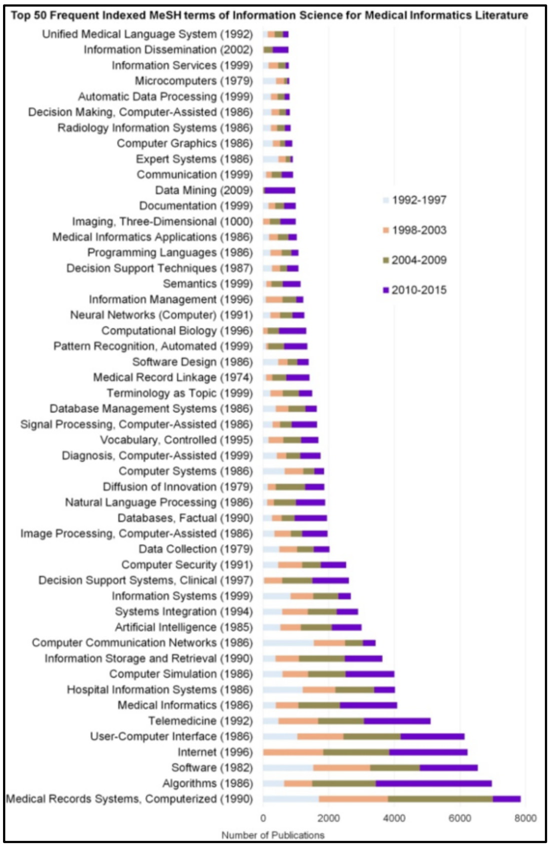

Figure 4– Core Topics in Information Science Category

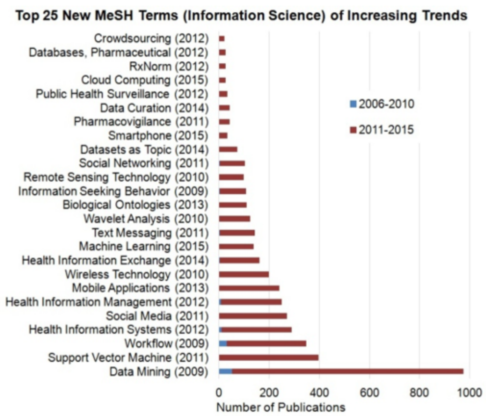

Figure 5– Novel Topics in the Information Science Category

Regardless of the stagnation period during 2006-2011, two booms in the number of publications were observed right before and after that period, 2004-2005 and 2011-2013 respectively. We suspect that these two waves may have partially been stimulated by the U.S. national health information technology (HIT) regulatory and incentive programs and similar efforts made by other countries.

In 2004, President Bush outlined a plan to ensure that most Americans have electronic health records (EHRs) within the next 10 years. It was stated in a 2005 report by the U.S. Department of Health and Human Services that "HIT is a high

Table 1– Core and New Topics for categories E and N

Cat.	E	N
Core	Electronic Health Records; Models, Theoretical; Models, Statistical	Reproducibility of Results; Surveys and Questionnaires; Delivery of Health Care; Confidentiality; Efficiency, Organizational; Attitude of Health Personnel; Pilot Projects; Models, Statistical; Organizational Case Studies; Remote Consultation; Medical Records; Quality Assurance, Health Care; Organizational Innovation; Practice Guidelines as Topic; Evaluation Studies as Topic; Patient Satisfaction; Quality of Health Care; Forms and Records Control; Patient Education as Topic
New	Health Records, Personal (2008); Precision Medicine (2008); Early Detection of Cancer (2009); Actigraphy (2010); Genome-Wide Association Study (2009); Accelerometry (2012); Brain-Computer Interfaces (2009); Clinical Alarms (2009); Health Smart Cards (2015); Drug Repositioning (2010); Clinical Decision-Making (2015); Epidemiological Monitoring (2012);	Meaningful Use (2012); Patient Safety (2011); Quality Improvement (2010); American Recovery and Reinvestment Act (2009); Clinical Coding (2010); Radio Frequency Identification Device (2009); Health Level Seven (2009); Accountable Care Organizations (2011); Patient Protection and Affordable Care Act (2010); Self Report (2010); Patient Preference (2009); Data Accuracy (2015); Patient Handoff (2012); Health Literacy (2009); Knowledge Management (2010); Comparative Effectiveness Research (2009); Medication Reconciliation (2010); Epidemics (2010); Checklist (2009); Patient Discharge Summaries (2013); Biometric Identification (2009); Tertiary Care Centers (2012); Health Communication (2010); Patient Outcome Assessment (2013); Medication Adherence (2008)

Cat. = Category, E = Analytical, Diagnostic and Therapeutic Techniques and Equipment [E], N = Health Care, Core = core medical informatics topics, New = New trending topics

priority for American healthcare system and the U.S. economy. IT is a pivotal part of transforming our healthcare system." [11] The 2011 and recent increase was perhaps stimulated by the Health Information Technology for Economic and Clinical Health (HITECH) Act, enacted as part of the American Recovery and Reinvestment Act (ARRA) of 2009, which aimed to promote the adoption and meaningful use of HIT. As a result, the initiatives have increased the adoption of EHRs across U.S. hospitals from about 30% in 2011 to almost 85% in 2015 [12]. The increased adoption in turn produced many more studies and projects, which has resulted in not only more publications but also new topics, such as meaningful use, quality improvement and ARRA, as demonstrated in Table 1.

For the MeSH terms analysis, focusing on a broader set of MeSH terms than DeShazo’s helped us to identify many core topics outside of MI MeSH terms. For example, in the L category, we found a large number of publications in topics, such as algorithms, user-computer interface, natural language processing, controlled vocabulary, internet, and computer simulation, which are beyond MI subcategory. By including the general health care (N) and techniques (E) categories, our findings reveal the continuously growing core topics and newly appeared topics that medical informaticiains have been dealing with in collaboration with professionals in healthcare and other technological fields.

Some core topics experienced increasing interests of the MI community in recent years (e.g., algorithms, data mining, database, pattern recognition, medical records linkage). The number of publications in several popular MeSH topics remains stable over the years (e.g., software, terminology and vocabulary, artificial intelligence). This indicates persistent interest in those topics. On the other hand, there are several topics that show a decreasing number of publications in recent years (e.g., computer or information systems, microcomputers, and computer communication networks). This might indicate that more specific MeSH terms (along with newly added ones) were used to index articles rather than using broader terms, such as computer or information systems. Alternatively, this might indicate outdated terms with only infrequent use today (e.g., microcomputers referring to "small computers using microprocessor chips") [13]. Overall, comparing the core topics extracted from the MI articles in the *L and E* categories, the majority of the terms were mentioned as part of the core content defined by Gardner *et al.*[8]

Our approach also detected new trending topics that have appeared only in the recent years. The most frequent new topic was the data mining, introduced in 2009 with almost 1,000 publications before the end of 2015. The second most frequent new topic was machine learning (including support vector machines). With the volume of health related data information growing exponentially, it is not surprising that new tools, such as data mining or machine learning are increasingly applied to process these data. Other literature also supports our findings on the increasing application of machine learning and data mining in healthcare [14–16]. Other frequent new topics in information science include concepts related to new technologies (e.g., wireless, smartphone, sensor, and mobile), as one might expect. In agreement with other literature [17,18], we found an increasing research focus on social media and related knowledge exploration techniques (e.g., crowdsourcing and pharmacovigilance) in healthcare. In terms of *Analytic, Diagnostic and Therapeutic Techniques*, electronic/personal health records are important tools to support analytics and statistical modeling, both of which show large increased interests to the MI community. In addition, recent advances in genetic testing and other molecular or cellular analysis gave birth to new topics such as precision medicine and genome-wide association study. In terms of topics in *Health Care*, two groups of trending topics were identified: (1) patient-centered care, which is indicated by the topics such as patient participation, patient preference, patient satisfaction, patient handoff, and health literacy; (2) medical safety, which is reflected by the trending terms like patient safety, medication reconciliation, medication adherence, and clinical alarms.

This study has several limitations: First, we collected articles from a set of selected MI journals. We could have missed MI articles published in journals in other fields, such as medicine, nursing, public health, library science, computer science, and engineering. However, articles published in a non-MI journal may not truly be MI, although it was assigned a MeSH term in the MI hierarchy. For example, "databases, fractal" is a MeSH term in the MI hierarchy and it can be assigned to many non-MI articles that used databases in their studies. As our study includes core MI journals, we believe that our findings capture the mainstream topics in the MI field. Second, using MeSH as part of our search and analysis strategy can be problematic. The MeSH terms may not totally reflect diverse topics in the domain of MI. The MeSH indexes can be subjective, although they were assigned by trained subject matter experts. Also, the addition of new MeSH terms often belatedly follows the actual research trend. Although these new MeSH terms may be assigned to literature retrospectively, our trend analysis based on MeSH may be chronologically behind the real trend. Third, our study does not include analyses of journal's impact factors and other bibliometrics, such as citations. Lastly, we focused on the MeSH term analysis over the last 24 years; the patterns of the trending and emerging topics detected in this study could be easily outdated as time passes.

Conclusions

We used MeSH terms to describe ongoing and emerging trends in MI publications. Our results indicate that over the past two decades, the body of MI literature has grown at an almost exponential pace. Our analysis indicates that while some popular topics in MI publications remain stable throughout the years (e.g., software, artificial intelligence), new topics also emerged in the past few years (e.g. data mining, social media). Our findings can be used to examine the existing MI educational programs (in terms of content coverage) and suggest new research avenues (e.g., emerging topics such as crowdsourcing).

Acknowledgements

This study was supported by the Agency for Healthcare Research and Quality (R01HS022728 and R01HS024264).

References

[1] https://www.nlm.nih.gov/hsrinfo/informatics.html
[2] M.F. Collen. Origins of medical informatics. *West J Med* 145(1986),778–85.
[3] https://www.amia.org/clinical-informatics-board-review-course/history
[4] https://www.amia.org/education/programs-and-courses?field_state_value=All&term_node_tid_depth=All
[5] J.P. DeShazo, D.L. Lavallie, F.M. Wolf, Publication trends in the medical informatics literature: 20 years of "Medical Informatics" in MeSH, *BMC Med Inform Decis Mak* 9(2009),7.
[6] https://www.nlm.nih.gov/bsd/disted/meshtutorial/me.
[7] A.R. Aronson, J.G. Mork, C.W. Gay, S.M. Humphrey, W.J. Rogers, The NLM indexing initiative's medical text indexer, *Stud Health Technol Inform* 107(2004),268–72.
[8] R.M. Gardner, J.M. Overhage, E.B. Steen, B.S. Munger, J.H. Holmes, J.J. Williamson, et al. Core Content for the Subspecialty of Clinical Informatics. *J Am Med Informatics Assoc* 16(2009):153–7.
[9] https://www.nlm.nih.gov/bsd/serfile_addedinfo.html.
[10] https://id.nlm.nih.gov/mesh/.
[11] The Lewin Group. Health information technolog leadership panel. 2005.
[12] https://dashboard.healthit.gov/evaluations/data-br.
[13] https://www.ncbi.nlm.nih.gov/mesh/?term=microcompu.
[14] I. Kadi, A. Idri, J.L. Fernandez-Aleman, Knowledge discovery in cardiology: A systematic literature review, *Int J Med Inform* 97(2017) 12–32.
[15] M. Herland, T.M. Khoshgoftaar, R. Wald, A review of data mining using big data in health informatics. *J Big Data* 1(2014),2.
[16] D.A. Clifton, K.E. Niehaus, P. Charlton, G.W. Colopy, Health Informatics via Machine Learning for the Clinical Management of Patients, *Yearb Med Inform* 10(2015),38–43.
[17] M.P. Hamm, A. Chisholm, J. Shulhan, A. Milne, S.D. Scott, T.P. Klassen, et al, Social media use by health care professionals and trainees: A scoping review, *Acad Med* 88 (2013) 1376–83.
[18] L. Sinnenberg, A.M. Buttenheim, K. Padrez, C. Mancheno, L. Ungar, R.M. Merchant, Twitter as a Tool for Health Research: A Systematic Review, *Am J Public Heal* 107(2017), e1-e8.

Address for correspondence

Liqin Wang: (857) 282-4089, lwang@bwh.harvard.edu

399 Revolution Drive, Suite 1315

Somerville, MA 02145