



A Scoping Review of Artificial Intelligence Algorithms in Clinical Decision Support Systems for Internal Medicine Subspecialties

Ploypun Narindrarangkura¹ Min Soon Kim² Suzanne A. Boren²

¹Institute for Data Science and Informatics, University of Missouri, Columbia, United States

²Department of Health Management and Informatics, University of Missouri Institute for Data Science and Informatics, University of Missouri, Columbia, United States

Address for correspondence Suzanne A. Boren, PhD, MHA, Department of Health Management and Informatics, University of Missouri School of Medicine, CE707 Clinical Support & Education Building, DC006.00, Columbia, MO 65212, United States (e-mail: BorenS@health.missouri.edu).

ACI Open 2021;5:e67–e79.

Abstract

Objectives Artificial intelligence (AI)-based clinical decision support systems (CDSS) have been developed to solve medical problems and enhance health care management. We aimed to review the literature to identify trends and applications of AI algorithms in CDSS for internal medicine subspecialties.

Methods A scoping review was conducted in PubMed, IEEE Xplore, and Scopus to determine articles related to CDSS using AI algorithms that use deep learning, machine learning, and pattern recognition. This review synthesized the main purposes of CDSS, types of AI algorithms, and overall accuracy of algorithms. We searched the original research published in English between 2009 and 2019.

Results Given the volume of articles meeting inclusion criteria, the results of 218 of the 3,467 articles were analyzed and presented in this review. These 218 articles were related to AI-based CDSS for internal medicine subspecialties: neurocritical care ($n = 89$), cardiovascular disease ($n = 79$), and medical oncology ($n = 50$). We found that the main purposes of CDSS were prediction (48.4%) and diagnosis (47.1%). The five most common algorithms include: support vector machine (20.9%), neural network (14.6%), random forest (10.5%), deep learning (9.2%), and decision tree (8.8%). The accuracy ranges of algorithms were 61.8 to 100% in neurocritical care, 61.6 to 100% in cardiovascular disease, and 54 to 100% in medical oncology. Only 20.1% of those algorithms had an explainability of AI, which provides the results of the solution that humans can understand.

Conclusion More AI algorithms are applied in CDSS and are important in improving clinical practice. Supervised learning still accounts for a majority of AI applications in internal medicine. This study identified four potential gaps: the need for AI explainability, the lack of ubiquity of CDSS, the narrow scope of target users of CDSS, and the need for AI in health care report standards.

Keywords

- ▶ clinical decision support systems
- ▶ artificial intelligence
- ▶ deep learning
- ▶ Machine learning
- ▶ medicine

received
October 28, 2020
accepted after revision
July 13, 2021

DOI <https://doi.org/10.1055/s-0041-1735470>.
ISSN 2566-9346.

© 2021. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution License, permitting unrestricted use, distribution, and reproduction so long as the original work is properly cited. (<https://creativecommons.org/licenses/by/4.0/>)
Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Background and Significance

Clinical Decision Support Systems

According to the Office of the National Coordinator for Health Information Technology, “clinical decision support (CDS) provides clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care.”¹ CDS can be used on a variety of tools and systems for clinical decision-making. Examples of CDS tools include alerts, reminders, clinical guidelines, recommendations, condition-specific order sets, data reports, documentation templates, diagnostic support, and databases.² CDS systems (CDSS) are computerized tools to help clinicians make clinical decisions and manage information.³ Examples of CDSS include automated laboratory alerting systems that help the user focus on key messages such as highlighting abnormal laboratory values,⁴ and pharmacy information systems that provide alerts for drug allergies or interactions.⁵ Advanced CDSS delivers more accurate information to clinicians, for instance, personalized drug dosage calculators, case-based recommendations, and suggestions for laboratory testing based on diseases. Because of the rapid growth of electronic health records (EHR), CDSS has been increasingly integrated in the EHR system and the existing workflow that the clinician can efficiently receive and act on system generated recommendations.⁶ To manage a large amount of clinical data and effectively transform health care systems, artificial intelligence (AI) and machine learning (ML) have been applied to computerized CDSS.⁷⁻⁹

Artificial Intelligence

AI was defined in 1955 by John McCarthy as “the science and engineering of making intelligent machines,” which has been designed to resolve complex challenges and hopefully someday will be as intelligent as humans.¹⁰ The first introduction of AI in health care was in the 1970s at Stanford University, California. They developed the MYCIN rule-based system to advise physicians regarding antimicrobial therapy. The MYCIN suggested possible pathogens and recommended a dosage of antibiotics based on body weight.^{11,12}

ML is a subset of AI defined as “the field of study that gives computers the ability to learn without being explicitly programmed” by Arthur Samuel.¹³ ML algorithms have four types: supervised learning, unsupervised learning, semisupervised learning, and reinforcement learning. Using data containing both inputs and target outcomes, supervised learning algorithms build a model. Conversely, unsupervised learning algorithms use data that contain only inputs to find the structure or pattern of the data. Semisupervised learning is an algorithm mixed between supervised and unsupervised learning algorithms to improve the accuracy of the model.¹⁴ Reinforcement learning does not require input/output pairs, and it focuses on a tradeoff between exploration and exploitation.¹⁵ ML models learn from training data to detect or predict outcomes with high accuracy. ML supports clinical work in prognosis, diagnosis, treatment, and clinical workflow.¹⁴ For example, ML was widely used in studies predict-

ing hospital readmission to reduce the payment for patients readmitted within 30 days of discharge. The most utilized algorithms in these studies were decision tree (DT)-based methods and support vector machine (SVM).¹⁶

Deep learning (DL) is a subset of ML that consists of layered sets of algorithms to progressively extract higher-level features from the raw input, inspired by neural networks (NN) of the human brain. The representation of one layer starting with the raw data input is fed and transformed into the next layer representation that enables learning highly complex functions.¹⁷ DL works very well at discovering complex structures in high-dimensional data in medicine. For example, DL was used to identify malignancy from pictures of skin lesions,¹⁸ detecting pneumonia from chest radiographs,^{19,20} and diagnosing diabetic retinopathy based on retinal photographs.²¹ These studies demonstrate that combining advanced computational methodologies with CDSS may reduce medical errors and improve care processes.^{6,22-24}

Explainable Artificial Intelligence

Explainable AI (XAI) was defined by Matt Turek from the Defense Advanced Research Projects Agency XAI program. Turek claims, “XAI proposes creating a suite of ML techniques that (1) produces more explainable models while maintaining a high level of learning performance (prediction accuracy) and (2) enables human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.²⁵” Many ML algorithms could not explain how and why a specific decision has been made. Thus, it raised the question: how can we make ML algorithms explainable? In 2018, the European Union General Data Protection Regulation discussed how to explain AI algorithms, and this discussion led to a debate among AI researchers regarding the “right to explanation.”²⁶ The right to explanation is a right to be given an explanation for the output of the algorithm. Because many AI algorithms, such as the output of the deep NN, are not easily explainable, XAI becomes more important and seeks to provide an explanation from AI algorithms. The explainability of AI could help to enhance the trust of AI-based systems from medical professionals.²⁷ Thus, AI-based CDSS requires not only good performance but also explainability that is trustworthy, transparent, and interpretable.²⁸

To analyze the explainability of AI-based CDSS, we can consider four perspectives from a multidisciplinary approach: technological, legal, medical, and patient perspectives.²⁹ The technological perspective considers the explainability of the model by characteristics of the algorithm. From the legal perspective, there are three issues needed to be considered for explainability: (1) informed consent, (2) certification and approval as medical devices from the Food and Drug Administration (FDA),³⁰ and (3) liability. Using unexplainable AI algorithms in CDSS for medical purposes has been controlled by the FDA; hence, it would affect the trend of using XAI and AI in the future. From the medical perspective, AI-based CDSS will be considered two levels of explainability: understanding the output

from the system and identifying feature importance. Last, from the patient perspective, explainability can provide personalized recommendations based on the patient's characteristics and risk factors. XAI-based CDSS could enhance patient engagement and provide an accurate risk perception.^{31,32}

AI in Internal Medicine

In medicine, AI is widely used to understand medical conditions, to predict diagnoses, to process extensive health data, and to aid physicians in making clinical decisions.³³ Examples of the current systems include IBM's Watson Health solutions³⁴ for the field of Clinical Medicine and MeVis medical solutions³⁵ for oncological radiology. Internal medicine is a medical specialty dealing with diagnosis, treatment, and prevention of adult diseases.³⁶ Internal medicine specialty has 20 subspecialties³⁷ and has the largest number of active physicians in the United States.³⁸

For the field of neurovascular disorders, Murray et al³⁹ reviewed the literature on acute stroke diagnostic-focused AI from 2014 to 2019 using the search terms: "artificial intelligence" or "machine learning or deep learning" and "ischemic stroke" or "large vessel occlusion." A total of 20 studies were identified, and the results show that random forest (RF) learning was used for the Alberta Stroke Program Early Computerized Tomography (CT) Score. In contrast, convolutional NN were used for detecting large vessel occlusions. The authors also identified platforms, including Brainomix, General Electric, iSchemaView, and Viz.ai. The authors suggested that AI improves stroke detection; however, the standardization of performance assessment is required.

For the field of cardiovascular diseases, Kilić⁴⁰ reviewed articles related to AI, ML, and cardiovascular health care that were published up to 2019. The author categorized ML algorithms into two major types, namely, supervised and unsupervised learning algorithms. Supervised learning algorithms include the Naïve Bayes theorem (NB), k-nearest neighbors, SVM, RF, extreme gradient boosting, and DT. Unsupervised learning algorithms include k-mean clustering, hierarchical clustering, principal component analysis, and singular value decomposition. The author summarized the potential application of ML in cardiovascular health care into three groups: (1) automated imaging interpretation, (2) natural language processing from EHR, and (3) predictive analytics. The author mentioned the challenges of implementing ML into clinical practice, including unexplainable results, privacy and ethical issues, validation and long-term evaluation, and the need for a large amount of data.

For an example of applied AI in the field of oncology, Jin et al⁴¹ conducted a systematic review on AI in gastric cancer using the search terms: "artificial intelligence" and "gastric cancer," and a total of 68 studies were included. The study reported that AI was used for omic data analyses, the identification of *Helicobacter pylori* infection and chronic atrophic gastritis, endoscopic diagnosis for gastric cancer, invasion depth prediction, digital pathology, bleeding detection, surgery (preoperative, intraoperative, and postoperative procedures), metastases and staging prediction, and prognosis prediction. The authors also grouped AI applica-

tions in gastric cancer, as mentioned above, into detection, treatment, and prognosis. The authors suggested that large randomized controlled trials (RCTs) are required to validate the AI models. However, it is difficult to conduct large RCTs in the rapidly changing environment of an EHR due to costs, interoperability, quality of data, and privacy and data security considerations.⁴²⁻⁴⁴

After reviewing several systematic reviews of AI in medicine, we concluded that AI applications in medicine could be grouped as prognosis/prediction, diagnosis/detection, treatment, and clinical workflow. Current ML implementation in clinical practice lacks the explainability of AI. Last, there is a need for the standardization to validate clinical performance of AI applications.

Objectives

There were many systematic review studies related to AI in Medicine. However, few studies reported the frequency and explainability of AI algorithms used in CDSS. We aimed to extract key information to identify a potential gap for further study.

In this study, we conducted a scoping review of literature in the past decade to analyze the implementation of applied AI in CDSS for subspecialties in internal medicine. Subspecialties in this study refer to the additional training to "subspecialize" in additional areas of internal medicine.³⁷ We aimed to answer three research questions (RQs), which are:

(RQ1) What is the frequency of applications regarding purposes of CDSS among prediction, diagnosis, treatment optimization, and clinical workflow optimization?

(RQ2) What is the frequency of applications regarding AI algorithms used in CDSS?

(RQ3) What is the overall accuracy of those algorithms?

Methods

Inclusion and Exclusion Criteria

Articles were included if they met the following criteria: (1) addressed CDSS using AI algorithms; (2) the AI algorithms studied include DL, ML, or automated pattern recognition; (3) they were related to the internal medicine specialty; (4) they were published between January 1, 2009 and December 31, 2019; (5) they were published in English; and (6) were original research.

We excluded articles using natural language or text processing that did not use AI algorithms. We also excluded articles proposing a new platform of CDSS without reporting results, technical reports of new algorithms without applications in medical research, and review papers.

Search Strategy

We searched three databases, including PubMed, IEEE Xplore, and Scopus, using the combination of search terms: "Clinical Decision Support Systems" AND ("Artificial Intelligence" OR "Deep Learning" OR "Machine Learning" OR "Automated Pattern Recognition") and limited results from January 1, 2009 to December 31, 2019. We included

automated pattern recognition in our search terms because pattern recognition is interchangeably used for ML.⁴⁵ We limited results from 2009 to 2019 because Meaningful Use introduced in 2009 in the United States promoted the electronic exchange of health information via certified EHR technology.^{46,47}

Study Selection

First, we reviewed the literature by screening the titles and abstracts and classified each paper as relevant, not relevant, or unclear. Second, the unclear category was revisited by reading the full-text and re-categorizing it as relevant or not relevant. Third, the full-text articles were read and key information was extracted. Those articles that met the inclusion criteria were included in the final set of articles. Last, we categorized all included articles into different internal medicine subspecialties including neurocritical care, cardiovascular disease, medical oncology, infectious disease, endocrinology, diabetes, and metabolism, critical care medicine, nephrology, gastroenterology, pulmonary disease, hematology, rheumatology, allergy and immunology, and geriatric medicine. We excluded articles related to other medical areas, including anesthesiology, dermatology, emergency medicine, obstetrics and gynecology, ophthalmology, orthopedic surgery, otolaryngology-head and neck surgery, pathology, pediatrics, physical medicine and rehabilitation, preventive medicine, psychiatry and neurocritical care, radiology, surgery, thoracic surgery, urology, orthodontics, and pharmacology from our review. Disagreements on inclusion, exclusion, and information extraction were resolved by consensus-based discussion among three authors (P.N., M.S.K. and S.A.B).

Data Extraction and Analysis

Key information was extracted from all articles by P.N. (→ **Appendix A**). The characteristics of articles included publication year, author, journal title, article title, study design (observational and experimental studies), purpose, decision, input data (a type of data, number of cases, and period of study), primary algorithms, comparison methods, balancing technique, explainability, accuracy, users, and ubiquity. The primary purpose of CDSS functions were categorized into four groups: prediction, diagnosis, treatment optimization, and clinical workflow optimization.¹⁴ XAI was determined from the included articles. If their methodology used an AI algorithm that maintained a high level of learning performance (prediction accuracy) and enabled human users to understand, appropriately trust, and effectively manage the emerging generation of AI partners,²⁵ we classified it as “explainable.” Otherwise, they were categorized as “unexplainable.” In → **Table 1**, P.N. and M.S.K. categorized those AI algorithms into four types: supervised ML, semisupervised ML, unsupervised ML, and DL.

Results

Identification of Eligible Articles

Our systematic searches identified 4,101 articles. There were 634 duplicate articles removed. The remaining 3,467 articles

Table 1 Categorization of AI algorithms

Types of AI algorithms	Algorithms
Supervised machine learning	<ul style="list-style-type: none"> • Support vector machine (SVM) • Decision tree (DT) • Various types of neural network (NN) • Regression • Random forest (RF) • Classifiers • k-Nearest neighbors (kNN) • Bayesian network (BN) • Naïve Bayes (NB) • Gradient boosting machine (GBM) • Fuzzy classifier • Genetic algorithm
Semisupervised machine learning	<ul style="list-style-type: none"> • Spectral regression Kernel discriminant analysis with semisupervised learning
Unsupervised machine learning	<ul style="list-style-type: none"> • Clustering • Text mining • Knowledge discovery • Rule-based reasoning (RBR)/case-based reasoning (CBR)/guideline-based • Various types of neural networks (NN)
Deep learning	<ul style="list-style-type: none"> • Deep NN (DNN) • Convolutional neural network (CNN) • Deep CNN • 3D CNN • Convolutional U-net with a two-dimensional gated recurrent NN (RNN) • Autoencoder network • CNN long-short-term memory (LSTM) • Hidden Markov Model, stacked denoising autoencoder, and Statistical Language Modeling

were screened using the inclusion criteria by titles, abstracts, and keywords. We excluded 1,973 articles based on exclusion criteria, which are articles proposing a new platform of CDSS without reporting results, technical reports of new algorithms without applications in medical research, and review papers. A full-text article assessment was conducted of 1,261 articles for eligibility. We removed 820 articles that were not related to the internal medicine specialty (→ **Fig. 1**). Out of 441 eligible articles, we considered the top three subspecialties composing 49.4% in internal medicine-related articles, which were neurocritical care ($n=89$), cardiovascular disease ($n=79$), and medical oncology ($n=50$) (→ **Table 2**). A total of 218 articles for these three subspecialties were further analyzed, and information was extracted to answer our RQ.

Purposes, AI Algorithms, Explainability of AI, and Target Users of CDSS

We used → **Table 3** to categorize CDSS into four groups based on the primary purpose of CDSS functions¹⁴: prediction,

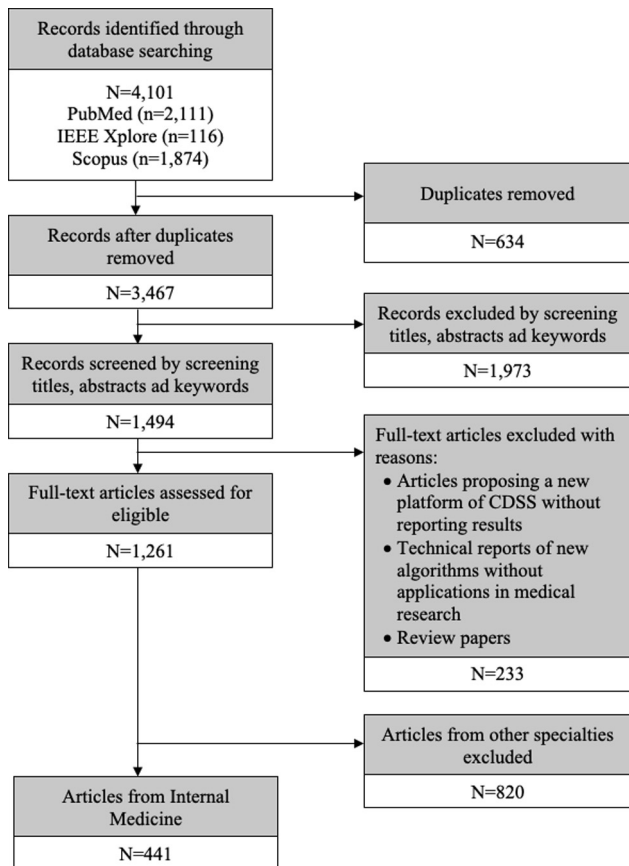


Fig. 1 Flow diagram of selecting studies for a scoping review.

diagnosis, treatment optimization, and clinical workflow optimization. The two most common purposes of CDSS were prediction (107, 48.4%) and diagnosis (104, 47.1%). In cardiovascular disease and medical oncology, CDSS was used for the prediction of 57.5 and 58.8%, respectively. Conversely,

Table 2 Number of studies in internal medicine subspecialties (total = 441)

Subspecialties (n = 13)	N
Neurocritical care	89
Cardiovascular disease	79
Medical oncology	50
Infectious disease	44
Endocrinology, diabetes, and metabolism	41
Critical care medicine	31
Nephrology	27
Gastroenterology	26
Pulmonary disease	24
Hematology	17
Rheumatology	5
Allergy and immunology	4
Geriatric medicine	4

CDSS in neurocritical care focused on diagnosis more than others (56, 62.2%).

For AI algorithms (→Table 3), generally, the top five common algorithms were SVM (50, 20.9%), NN (35, 14.6%), RF (25, 10.5%), DL (22, 9.2%), and DT (21, 8.8%). Regarding the neurocritical care subspecialty, the top five common algorithms were SVM (23, 23.5%), DL (12, 12.2%), RF (11, 11.2%), NN (10, 10.2%), and DT (10, 10.2%). Regarding the cardiovascular disease subspecialty, the top five common algorithms were SVM (16, 19.0%), NN (13, 15.5%), RF (11, 13.2%), others (9, 10.7%), and DT (8, 9.5%). The common algorithms of the medical oncology subspecialty were NN (12, 21.1%), SVM (11, 19.3%), DL (9, 15.8%), RBR (rule-based reasoning)/CBR (case-based reasoning)/guideline-based (5, 8.8%), Bayesian network (4, 7.0%), and NB (4, 7.0%).

AI algorithms applied to CDSS for subspecialties in medicine had a wide range covering supervised ML, semisupervised ML, unsupervised ML, and DL. Although DL is a part of ML, we separated DL into a specific category because we wanted to compare the prevalence of DL applications to other types of algorithms. Of the 18 AI algorithms in →Table 3, 85.8% were supervised ML, of which 79.5% of those algorithms were unexplainable AI. The majority of CDSS were developed for physician use (218, 96.9%), followed by patient use (4, 1.8%) and nurse use (3, 1.3%).

The trend of using AI algorithms has been changing over time, as shown in →Figs. 2 and 3. →Fig. 2 shows the number of types of AI algorithms by year, and →Fig. 3 shows the number of five common AI algorithms: SVM, NN, RF, DT, and DL by year. The use of ML in CDSS increased from two articles in 2009 to 42 articles in 2019. Specifically, DL algorithms, a newer technique, showed a sharp increase in published articles in 2018 (n = 8) and 2019 (n = 8). In the neurocritical care research area, DL has been used earlier than cardiovascular disease and medical oncology research areas for 2 years.

Answers to Research Questions

After synthesizing findings from 218 included articles, we attempted to answer our RQs as follows:

- *RQ1: What is the frequency of applications regarding purposes of CDSS among prediction, diagnosis, treatment optimization, and clinical workflow optimization?*

We grouped the purposes of CDSS into four categories: prediction, diagnosis, treatment optimization, and clinical workflow optimization. This review showed that the majority of CDSS were developed for prediction (48.4%) and diagnosis (47.1%) purposes.

- *RQ2: What is the frequency of applications regarding AI algorithms used in CDSS?*

There were wide ranges of AI algorithms used in medical research. After categorization, we found 18 different types of algorithms and the top five common algorithms among all subspecialties were SVM (20.9%), NN (14.6%), RF (10.5%), DL (9.2%), and DT (8.8%).

As the breadth of these data demonstrates, each model has its pros and cons and is potentially suited for different subspecialties (→Table 4). From our results, we found that

Table 3 Characteristics of the reviewed literature

	Total (n = 218)		Neurocritical care (n = 89)		Cardiovascular disease (n = 79)		Medical oncology (n = 50)	
	n	%	n	%	n	%	N	%
Purposes (n = 221)								
Prediction	107	48.4%	31	34.4%	46	57.5%	30	58.8%
Diagnosis	104	47.1%	56	62.2%	30	37.5%	18	35.3%
Treatment optimization	7	3.2%	3	3.3%	1	1.3%	3	5.9%
Clinical workflow optimization	3	1.4%	0	0.0%	3	3.8%	0	0.0%
Algorithms (n = 239)								
SVM	50	20.9%	23	23.5%	16	19.0%	11	19.3%
NN	35	14.6%	10	10.2%	13	15.5%	12	21.1%
RF	25	10.5%	11	11.2%	11	13.1%	3	5.3%
DL	22	9.2%	12	12.2%	1	1.2%	9	15.8%
DT	21	8.8%	10	10.2%	8	9.5%	3	5.3%
Others	17	7.1%	7	7.1%	9	10.7%	1	1.8%
Classifiers	10	4.2%	3	3.1%	6	7.1%	1	1.8%
RBR/CBR/guideline-based	10	4.2%	3	3.1%	2	2.4%	5	8.8%
kNN	9	3.8%	5	5.1%	3	3.6%	1	1.8%
BN	9	3.8%	2	2.0%	3	3.6%	4	7.0%
Regression	8	3.3%	5	5.1%	2	2.4%	1	1.8%
NB	8	3.3%	1	1.0%	3	3.6%	4	7.0%
Fuzzy	6	2.5%	3	3.1%	2	2.4%	1	1.8%
Clustering	4	1.7%	2	2.0%	2	2.4%	0	0.0%
GBM	2	0.8%	0	0.0%	1	1.2%	1	1.8%
Text mining	1	0.4%	1	1.0%	0	0.0%	0	0.0%
Genetic algorithm	1	0.4%	0	0.0%	1	1.2%	0	0.0%
Knowledge discovery	1	0.4%	0	0.0%	1	1.2%	0	0.0%
Types of AI (n = 239)								
Supervised ML	206	86.2%	78	79.6%	76	91.6%	52	91.2%
DL	16	6.7%	12	12.2%	1	1.2%	3	5.3%
Unsupervised ML	16	6.7%	7	7.1%	7	8.4%	2	3.5%
Semisupervised ML	1	0.4%	1	1.0%	0	0.0%	0	0.0%
Explainable AI								
Unexplainable	174	79.5%	69	77.5%	69	87.3%	36	72.0%
Explainable	44	20.1%	20	22.5%	10	12.7%	14	28.0%
Users (n = 225)								
Physician	218	96.9%	89	97.8%	79	95.2%	50	98.0%
Patient	4	1.8%	2	2.2%	1	1.2%	1	2.0%
Nurse	3	1.3%	0	0.0%	3	3.6%	0	0.0%

Abbreviations: AI, artificial intelligence; BN, Bayesian network; CBR, case-based reasoning; DL, deep learning; DT, decision tree; GBM, gradient boosting machine; kNN, k-nearest neighbors; ML, machine learning; NB, naïve Bayes theorem; NN, neural network; RBR, rule-based reasoning; RF, random forest; SVM, support vector machine.

SVM and NN were common among those three subspecialties. The reason could be that SVM can handle multiple-class classification and small datasets. Moreover, SVM and NN are easier to use for prediction or classification

and more stable than DT. However, the results from SVM and NN can be hard to explain. We also found that DL is more prevalent in neurocritical care and medical oncology than in cardiovascular disease.

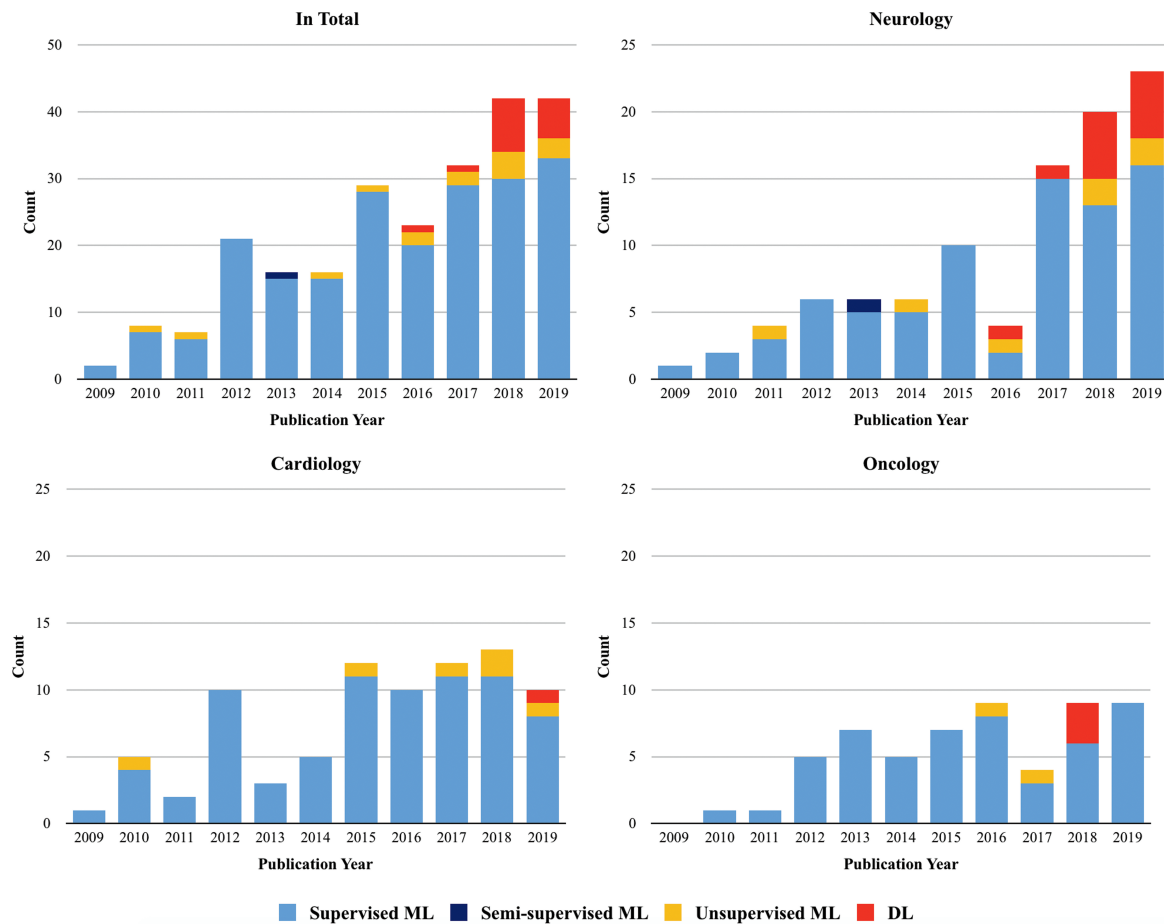


Fig. 2 The number of types of AI algorithms by year. AI, artificial intelligence.

After further examination of the data modalities used in the original studies, we found that, in neurocritical care, several frequently applied data types are suitable for using DL, such as intracranial electroencephalogram,⁴⁸ facial video clips,⁴⁹ electroencephalogram,^{50–58} and magnetic resonance imaging (MRI).^{59–70} Similarly, in medical oncology, the DL method is mostly applied to the image data.^{71,72} This is reasonable as images are used more in diagnosis in these two subspecialties than cardiovascular disease, and DL is suited to the analysis of image data, such as MRI, CT, positron emission tomography scans, and ultrasound images.

- RQ3: What is the overall accuracy of those algorithms?**
 Accuracy is the percentage of correct predictions for the input data and is calculated by the number of correct predictions divided by the total number of predictions made. In a simple way, accuracy is the percentage of our model got right.⁷³ The accuracy of CDSS should be tested because inaccurate recommendations can endanger the safety or well-being of patients.⁹ It is challenging to report the average accuracy of AI algorithms because various metrics have been used to measure accuracy in these articles. For the articles reporting accuracy scores, we found that the accuracy ranges of AI algorithms in neurocritical care, cardiovascular disease, medical oncology were 61.8 to 100%, 61.6 to 100%, and 54 to 100%, respectively.

Because of the inconsistency in reporting results of individual articles, it is particularly challenging to synthesize and report the results from included articles. To address this issue, Hernandez-Boussard et al⁷⁴ presented MINIMUM Information for Medical AI Reporting or MINIMAR to standardize the report on AI in health care. The standard report should satisfy four essential requirements: (1) study population and setting, (2) patient demographic characteristics, (3) model architecture, and (4) model evaluation. The study population and setting include population, study setting, data source, and cohort selection. The patient demographic characteristics are age, sex, race, ethnicity, and socioeconomic status. For the model architecture, researchers should report model output, target user, data splitting, gold standard, model task, model architecture, features, and missingness. The report should include optimization, internal model validation, external validation, and transparency for the model evaluation. This standard would help provide an accurate and responsible report on AI in health care.

Discussion

We conducted a scoping review to find evidence of applied AI algorithms in CDSS for internal medicine subspecialties. Accordingly, our study found that neurocritical care,

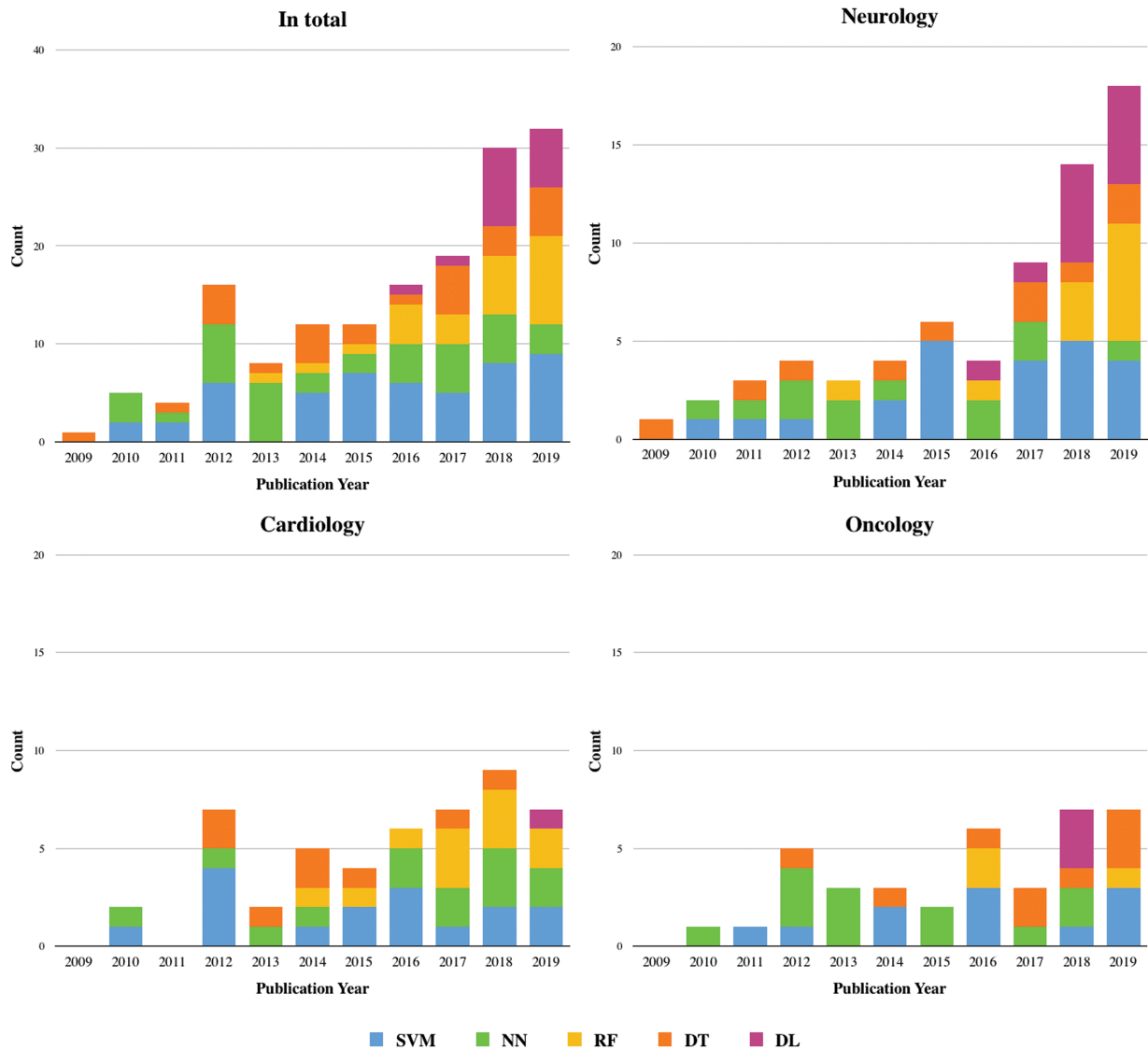


Fig. 3 The number of AI algorithms by year. AI, artificial intelligence.

Table 4 Advantages and disadvantages of AI algorithms

Algorithms	Advantages	Disadvantages
Support vector machine ¹⁰²⁻¹⁰⁷	<ul style="list-style-type: none"> • Binary or multiple-class classification • Do not need a large dataset 	<ul style="list-style-type: none"> • Difficult to pick kernel function/ parameters • Difficult to explain • Long-time training on a large dataset
Neural network ¹⁰⁸⁻¹¹⁰	<ul style="list-style-type: none"> • Able to do complex classification 	<ul style="list-style-type: none"> • Difficult to explain • Difficult to tune parameters
Random forest ¹¹¹⁻¹¹⁴	<ul style="list-style-type: none"> • Able to deal with missing values • Able to deal with high-dimension large dataset 	<ul style="list-style-type: none"> • Difficult to explain • Prone to have overfitting problem
Decision tree ¹¹⁵⁻¹¹⁷	<ul style="list-style-type: none"> • Able to be explained • Able to deal with complex data 	<ul style="list-style-type: none"> • Unstable • Accuracy not high
Deep learning ^{17,118-120}	<ul style="list-style-type: none"> • Able to do complex classification • Could be expensive (graphics processing unit needed) 	<ul style="list-style-type: none"> • Difficult to explain • Need large dataset

cardiovascular disease, and medical oncology subspecialties had the high volume of applied AI algorithms in CDSS studies, approximately 49%. According to the World Health Organization and the Centers for Disease Control and Prevention, cardiovascular diseases were the first leading cause of death, and neurological disorders were the second leading cause of death worldwide.^{75–77} Cancer is a major health problem worldwide and was the second leading cause of death in the United States in 2019.⁷⁸ This review provided significant value regarding CDSS using AI algorithms in internal medicine and globally major health problems. The volume of applied AI algorithms to solve medical problems has continuously increased from 2009 to 2019, with a substantial change in 2018 and 2019. We also observed a significantly growing number of articles involving DL from 2016 to 2019.

Explainability of AI Algorithms

This review shows that most articles have used unexplainable algorithms (79.5%). The use of unexplainable AI models has been debated and discussed in many articles, with an ongoing controversy in current medical practices. We believe that in the future, researchers should move forward applying XAI algorithms, which are AI algorithms that provide results that are understood by human experts.²⁵

AI explainability is examined primarily from a clinical point of view, highlighting the ability of humans to understand which clinical characteristics drive the prediction. This is important, as the main objective of clinical predictive modeling is the development of CDSS, assisting health professionals in their clinical decision-making, predicting diagnoses, risks, and results.^{27,79} It is important to keep in mind that the requirements for CDSS go far beyond the performance of the model.⁸⁰ It is established that CDSS for the clinical environment needs to exhibit proven safety and accuracy.⁸⁰ The explainability of AI systems is crucial to understand why they do what they do, but more importantly, to understand why and when they may not do what is planned. This transparency is important in light of the growing awareness of potential biases in the models used for health discrimination. An XAI system is essential to provide: a safe interpretation and verification of the results acquired during development; better evaluation of the safety and justice of medical products, especially concerning bias, during the regulatory process; interpretation supported by domain knowledge leading to increased confidence on the part of doctors, other health professionals, and patients. The explainability of AI can help to increase the confidence of medical professionals in future AI systems.

Ubiquity and Usability

We identified information on developed CDSS for ubiquity, i.e., if the CDSS are made to appear anytime and everywhere. Some articles had developed ubiquity, such as software, Web-based tools, and mobile apps. The ubiquity includes neuroQWERTY platform,⁸¹ Heart Failure Manager tool,⁸² Chest Pain Rule Out (CPRO) Calculator,⁸³ the HEARTFAID platform,⁸⁴ PaDEL-Survival,⁸⁵ OncoMortality,⁸⁶ PrediWeb,⁸⁷

and The-Optimal-Lymph-Flow (TOLF).⁸⁸ Most of the included articles did not report about model applications.

In a CDSS, the outcome of the system can be related to the user interface directly. A successful CDSS should offer an efficient user interface to clinicians to get the most proper consultation results. Miller et al⁸⁹ described simplification as including only the elements that are most important for communication. Use of consistent terminology, concise and unambiguous language, and effective visualization improved usability and reduced information density. To improve usability, it is suggested to consider using appropriate font sizes, using meaningful colors, ensuring acceptable contrast between the text and background, and making the icons bold or larger. Space-filling techniques help to maximize the amount of information that can be displayed in the available display space. Visibility factors consider human factors and cognitive computing. A user-centered design process also should be considered during the CDSS development. The user-centered design aims to create the system based on user characteristics using interdisciplinary approaches of cognitive science, psychology, and computer science.^{90,91} The user-centered design helps identify the potential deficiencies of CDSS, such as substantial variability in the usability, efficacy, and safety of CDSS.^{92–94}

Study Limitations

Our study has several limitations. First, we conducted a scoping review, which did not require an assessment of methodological limitations or risk of bias of the evidence⁹⁵; however, we collected study design (→ **Appendix A**), which can provide the level of evidence of individual studies. Second, we excluded non-English papers, which may constitute a selection bias. Last, we limited the year of publications based on EHR implementation in the United States and associated applications of AI-based CDSS, which may lead to publication bias. However, we believe that the findings of our review were able to answer our RQ.

Conclusion

With the continued advancement of medical techniques and devices, the size, variety, and complexity of data also continue to increase. Many ML and data mining methods have been used in the medical field to help with disease diagnosis, prediction, and treatment optimization. This demonstrates that AI can provide more accurate diagnostic results. We identified four potential research gaps to fill in from this study. First, we found that only 44 articles (20.1%) of the included articles have used XAI algorithms resulting in distrust from clinicians because of the lack of effectiveness and learning performance. We suggest future CDSS should increase the utilization of XAI algorithms, which can help to enhance trust and confidence in using the CDSS among clinicians. Second, we found that there was a lack of ubiquity among the reviewed articles. The CDSS should be available for users anytime and anywhere to make clinical decisions at the point of care; however, only 21 articles (9.6%) developed platforms (i.e., software, web-based tools, and mobile apps)

that clinicians and patients can access. Most of the articles did not report the platform development or implementation. We suggest future CDSS should consider not only the model performance but also ubiquity improvement. The ubiquity will increase accessibility for clinicians and patients and lead to opportune use of CDSS in clinical practice. Third, the majority of CDSS were developed for physician users (96.9%). Developers should consider expanding the scope of target users and enhancing engagement in shared decision-making among health care providers and patients to achieve the delivery of patient-centered care. Last, we observed a lack of standardized reporting structure in AI-based CDSS that resulted in inconsistent data extraction. The reviewed articles did not follow the MINIMAR standards when they reported information and failed to provide an accurate, unbiased, and meaningful report. We suggest future articles related to AI in health care should report information following the MINIMAR standards.

Although there are many studies showing the success of using CDSS in health care management, implementation is a significant challenge because of unreliability and inability to exchange EHR data between systems, unfriendly user interfaces, limited choices of implementation and workflow, and technical issues.^{96,97} Moreover, in the real world, EHR data can be inaccurate, unreliable, transformed, and insufficient.^{98–101} Hence, the quality of data is an important challenge for applied AI in medicine.

Clinical Relevance Statement

This scoping review showed the trends of utilizing AI algorithms in CDSS for subspecialties in internal medicine between 2009 and 2019. The most frequent numbers of articles related to CDSS using AI algorithms among internal medicine subspecialties were neurocritical care, cardiovascular disease, and medical oncology. This review showed a substantial change in utilizing DL in published articles in 2018 and 2019. This review indicated four potential gaps of CDSS development: the need for AI explainability, the lack of ubiquity of CDSS, the narrow scope of target users, and the need for AI in health care report standards.

Appendix A: Characteristics of the reviewed literature

Author Contributions

S.A.B. and M.S.K. contributed the study design, critical revision of this article, and final approval of the version to be published. P.N. searched the literature, synthesized included studies, and drafted the article.

Funding

None.

Conflict of interest

None declared.

References

- 1 HealthIT.gov. Clinical decision support. Accessed August 23, 2020 at: <https://www.healthit.gov/topic/safety/clinical-decision-support>
- 2 Agency for Healthcare Research and Quality. Clinical decision support. Published June. 2019. Accessed February 11, 2021 at: <http://www.ahrq.gov/cpi/about/otherwebsites/clinical-decision-support/index.html>
- 3 Alther M, Reddy CK. *Clinical Decision Support Systems*. Boca Raton, FL: CRC Press; 2015
- 4 Tate KE, Gardner RM, Weaver LK. A computerized laboratory alerting system. *MD Comput* 1990;7(05):296–301
- 5 Alanazi A, Al Rabiah F, Gadi H, Househ M, Al Dosari B. Factors influencing pharmacists' intentions to use Pharmacy Information Systems. *Inform Med Unlocked*. 2018;11:1–8
- 6 Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005;330(7494):765
- 7 Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health* 2018;8(02):020303
- 8 Kim E, Rubinstein SM, Nead KT, Wojcieszynski AP, Gabriel PE, Warner JL. evolving use of electronic health records (EHR) for research. *Semin Radiat Oncol* 2019;29(04):354–361
- 9 Mahadevaiah G, Rv P, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: Selection, acceptance, commissioning, and quality assurance. *Med Phys* 2020;47(05):e228–e235
- 10 Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism* 2017;69S:S36–S40
- 11 Kulikowski CA. An opening chapter of the first generation of artificial intelligence in medicine: the First Rutgers AIM Workshop, June 1975. *Yearb Med Inform* 2015;10(01):227–233
- 12 Shortliffe EH, Axline SG, Buchanan BG, Merigan TC, Cohen SN. An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Comput Biomed Res* 1973;6(06):544–560
- 13 Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Develop* 1959;3(03):210–229
- 14 Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019;380(14):1347–1358
- 15 Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. *J Artif Intell Res* 1996;4:237–285
- 16 Artetxe A, Beristain A, Graña M. Predictive models for hospital readmission risk: A systematic review of methods. *Comput Methods Programs Biomed* 2018;164:49–64
- 17 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–444
- 18 Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Invest Dermatol* 2018;138(07):1529–1538
- 19 Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheX-NeXt algorithm to practicing radiologists. *PLoS Med* 2018;15(11):e1002686
- 20 Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJPC. Identifying pneumonia in chest X-rays: a deep learning approach. *Measurement* 2019;145:511–518
- 21 Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316(22):2402–2410
- 22 Castaneda C, Nalley K, Mannion C, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *J Clin Bioinforma* 2015;5:4
- 23 Moja L, Kwag KH, Lytras T, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. *Am J Public Health* 2014;104(12):e12–e22

- 24 Nuckols TK, Smith-Spangler C, Morton SC, et al. The effectiveness of computerized order entry at reducing preventable adverse drug events and medication errors in hospital settings: a systematic review and meta-analysis. *Syst Rev* 2014;3:56
- 25 Turek M. Explainable artificial intelligence. Accessed October 30, 2019 at: <https://www.darpa.mil/program/explainable-artificial-intelligence>
- 26 Edwards L, Veale M. Slave to the algorithm? Why a 'Right to an Explanation' is probably not the remedy you are looking for. *Social Science Research Network* 2017. Accessed February 27, 2020 at: <https://papers.ssrn.com/abstract=2972855>
- 27 Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 2019;9(04):e1312
- 28 Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? *ArXiv171209923 Cs Stat*. Published online December 28, 2017. Accessed March 2, 2021 at: <http://arxiv.org/abs/1712.09923>
- 29 Amann J, Blasimme A, Vayena E, Frey D, Madai VIPrecise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* 2020;20(01):310
- 30 FDA. Artificial intelligence and machine learning in software as a medical device. FDA. Published online January 11, 2021. Accessed February 16, 2021 at: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>
- 31 Politi MC, Dizon DS, Frosch DL, Kuzemchak MD, Stiggelbout AM. Importance of clarifying patients' desired role in shared decision making to match their level of engagement with their preferences. *BMJ* 2013;347(dec02 1):f7066-f7066
- 32 Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* 2017;4:CD001431
- 33 Stoeklé HC, Charlier P, Hervé C, Deleuze JF, Vogt G. Artificial intelligence in internal medicine: between science and pseudo-science. *Eur J Intern Med* 2018;51:e33-e34
- 34 IBM Watson Health | AI healthcare solutions. IBM Watson Health. Published February 19, 2021. Accessed March 2, 2021 at: <https://www.ibm.com/watson-health>
- 35 MeVis medical solutions AG. Accessed March 2, 2021 at: <https://www.mevis.de/en>
- 36 ACP. What is a doctor of internal medicine, or internist? Accessed March 9, 2021 at: <https://www.acponline.org/acp-newsroom/what-is-a-doctor-of-internal-medicine-or-internist-0>
- 37 ACP. Internal medicine subspecialties career information. Accessed October 21, 2020 at: <https://www.acponline.org/about-acp/about-internal-medicine/subspecialties>
- 38 AAMC. Physician specialty data report, 2019. Accessed March 9, 2021 at: <https://www.aamc.org/data-reports/workforce/interactive-data/active-physicians-largest-specialties-2019>
- 39 Murray NM, Unberath M, Hager GD, Hui FK. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J Neurointerv Surg* 2020;12(02):156-164
- 40 Kilic A. Artificial intelligence and machine learning in cardiovascular health care. *Ann Thorac Surg* 2020;109(05):1323-1329
- 41 Jin P, Ji X, Kang W, et al. Artificial intelligence in gastric cancer: a systematic review. *J Cancer Res Clin Oncol* 2020;146(09):2339-2350
- 42 Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *Br J Cancer* 2014;110(03):551-555
- 43 Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. *Am J Prev Med* 2007;33(02):155-161
- 44 Mc Cord KA, Hemkens LG. Using electronic health records for clinical trials: Where do we stand and where can we go? *CMAJ* 2019;191(05):E128-E133
- 45 Bishop CM. *Pattern Recognition and Machine Learning*. New York, NY: Springer; 2006
- 46 CDC. Introduction | Meaningful use. Published September 17, 2020. Accessed February 16, 2021 at: <https://www.cdc.gov/ehrmeaningfuluse/introduction.html>
- 47 Murphy EV. Clinical decision support: effectiveness in improving quality processes and clinical outcomes and factors that may influence success. *Yale J Biol Med* 2014;87(02):187-197
- 48 Wang F, Sun M, Min T, et al. Analysis for early seizure detection system based on deep learning algorithm. In: Schmidt H, Griol DZL, Wang H, Baumbach J, Zheng H, Callejas Z, Hu X, Dickerson J, eds. *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*. Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc.; 2019:2382-2389
- 49 Ahmedt-Aristizabal D, Fookes C, Nguyen K, Denman S, Sridharan S, Dionisio S. Deep facial analysis: a new phase I epilepsy evaluation using computer vision. *Epilepsy Behav* 2018;82:17-24
- 50 Abibullaev B, Kim MS, Seo HD. Seizure detection in temporal lobe epileptic EEGs using the best basis wavelet functions. *J Med Syst* 2010;34(04):755-765
- 51 Kharat PA, Dudul SV. Clinical decision support system based on Jordan/Elman neural networks. In: *2011 IEEE Recent Advances in Intelligent Computational Systems, RAICS 2011*. Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc.; 2011:255-259
- 52 Kharat PA, Dudul SV. Epilepsy diagnosis based on generalized feed forward neural network. *Interdiscip Sci* 2012;4(03):209-214
- 53 Trambaioli LR, Spolađr N, Lorena AC, Anghinah R, Sato JR. Feature selection before EEG classification supports the diagnosis of Alzheimer's disease. *Clin Neurophysiol* 2017;128(10):2058-2067
- 54 Sun M, Wang F, Min T, Zang T, Wang Y. Prediction for high risk clinical symptoms of epilepsy based on deep learning algorithm. *IEEE Access* 2018;6:77596-77605
- 55 Golmohammadi M, Harati Nejad Torbati AH, Lopez de Diego S, Obeid I, Picone J. Automatic analysis of eegs using big data and hybrid deep learning architectures. *Front Hum Neurosci* 2019;13:76
- 56 Subasi A, Ahmed A, Aličković E, Rashik Hassan A. Effect of photic stimulation for migraine detection using random forest and discrete wavelet transform. *Biomed Signal Process Control* 2019;49:231-239
- 57 Rodríguez Aldana Y, Marañoń Reyes EJ, Macias FS, et al. Non-convulsive epileptic seizure monitoring with incremental learning. *Comput Biol Med* 2019;114:103434
- 58 Chiang HS, Chen MY, Huang YJ. Wavelet-based EEG processing for epilepsy detection using fuzzy entropy and associative petri net. *IEEE Access* 2019;7:103255-103262
- 59 Abdulkadir A, Mortamet B, Vemuri P, Jack CR Jr, Krueger G, Klöppel S Alzheimer's Disease Neuroimaging Initiative. Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. *Neuroimage* 2011;58(03):785-792
- 60 Zhang X, Hu B, Ma X, Moore P, Chen J. Ontology driven decision support for the diagnosis of mild cognitive impairment. *Comput Methods Programs Biomed* 2014;113(03):781-791
- 61 Munsell BC, Wee CY, Keller SS, et al. Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. *Neuroimage* 2015;118:219-230
- 62 Pillai PS, Leong TY Alzheimer's Disease Neuroimaging Initiative. Fusing heterogeneous data for Alzheimer's disease classification. *Stud Health Technol Inform* 2015;216:731-735
- 63 Wu X, Zou Q, Hu J, et al. Intrinsic functional connectivity patterns predict consciousness level and recovery outcome in acquired brain injury. *J Neurosci* 2015;35(37):12932-12946

- 64 Hu C, Ju R, Shen Y, Zhou P, Li Q. Clinical decision support for Alzheimer's disease based on deep learning and brain network. In: 2016 IEEE International Conference on Communications, ICC 2016. Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc.; 2016
- 65 Zhang J, Fan Y, Li Q, Thompson PM, Ye J, Wang Y. Empowering cortical thickness measures in clinical diagnosis of Alzheimer's disease with spherical sparse coding. *Proc IEEE Int Symp Biomed Imaging* 2017;2017:446–450
- 66 Yang H, Zhang J, Liu Q, Wang Y. Multimodal MRI-based classification of migraine: using deep learning convolutional neural network. *Biomed Eng Online* 2018;17(01):138
- 67 Nielsen A, Hansen MB, Tietze A, Mouridsen K. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. *Stroke* 2018;49(06):1394–1401
- 68 Pinto A, Mckinley R, Alves V, Wiest R, Silva CA, Reyes M. Stroke lesion outcome prediction based on MRI imaging combined with clinical information. *Front Neurol* 2018;9:1060
- 69 Ju R, Hu C, Zhou P, Li Q. Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Trans Comput Biol Bioinformatics* 2019;16(01):244–257
- 70 Law MT, Traboulsee AL, Li DK, et al. Machine learning in secondary progressive multiple sclerosis: an improved predictive model for short-term disability progression. *Mult Scler J Exp Transl Clin* 2019;5(04):2055217319885983
- 71 Li Y, Charalampaki P, Liu Y, Yang GZ, Giannarou S. Context aware decision support in neurosurgical oncology based on an efficient classification of endomicroscopic data. *Int J CARS* 2018;13(08):1187–1199
- 72 Cha YJ, Jang WI, Kim MS, et al. Prediction of response to stereotactic radiosurgery for brain metastases using convolutional neural networks. *Anticancer Res* 2018;38(09):5437–5445
- 73 Mitchell TM. *Machine Learning*. New York, NY: McGraw-Hill; 1997
- 74 Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc* 2020;27(12):2011–2015
- 75 GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 2019;18(05):459–480
- 76 WHO. Cardiovascular diseases (CVDs). Accessed March 7, 2021 at: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- 77 WHO. The top 10 causes of death. Accessed March 7, 2021 at: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- 78 Kochanek KD, Xu JQ, Arias EMortality in the United States, 2019. National Center for Health Statistics. 2020. Accessed March 7, 2021 at: <https://www.cdc.gov/nchs/products/databriefs/db395.htm>
- 79 Khamparia A, Singh KM. A systematic review on deep learning architectures and applications. *Expert Syst* 2019;36(03):e12400
- 80 Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019;28(03):231–237
- 81 Arroyo-Gallego T, Ledesma-Carbayo MJ, Butterworth I, et al. Detecting motor impairment in early Parkinson's disease via natural typing interaction with keyboards: validation of the neuroQWERTY approach in an uncontrolled at-home setting. *J Med Internet Res* 2018;20(03):e89
- 82 Guidi G, Pettenati MC, Melillo P, Iadanza E. A machine learning system to improve heart failure patient assistance. *IEEE J Biomed Health Inform* 2014;18(06):1750–1756
- 83 Isma'eel HA, Cremer PC, Khalaf S, et al. Artificial neural network modeling enhances risk stratification and can reduce downstream testing for patients with suspected acute coronary syndromes, negative cardiac biomarkers, and normal ECGs. *Int J Cardiovasc Imaging* 2016;32(04):687–696
- 84 Costanzo D. Biomedical data acquisition and processing in the decision support services of HEARTFAID platform. In: Proceedings of the 5th IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS'2009. Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc.; 2009:291–296
- 85 Ng T, Chew L, Yap CW. A clinical decision support tool to predict survival in cancer patients beyond 120 days after palliative chemotherapy. *J Palliat Med* 2012;15(08):863–869
- 86 Bertsimas D, Dunn J, Pawlowski C, et al. Applied informatics decision support tool for mortality predictions in patients with cancer. *JCO Clin Cancer Inform* 2018;2:1–11
- 87 De Bari B, Vallati M, Gatta R, et al. Development and validation of a machine learning-based predictive model to improve the prediction of inguinal status of anal cancer patients: a preliminary report. *Oncotarget* 2016;8(65):108509–108521
- 88 Fu MR, Axelrod D, Guth AA, et al. mHealth self-care interventions: managing symptoms following breast cancer treatment. *mHealth* 2016;2:28–28
- 89 Miller K, Mosby D, Capan M, et al. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. *J Am Med Inform Assoc* 2018;25(05):585–592
- 90 Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. *J Biomed Inform* 2005;38(01):75–87
- 91 Brunner J, Chuang E, Goldzweig C, Cain CL, Sugar C, Yano EM. User-centered design to improve clinical decision support in primary care. *Int J Med Inform* 2017;104:56–64
- 92 Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. *J Biomed Inform* 2004;37(01):56–76
- 93 Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. *Ann Intern Med* 2012;157(01):29–43
- 94 Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. *Ann Intern Med* 2014;160(01):48–54
- 95 Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *Int J Evid-Based Healthc* 2015;13(03):141–146
- 96 Tcheng JE, Bakken S, Bates DW, et al. (eds) *Optimizing Strategies for Clinical Decision Support: Summary of a Meeting Series*. Washington, DC: National Academy of Medicine; 2017
- 97 Sittig DF, Wright A, Osheroff JA, et al. Grand challenges in clinical decision support. *J Biomed Inform* 2008;41(02):387–392
- 98 Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care* 2013;51(08, Suppl 3):S30–S37
- 99 Bhise V, Rajan SS, Sittig DF, et al. Electronic health record reviews to measure diagnostic uncertainty in primary care. *J Eval Clin Pract* 2018;24(03):545–551
- 100 Dexheimer JW, Taylor RG, Kachelmeyer AM, Reed JL. The reliability of computerized physician order entry data for research studies. *Pediatr Emerg Care* 2019;35(03):e61–e64
- 101 Murphy DR, Satterly T, Rogith D, Sittig DF, Singh H. Barriers and facilitators impacting reliability of the electronic health record-facilitated total testing process. *Int J Med Inform* 2019;127:102–108
- 102 Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V. Feature selection for SVMs. *Adv Neural Inform Process Syst* 13 (NIPS 2000) 2000;13:668–674
- 103 Lorena AC, de Carvalho ACPLF. Evolutionary tuning of SVM parameter values in multiclass problems. *Neurocomputing* 2008;71(16–18):3326–3334

- 104 Yu H, Yang J, Han J. Classifying large data sets using SVMs with hierarchical clusters. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '03. New York, NY: ACM Press; 2003:306
- 105 Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter MA, Kagal L. Explaining Explanations: An Overview of Interpretability of Machine Learning. 2018 IEEE 5th International Conference on Data Science and Advanced Analytics DSAA 2018:80–89
- 106 Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995; 20(03):273–297
- 107 Hsu CW, Lin CJ. A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw* 2002;13(02):415–425
- 108 Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996;49(11):1225–1231
- 109 Livingstone DJ, Manallack DT, Tetko IV. Data modelling with neural networks: advantages and limitations. *J Comput Aided Mol Des* 1997;11(02):135–142
- 110 Specht DF. A general regression neural network. *IEEE Trans Neural Netw* 1991;2(06):568–576
- 111 Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 2018;19(01):270
- 112 Li B, Chen X, Li MJ, Huang JZ, Feng S. Scalable random forests for massive data. In: Tan P-N, Chawla S, Ho CK, Bailey J, eds. *Advances in Knowledge Discovery and Data Mining*. Vol 7301. Lecture Notes in Computer Science Berlin, Heidelberg: Springer; 2012:135–146
- 113 Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 1998;20(08):832–844
- 114 Breiman L. Random forests. *Mach Learn* 2001;45(01):5–32
- 115 Kamiński B, Jakubczyk M, Szufel P. A framework for sensitivity analysis of decision trees. *Cent Eur J Oper Res* 2018;26(01):135–159
- 116 Kotsiantis SB. Decision trees: a recent overview. *Artif Intell Rev* 2013;39(04):261–283
- 117 Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1(01):81–106
- 118 Samek W, Binder A, Montavon G, Lapuschkin S, Müller K-R. Evaluating the visualization of what a deep neural network has learned. *IEEE Trans Neural Netw Learn Syst* 2017;28(11):2660–2673
- 119 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *ArXiv14090473 Cs Stat*. Published online May 19, 2016. Accessed October 31, 2019 at: <https://arxiv.org/abs/1409.0473>
- 120 Donahue J, Hendricks LA, Rohrbach M, et al. Long-term recurrent convolutional networks for visual recognition and description. *IEEE Trans Pattern Anal Mach Intell* 2017;39(04):677–691