Application of Machine and Deep Learning Algorithms in Intelligent Clinical Decision Support Systems in Healthcare

Jong Taek Kim*

Department of Pathology, Immunology and Laboratory Medicine, University of Florida College of Medicine, Gainesville, United States

Abstract

Objective: The purpose of this paper is to review the PubMed/MEDLINE literature for articles that discuss the use of machine learning (ML) and deep learning (DL) for clinical decision support systems (CDSSs).

Materials and Methods: To identify relevant articles, we searched PubMed/MEDLINE through December 2nd, 2017. We identified a total of 283 studies.

Results: The number of ML and DL associated CDSS articles increased significantly beginning around 2010. The most common type of advanced artificial intelligence (AI) methodologies that the articles evaluated was neural networks also known as DL (n=109) followed by ML (n=86). The most common types of ML algorithm were support vector machines (n=78), logistic regression analysis (n=38), random forest (n=26), decision tree (n=25), and k-nearest neighbour (n=21). Cardiology, oncology, radiology, surgery, and critical care/ED were the most commonly represented specialties. Only 19 out of 283 (6.7%) ML and DL associated CDSS articles reported an effect on the process of care or patient outcomes.

Discussion: The current decade has seen research efforts and attention increase significantly in creating CDSS tools with the advanced AI methodologies of DL and ML. Although the research experiments demonstrate success, the scope of AI technology is still limited to a well-defined task. Also, most of these studies lack patient-oriented outcomes necessary to justify its widespread application in healthcare.

Conclusion: There is a clear upwards trend in ML and DL research in healthcare. However, in order to effectively translate successful AI research into the patient care, more clinically-relevant studies must be pursued.

Keywords: Clinical decision support; Artificial intelligence; Machine learning; Deep learning

Introduction

The last several years have seen a significant resurgence in optimism for using AI tools in healthcare. It is very common to see popular outlet publishers such as Harvard Business Review and Forbes publish articles that predict-however unrealistically-that AI technology will soon replace doctors [1,2]. This renewed optimism and hopefulness has its origins-at least in part-with recent advances and successes in machine learning (ML) and deep learning (DL) research in the non-healthcare sector of industry. The concept and design of ML and DL algorithms are not new, but the increased availability of large quantities of data, coupled with equally impressive computing power, enabled the kinds of ML and DL success seen this decade [3]. Examples include IBM's Deep Blue beating the world's best Chess player (Garry Kasparov) 20 years ago, IBM's Watson winning Jeopardy by beating its best player, Google's AI computer AlphaGo beating the world's best Go player (considered a much more complex game than Chess), Google's success in building a safe self-driving car, and remarkable results in Google's FaceNet and Facebook's DeepFace facial recognition research [4-9].

This success in ML and DL research in non-healthcare areas has inspired researchers to apply the technology to the healthcare domain. In dermatology, convolutional neural network (CNN) algorithms performed skin cancer classification as well as board-certified dermatologists [10]. In pathology, Google researchers developed an algorithm that outperformed board-certified pathologists in detecting a lymph node metastasis on hematoxylin and eosin slides in the Camelyon16 Challenge [11]. In radiology, CheXNet-based on a CNN algorithm-was able to detect pneumonia on a chest X-ray better than board-certified radiologists [12].

Although AI technology in healthcare is celebrated, it is more important than ever to understand what AI is and how it might enable medical professionals to deliver better healthcare. The AI technology available today is too narrow in scope to fully replace a doctor's role in healthcare (Table 1). The doctor's role in clinical work is social-technically multi-faceted, involving multiple levels of interaction and collaboration from different teams (other physicians, nurses, social workers, pharmacist, therapists, etc.) [13]. It is inherently more complex than mastering and outperforming humans at one specific and narrow medical task. With today's available AI technology, a self-sufficient, self-aware, autonomous AI doctor is not simply feasible. However, currently available AI technologies are most suited to enable doctors as a form of clinical decision support system (CDSS) rather than replacing them [14].

Shortliffe and Cimino define CDSS as a set of computer applications within the clinical information system (CIS) and electronic health record (EHR) that empowers healthcare professionals in making improved clinical decisions [15]. Most traditional types of CDSS include order

*Corresponding author: Jong Taek Kim, Department of Pathology, Immunology and Laboratory Medicine, University of Florida College of Medicine, Gainesville, United States, Tel: 251-622-3978; E-mail: jtk622@gmail.com

Artificial Intelligence (AI)	An ability for a computer machine to simulate human intelligence. Narrow AI: A specific or well-defined task. Artificial General Intelligence (AGI): Human-level intelligence (and beyond).
Machine teaming (ML)- A subset of AI	ML refers to a subset of AI that can learn and improve at tasks with experience without explicitly programmed to do so.
Deep Learning (DL)- A subset of ML, also known as Deep Neural Net or Artificial Neural Net	DL is a set of algorithms based on a multi-layered neural network that allows the system to learn representation of data. Examples: Convolutional Neural Network and Recurrent Neural Network.

Table 1: Artificial intelligence, machine learning, and deep learning definitions [3,18].

sets, documentation templates, computerized guidelines, alerts, advice and reminders, and inference engines while more advanced CDSS include ML algorithms, DL algorithms, or other elaborate software systems such as Bayesian networks and natural language processing (NLP) [16]. This is summarized in Table 2.

A few comprehensive reviews and evaluations of CDSS effectiveness exist, published in 1998, 2005, and 2011, that conclude that CDSSs improve practitioner performance and patient outcomes [17-19]. But these studies predate recent ML and DL research successes and breakthroughs. In addition, their focus was not the application of ML, DL, and AI technologies to CDSSs. Although more recent, comprehensive HIT reviews appeared in 2016, they also did not remark on the recent incredible advances in ML and DL methodologies in regards to CDSS designs [20,21]. This observation prompted us to conduct a PubMed/MEDLINE review and survey CDSS research that integrates ML and DL methodologies.

The purpose of this paper is to provide a survey and review of the PubMed/MEDLINE literature to gauge the extent to which ML and DL methodologies have been incorporated into CDSS research. In addition, the clinically-oriented studies will be selected for further analysis. By so doing, we hope to present an accurate and realistic perspective regarding current trends in applying ML and DL methodologies in CDSS biomedical research, and the results attained thus far.

Materials and Methods

Author accessed PubMed/MEDLINE (https://www.ncbi.nlm.nih.gov/pubmed/) on 12/02/2017 to search relevant articles for this study. Author focused on the following keywords in the Title and Abstract: clinical decision support (CDS), AI, ML, DL, software, and algorithm (Table 3). Inclusion criteria are as follows: 1) Articles published up to 12/02/2017 in English language, 2) Articles with research focus on CDSS, 3) Articles with method consisting of machine, deep learning, and complex software algorithms. Exclusion criteria are as follows: 1) Articles not in English language, 2) Articles with research focus other than CDSS, 3) Articles with research methodologies other than machine and deep learning systems, 4) Abstract and full-text articles were not available.

The search diagram is shown in Figure 1. Author identified additional ML and DL articles from the "Reference Review" and "Seminal Paper Citation Index Search" as shown. Seminal papers are defined as those papers that have been cited at least 300 times [17-22].

Type I: "Simple" CDSS	Order Sets, Documentation Templates, Inference Knowledge Engine, Alerts, Reminders, Recommendations Based on Computerized Guidelines (Not subject to FDA regulations) [23]
Type 2: "Intelligent" CDSS	Machine Learning Algorithms, Deep Learning Algorithms, or Multi-Faceted Software Algorithms such as Bayes Networks and Support Vector Machines

Table 2: Two types of CDSS defined- "Simple" CDSS and "Intelligent" CDSS.

CDS and AI	(Clinical decision support[Title/Abstractj) AND Artificial intelligence[Title/Abstract]
CDS and ML	(Clinical decision support[Title/Abstract]) AND Machine learning[Title/Abstract]
CDS and DL	(Clinical decision support[Title/Abstract]) AND Deep learning[Title/Abstract]
CDS and Software	(Clinical decision support[Title/Abstract]) AND Software[Title/Abstract]
CDS and Algorithm	(Clinical decision support[Title/Abstract]) AND Algorithmride/Abstract)
CDS and Bayesian	(Clinical decision support[Title/Abstract]) AND Bayesian[Title/Abstract]

Table 3: PubMed/MEDLINE Keyword search method.
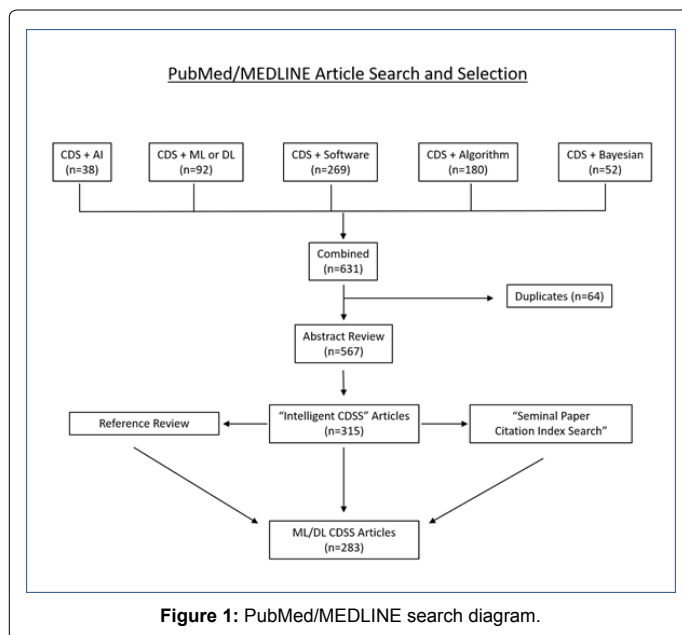
Figure 1: PubMed/MEDLINE search diagram.

Author plotted the number of included articles by year to identify any trends. We also tallied the types of ML and DL methodologies used. For ML methodology, we also tallied the types of ML algorithms the studies used (e.g., random forests, k-nearest neighbor, etc.). Then, we categorized the articles by medical specialty and types of condition or disease investigated. We also extracted any information on the CDSS' effect on the process of care or patient outcomes.

Results

The keyword search for "Clinical Decision Support" (CDS) in the title and abstract were combined with other relevant keywords. This step yielded 38 articles with "CDS+Artificial Intelligence", 92 articles with "CDS+Machine Learning" or "Deep Learning", 269 articles with "CDS+Software", 180 articles with "CDS+Algorithm" and 52 articles with

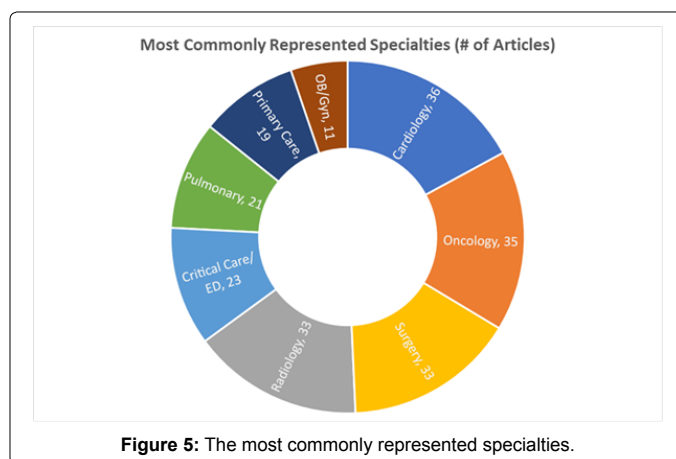"CDS+Bayesian". After pooling the results and removing duplicates, there was a total of 567 articles (Figure 1).
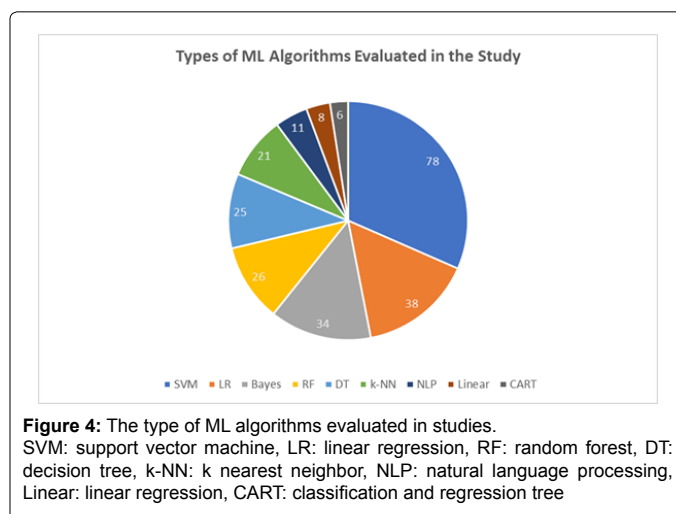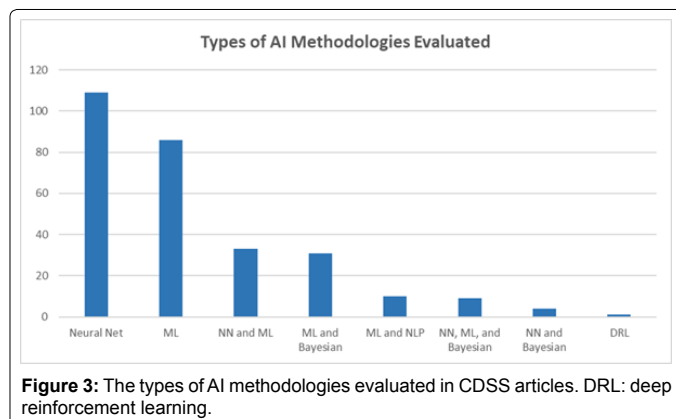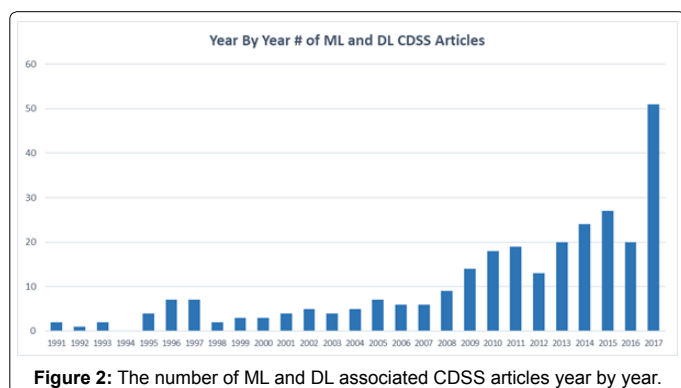
Author then reviewed the abstract or texts to determine eligibility for this review (CDSS research having ML or DL methodologies). Overall, 315 articles met the definition of Type 2 "Intelligent CDSS". Author further narrowed the articles with a focus on ML and DL methodologies (n=92). After combining these articles with additional ML and DL articles identified from the "Reference Review" (n=185) and "Seminal Paper Citation Index Search" (n=6) there was a final total of 283 articles included in this review.

The number of ML/DL in CDSS articles were relatively few from 1991 to 2008, and then began to increase noticeably in 2008 and more significantly beginning around 2010 (Figure 2). The most popular AI methodology was DL, historically referred to as artificial neural networks or deep neural networks, (n=109) followed by ML (n=86). Many researchers simultaneously evaluated both ML and DL methodologies (n=33). ML and Bayesian methodologies were also commonly studied together (n=31). The remainder include ML and NLP together (n=11), DL, ML, and Bayesian together (n=9), DL and Bayesian together (n=4), and deep reinforcement learning (n=1). This is illustrated in Figure 3.

We broke ML methodology further down into various types of algorithms in Figure 4. Support vector machine was the most commonly used ML algorithm (n=79) followed by logistic regression analysis (n=38). Other popular methodologies included Naïve Bayes (n=34), random forest (n=27), decision tree (n=25), k-nearest neighbor (n=25), natural language processing (n=11), linear regression (n=8), and classification and regression tree (n=6).

Regarding medical specialties represented by these articles, as shown in Figure 5, cardiology (n=36), oncology (35), radiology (n=34), and surgery (n=33) were the most common over thirty articles each. Other notable specialties included critical care/ED (n=23), pulmonary (n=21), primary care (n=19) and Ob/Gyn (11). Also, the commonly studied conditions or variables are summarized in Table 4 for the cardiology, oncology, radiology, surgery, and critical care/ED specialties.

Out of 283 articles, 18 research studies reported an effect on the process of care. One research study reported the effect on both the process and outcome of care (Figure 6). Out of 283 articles, only 22 studies were able to collect prospective data from the patients. The remaining studies (n=260) relied on retrospectively collected data or data from a public data repository. For one study, the data collection method was not available for review. The complete set of information

Figure 3: The types of AI methodologies evaluated in CDSS articles. DRL: deep reinforcement learning.

Figure 4: The type of ML algorithms evaluated in studies.
SVM: support vector machine, LR: linear regression, RF: random forest, DT: decision tree, k-NN: k nearest neighbor, NLP: natural language processing, Linear: linear regression, CART: classification and regression tree

Figure 5: The most commonly represented specialties.

Figure 2: The number of ML and DL associated CDSS articles year by year.
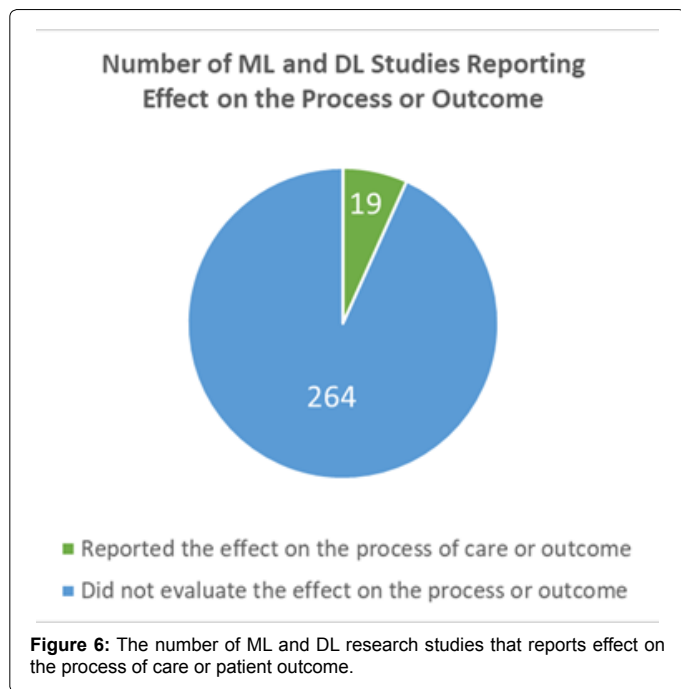
can be found in the supplementary material online where we summarize the entire 283 studies (Supplementary Table 1).

Discussion

Advances in AI research and technology over the current decade have been remarkable. However, these advances and breakthrough successes are stilled considered narrow type AI, that is, achieving a human-level competency for a specific task. Likewise, the application of ML and DL algorithms in healthcare has been quite remarkable as well. The enthusiasm and optimism are evident from the number

Cardiology	Oncology	Radiology	Surgery	Critical Care/ED
– Heart failure – Coronary artery disease – Myocardial infarction – Arrhythmia and ECG	– Treatment planning – Cancer diagnosis – Prognosis – Outcome prediction – Survival prediction	– MRI protocol selection – Diagnosis – Breast cancer – Lung cancer – Brain cancer	– Surgical prediction – Complication risk – Mortality rate – Outcome – Endocrine disease	– Detection of sepsis – ICU prediction – Mortality – Outcome

Table 4: Common types of conditions or variables studied with Mack!WE and deep learning algorithms.

Figure 6: The number of ML and DL research studies that reports effect on the process of care or patient outcome.

of publications that met our criteria for review. Nevertheless, these amazing feats of AI research are still narrow, for example being trained to render a specific diagnosis or predict one or more outcomes of a given disease.

Although the number of articles published about ML and DL CDSS methodologies has multiplied many folds this decade, only a fraction has reported on patient outcomes. Most studies used data sets from a public repository or one institution's retrospective health records to train the algorithms. Most importantly, these studies lack information about the efficacy in a clinical environment and patient care setting. Thus, although ML and DL algorithms may perform extremely well in a controlled and non-clinical situation, whether that success will translate into clinical patient care is not certain nor guaranteed. One limitation not addressed by this body work is that the healthcare and patient care system is much more than a cleanly preprocessed and well-annotated data set. There are intrinsic uncertainties and complex clinical contexts that cannot be easily reproduced by a set of clean and annotated data sets [23]. More clinically oriented studies and trials are necessary to accurately evaluate the ML and DL CDSS's efficacy and value in healthcare [24].

The "technological singularity" or AGI refers to a point in time when AI will match and surpass human intelligence [25]. The topic of how one can achieve such AGI in healthcare is much discussed and debated. It is difficult to avoid reading posts or reports from popular media outlets where doctor's profession is allegedly threatened by an AI system. There is no precedent in achieving AGI in healthcare but

Guruduth Banavar, then-IBM Watson's Chief Science Officer, discussed what it might entail in a recent conference with other AI experts [26]. He argued that "One cannot achieve AGI by going straight after AGI, but by repeatedly achieving narrow AI". He opined that a narrow AI has to be done many times over systematically while finding a common interface, essentially creating an AGI platform in the process. In essence, narrow AI would augment and enable humans' abilities one by one until the entire repertoire of human intelligence is simulated. Of note, this is but one of several positions shared by AI experts and further discussion is outside the scope of this paper.

In medicine, the amount of information that needs to be processed by a doctor to make a well-informed and best clinical decision can be overwhelming. There is a recognized mismatch between the complexity of medicine and the doctor's ability to process it all [27]. However, many HIT and CDSS tools exist to help doctors navigate through a sea of health information and data to make the best clinical decisions. Our review has shown that several successful ML and DL CDSS studies have emerged that hope to augment and enable doctor's innate abilities in real clinical healthcare settings. There is clear upward trend for ML and DL research in healthcare.

The top two most commonly represented specialties were cardiology and oncology. In the field of cardiology, the interest in AI and ML research can be seen in early 1990's [28,29]. Over the years, the research has shown good prediction performances in cardiology (89.23 ± 8.87% classification accuracy and 84.84 ± 8.68% area-under-curve). The popular topics included ECG, myocardial infarction, and heart failure. They showed promise as useful clinical decision support, but most of these researches were not clinically evaluated nor validated. In addition, the regulatory guidelines for building, evaluating, and validating clinical decision support tools were not entirely clear until recently. In December of 2017, the Food and Drug Administration (FDA) published draft guidelines on how it intends to regulate clinical decision support tools for both clinicians and patients [30]. These will be extremely beneficial for biomedical researchers and clinicians involved in developing and implementing AI and ML-driven CDSS. As a result, we are beginning to see the FDA granting approval for clearance for various AI and ML-based CDSS in cardiology. One of the first FDA-approval was Arterys Cardio DL medical imaging, which is based on deep learning algorithms [31]. Other CDSS tools cleared by the FDA include CADence (stethoscope and ECG in one device), AliveCor Heart Monitor, and KardiaBand (the first ECG medical device accessory for Apple Watch) [32-34].

The second most common specialty represented in the review was oncology. Similar to cardiology, AI and ML research in oncology has shown good prediction performances over the years (91.52 ± 8.97% classification accuracy and 90.3 ± 7.38% area-under-curve). Treatment planning, diagnostic and prognostic areas were the most studied variables. The authors of the study concluded that their models could be useful and assist clinicians in decision making. However, these studies in oncology did not evaluate its AI and ML-based CDSS research

for clinical and patient outcomes. None of the studies were bridged to clinical trials and prospective studies that must take place before obtaining the FDA clearances. It is likely that translating biomedical research into clinical trials is expansive and challenging with regulatory hurdles, particularly in the field of oncology.

IBM Watson for Oncology is relatively well-known for its efforts for developing guidance for cancer treatments using the supercomputer, one of the popular areas of AI and ML research. However, it has struggled and failed to live up to expectations. The collaboration Between M.D. Anderson and IBM Watson Oncology was recently discontinued; citing challenges with integrating the algorithms into the patient care environment [35]. The shortcomings of IBM Watson Oncology so far underscore the difficulties in exploring and implementing AI and ML-based CDSS into the healthcare model. There is a need for a better and improved methodology for translating promising AI-based biomedical research into a clinical care model.

The AI and ML research in CDSS has been taking place for a long time. There have been both successes and failures in translating biomedical research into useful tools for both clinicians and patients. The future studies would benefit by collaborating with clinicians early on while developing the framework for the CDSS designs. By involving healthcare professionals in the early stage, there is a better chance of successfully guiding AI and ML-based CDSS thru the creation, validation, and deployment within the clinical care setting. There are legal, ethical, and societal implications that would be better off if carefully thought out in the beginning. The AI and ML research has shown very promising results in literature so far. By overcoming many challenges associated with integrating the models into patient care, there is chance that AI and ML-based CDSS can assist clinicians and improve patient outcomes at the same time.

Limitations

Our review has several limitations. First, a single author reviewed abstracts and papers against the inclusion criteria. Thus it is possible that this review missed relevant articles and included potentially irrelevant articles. However, given that the technologies for which he was searching are very specific, and given that there is a specific MeSH heading for CDSSs, the likelihood that he missed a significant number-or inappropriately included a significant number-of articles is low. The trends in the number of articles per year are also unlikely to be affected. Second, a single author carried out data abstraction. It is possible that there were errors. Nevertheless, our high-level survey of specialties, diseases/conditions, and effects on processes and outcomes of care demonstrated clear trends and tendencies that are also unlikely to be impacted by such errors. Finally, limiting the search to PubMed/Medline could exclude some of the relevant literature published in computer science and engineering journals. However, the PubMed/Medline search was supplemented with "Reference Review" and "Seminal Paper Citation Index Search".

Conclusion

Experimental research into ML and DL methodologies for CDSSs has demonstrated promise in the current decade. Our review identifies reasons to be optimistic, but also a basis to be realistic about the near and medium term possibilities that AI technologies might bring to healthcare. Perhaps the most important requirement of CDSS research is demonstrating improved patient outcomes or the process of care. As clearer regulatory guidelines have emerged recently this should also help biomedical researchers, healthcare organizations, and technology companies in choosing the most proper paths in designing and conducting CDSS research that can be bridged into clinical practice.

Supplementary Material

Supplementary Table 1 is available at Journal of Health & Medical Informatics Website.

References

1. https://hbr.org/2016/10/robots-will-replace-doctors-lawyers-and-other-professionals

2. https://www.forbes.com/sites/haroldstark/2017/07/10/prepare-yourselves-robots-will-soon-replace-doctors-in-healthcare/#7b9716a252b5

3. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagation errors. Nature 323: 533-536.

4. https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/

5. http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?pagewanted=all

6. https://www.nytimes.com/2016/03/16/world/asia/korea-alphago-vs-lee-sedol-go.html

7. Teoh ER, Kidd DG (2017) Rage against the machine? Google's self-driving cars versus human drivers. J Safety Res 63: 57-60.

8. Schroff F, Kalenichenko D, Philbin J (2015) FaceNet: A unified embedding for face recognition and clustering. IEEE arXiv: 1503.03832.

9. Taigman Y, Yang M, Ranzato M, Wolf L (2014) DeepFace: Closing the gap to human-level performance in face verification. IEEE 1701-1708.

10. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, et al. (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542: 115-118.

11. Liu Y, Gadepalli K, Norouzi M, Dahl GE, Boyko A, et al. (2017) Detecting cancer metastases on gigapixel pathology images.

12. https://stanfordmlgroup.github.io/projects/chexnet/

13. Wears RL, Berg M (2005) Computer technology and clinical work: Still waiting for Godot. JAMA 293: 1261-1263.

14. Takahashi R, Kajikawa Y (2017) Computer-aided diagnosis: A survey with bibliometric analysis. Int J Med Inform 101: 58-67.

15. Shortliffe EH, Cimino JJ. Biomedical informatics: Computer applications in health care and biomedicine.

16. Aljaaf AJ, Al-Jumeily D, Hussain AJ, Fergus P, Al-Jumaily M, et al. (2015) Toward an optimal use of artificial intelligence techniques within a clinical decision support system. IEEE: 548-554.

17. Hunt DL, Haynes RB, Hanna SE, Smith K (1998) Effects of computer-based clinical decision support systems on physician performance and patient outcomes: A systematic review. JAMA 280: 1339-1346.

18. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, et al. (2005) Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. JAMA 293: 1223-1238.

19. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW (2011) Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Informatics Assoc 18: 327-334.

20. Brenner SK, Kaushal R, Grinspan Z, Joyce C, Kim I, et al. (2016) Effects of health information technology on patient outcomes: a systematic review. J Am Med Informatics Assoc 23: 1016-1036.

21. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, et al. (2006) Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 144: 742-752.

22. Kawamoto K, Houlihan CA, Balas EA, Lobach DF (2005) Improving clinical

practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 330: 765.

23. Cabitza F, Rasoini R, Gensini GF (2017) Unintended consequences of machine learning in medicine. JAMA 318: 517-518.

24. Liu JLY, Wyatt JC (2011) The case for randomized controlled trials to assess the impact of clinical information systems. J Am Med Inform Assoc 18: 173-180.

25. Russell SJ (2010) Artificial intelligence: A modern approach prentice hall.

26. https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1

27. Obermeyer Z, Lee TH (2017) Lost in thought- The limits of the human mind and the future of medicine. N Engl J Med 377: 1209-1211.

28. Xue Q, Hu YH, Tompkins WJ (1992) Neural-network-based adaptive matched filtering for QRS detection. IEEE Trans Biomed Eng 39: 317-329.

29. Furlong JW, Dupuy ME, Heinsimer JA (1991) Neural network analysis of serial cardiac enzyme data. A clinical application of artificial machine intelligence. Am J Clin Pathol 96: 134-141.

30. https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm587819.pdf

31. Ochs R (2017) Arterys Cardio DL 510(k) Premarket Notification. Silver Spring.

32. Zuckerman B (2017) CADence System 510(k) Premarket Notification. Silver Spring.

33. Boniske A, Zuckerman BD, Cavanaugh KJ (2017) Alivecor Heart Monitor 510(k) Premarket Notification. Silver Spring.

34. Zuckerman BD (2017) Kardia Band System 510(k) Premarket Notification. Silver Spring.

35. Schmidt C (2017) Anderson Breaks with IBM Watson, raising questions about artificial intelligence in oncology. J Natl Cancer Inst 109.