Review

Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine

Zeeshan Ahmed[1,2,3,4,*], Khalid Mohamed[3], Saman Zeeshan[5] and XinQi Dong[1,2]

[1]Institute for Health, Health Care Policy and Aging Research, Rutgers, The State University of New Jersey, 112 Paterson Street, New Brunswick, NJ, USA, [2]Department of Medicine, Rutgers Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, 125 Paterson Street, New Brunswick, NJ, USA, [3]Department of Genetics and Genome Sciences, School of Medicine, University of Connecticut Health Center, 263 Farmington Ave., Farmington, CT, USA, [4]Institute for Systems Genomics, University of Connecticut, 67 North Eagleville Road, Storrs, CT, USA and [5]The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, USA

*Corresponding author: Email: zahmed@ifh.rutgers.edu

Abstract

Precision medicine is one of the recent and powerful developments in medical care, which has the potential to improve the traditional symptom-driven practice of medicine, allowing earlier interventions using advanced diagnostics and tailoring better and economically personalized treatments. Identifying the best pathway to personalized and population medicine involves the ability to analyze comprehensive patient information together with broader aspects to monitor and distinguish between sick and relatively healthy people, which will lead to a better understanding of biological indicators that can signal shifts in health. While the complexities of disease at the individual level have made it difficult to utilize healthcare information in clinical decision-making, some of the existing constraints have been greatly minimized by technological advancements. To implement effective precision medicine with enhanced ability to positively impact patient outcomes and provide real-time decision support, it is important to harness the power of electronic health records by integrating disparate data sources and discovering patient-specific patterns of disease progression. Useful analytic tools, technologies, databases, and approaches are required to augment networking and interoperability of clinical, laboratory and public health systems, as well as addressing ethical and social issues related to the privacy and protection of healthcare data with effective balance. Developing multifunctional machine learning platforms for clinical data extraction, aggregation, management and analysis can support clinicians by efficiently stratifying

subjects to understand specific scenarios and optimize decision-making. Implementation of artificial intelligence in healthcare is a compelling vision that has the potential in leading to the significant improvements for achieving the goals of providing real-time, better personalized and population medicine at lower costs. In this study, we focused on analyzing and discussing various published artificial intelligence and machine learning solutions, approaches and perspectives, aiming to advance academic solutions in paving the way for a new data-centric era of discovery in healthcare.

Introduction

Over the centuries, quests for answers have led us to take giant leaps. It was only in the last century that the discovery of antibiotics freed us from many of the dreaded diseases of the past. Still, in the context of recent published literature (e.g. accessible through PubMed), over 138 000 studies discuss medication errors, and over 450 000 include delayed treatment (date accessed 10 October 2019, using search query keywords including 'medication error' and 'delayed treatment'). Still, the problem of people dying from medical care gone wrong has been vastly underappreciated and not well recognized. Today, we stand on the threshold of the new medical revolution, just as big and far-reaching. Despite all of our scientific knowledge, much of medicine is still based on the treatment of symptoms and performing learned trials based on treatments, which works for most patients to bring symptom relief, reduce the risk of complications and improve survival chances, but not for all. To get new insights into disease taxonomy, etiology and pathogenesis, it is important to understand how diseases are related to each other. Breakthroughs in prescription medication, surgical treatment and mental health interventions are among the reasons we live longer. However, providing the correct in-time treatment plan for patients with knowledge about their current medications and drug allergies is currently a tedious and error-prone task [1]. The widespread growth of prescribing and consuming medications has increased the need for applications that support medication reconciliation. Furthermore, living healthcare issues include misdiagnosis, overtreatment, decreased productivity, under-utilized clinical data handling, significant cost and spending (Figure 1). These miscalculations can be reduced to a great extent with the use of advancements in information technology at every level of care.

Medical error is the third leading cause of death after heart failure and cancer (2). According to recent studies, approximately 180 000 to 251 000 people are dying every year in the USA due to medical errors (2). This number has been rising due to increasing complexity and reduced quality of our current medical system, which includes communication breakdown, misdiagnosis, poorly coordinated care and growing cost. In recent years, the concept of precision medicine has evolved as a central innovation pillar for leading research in transforming health and holds great promise in patient treatment (3,4). Precision medicine has the potential to improve the traditional symptom-driven practice of medicine by intelligently integrating multi-omics profiles with clinical, imaging, epidemiological and demographic details to allow a wide range of earlier interventions for advanced diagnostics and tailoring better and economical personalized treatment (Figure 1). This requires a progressive healthcare environment that can enable clinicians and researchers to gain a complete picture of the patient to deepen their understanding, using additional basic details from healthcare data e.g. phenotypic information, life style factors and social determinants that can impact treatment decisions. It is primarily based on '4Ps'—Predictive, Preventive, Personalized and Participatory—treatment of each individual patient and aims to enable clinicians to efficiently understand how personalized clinical data variations can contribute to health and accurately diagnose and predict the most appropriate course of action for a patient (5). While the complexities of diseases at the individual level have made it difficult to utilize healthcare information in clinical decision-making, some of the existing constraints have been minimized by technological advancements (6). To implement effective personalized and population health with enhanced ability to positively impact patient outcomes, it is important to harness the power of electronic health records (EHR) by integrating disparate data sources and discovering patient-specific patterns of disease progression to provide real-time decision support. The significance of healthcare data mining cannot be denied, but the challenges of big data management loom large (7).

Over the years, biotechnology has evolved immensely. Computers are becoming faster in speed and micro in size, heterogeneity is increasing in datasets and their volume is growing robustly. These expansions are fueling the engine of artificial intelligence (AI) for discovering many technical refinements to solve complex problems in almost every field of life, including science and medicine. AI is the branch of computer science with the capability of a machine to imitate and even enhance intelligent human behavior. One of the expected roles in life and medical sciences is to
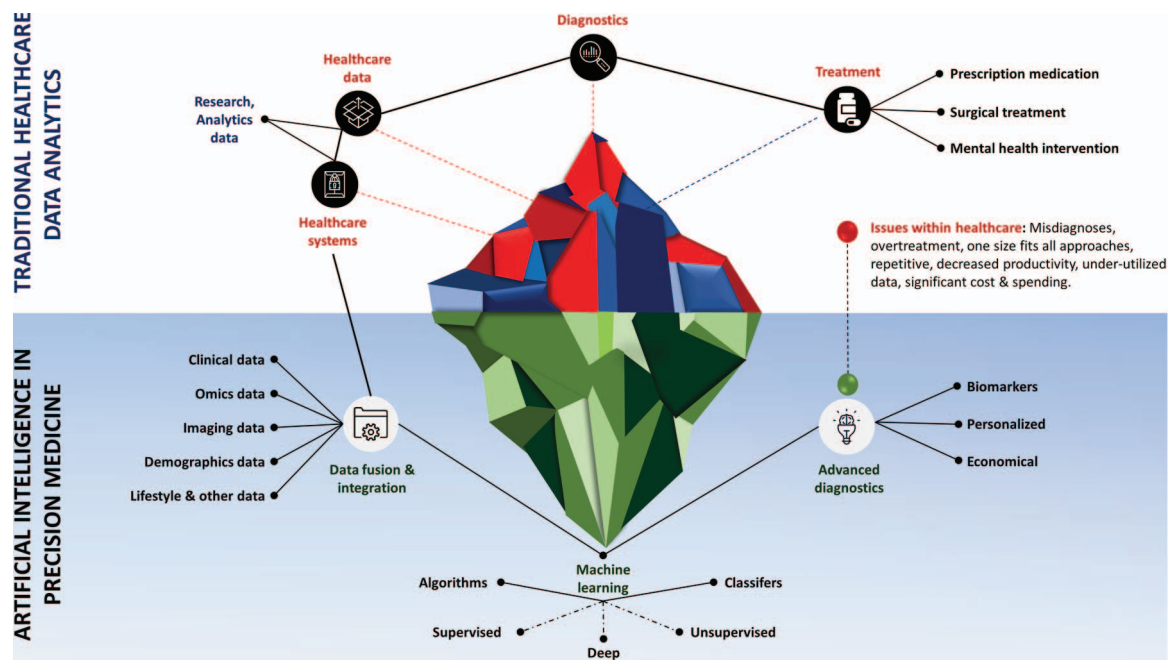
Figure 1. Role of artificial intelligence in traditional healthcare data analytics, and in precision medicine. Addressing key issues in healthcare (e.g. misdiagnoses, overtreatment, one-size-fits-all approaches, repetitive, decreased productivity, under-utilized data, significant cost & spending), and finding key biomarkers to provide economic and personalized treatment by intelligently analyzing heterogeneous data.

deal with extensive research studies aimed at supporting real-time decision-making and producing solutions to complex problems through knowledge and data intensive computational and simulated analysis (8). Healthcare data includes information about a patient's lifestyle, medical history, encountered visits with practices, laboratory and imaging tests, diagnoses, prescribed medications, performed surgical procedures and consulted providers (9). Adequate, analytical and intelligent access to healthcare data has the potential to revolutionize the field of medicine by improving the quality and transition of care, improving outcomes by reducing cost, detecting diseases at earlier stages (10,11) and developing a better understanding of biological mechanisms by modeling complex biological interactions through a holistic integration and analysis of knowledge (12). The ability to stratify patients, understand scenarios and optimize decision-making would consistently improve based on the myriad data obtained during the care-delivery process. Innovative and robust big data platforms are necessary to improve the quality and transition of healthcare by analyzing heterogeneous healthcare, which can be of huge volume, velocity, variety and veracity.

To effectively implement healthcare data analytic processes, various big data management challenges (1,9,13,14, 15,16,17,18,19) have to be overcome, which include inadequacy of analyzable clinical data (20); existence of multiple data standards, structures, types and formats; rapid growth in heterogeneous data; understanding of anal-

ysis algorithms for clinical data interpretation, exploration and drawing inference; unavailability of effective open-source tools that combine various approaches to model biological interactions; integration of clinical and analytic systems; interdisciplinary field barriers; high cost (21); implementation of secure frameworks for data collection, simplification, conversion from raw form to knowledge, management and distribution (22,23); automatic cleansing of faulty and error-prone EHRs; correctly identifying prescription medication; and implementing predictive diagnostics (10). It is not possible to easily track and prospectively follow the clinical progress and outcomes in patients over time (e.g. having a critical predictor of future clinical events where a patient may show the disease months or years down the road). In past decades, various systems have been developed in both commercial and academic sectors (11,24,25,26,27,28,29,30) for this purpose. Academic systems put significant values on analytics, while commercial systems focus on supporting clinical operations. However, independently utilizing traditional approaches, both sectors are unable to identify problems by their effects and significantly help in clinical decision-making. However, major concerns include handling and evaluation of electronic medical records (EMR), and repetitive tasks; patient medication adherence; inefficient therapeutic treatment for cancer and other critical diseases; cost ineffectiveness; and addressing ethical issues related to AI and ML implementation in healthcare.

AI & ML in health intelligence, precision medicine and resource management

Intelligent big data platforms are necessary to improve the quality and transition of healthcare by expediting investigation of active hidden factors in clinical data with machine learning algorithms to obtain actionable gap-based information about patients for early detection and prevention of constitutional disorders like cancer, and streamlining data sharing by developing efficient communication across healthcare units and scientific laboratories. Its application in healthcare could be another great leap in medicine and a transformational force for guiding personalized and population medicine with several computational benefits. In the recent past, multiple AI and ML-based efforts have been made for deciphering diseases to facilitate predictive diagnosis and thereby guide treatment factors, e.g. drawing disease relationships using clinical manifestations, EHR and data generated using wearable technology. To get a detailed overview of available academic solutions, we reviewed contributions and compared various AI and ML claimed solutions, approaches and discussions (Table 1), and real-time examples (Table 2) published within the last 5 years [31–63]. Our focus here was to discuss valued contributions of all mentioned AI and ML algorithms (section   Theoretical background of AI, ML and examples in healthcare and approaches. We provide detailed and individualized overview of presented approaches. Overall, the review study and the contributions of AI and ML are divided into three categories: Health Intelligence, Precision Medicine, and Healthcare Resource Management and Ethical Challenges.

Health intelligence approaches

Health intelligence can play a vital role at various levels of clinical research and analytics that can lead to significant improvements in achieving goals for the provision of better personalized and population healthcare. In the past decade, various operational and research-based healthcare data management and analytic systems have been developed in both academia and commercial sectors. Our interest includes comprehensive solutions that implement healthcare data analytics process; provide features to manage, analyze, visualize and share EHR in de-identified form; help in automatically capturing information about patient demographics, scheduled appointments, pre-exam questionnaire results, consulted providers, conducted lab tests, diagnoses, treatment plans, objective test results, medications, surgical procedures and claims; support the clinical decision-making process with AI techniques to create classifiers, which can be trained on structured clinical data generated from different clinical activities and can learn

similar groups of subjects, associations between subject features and outcomes of interest; and apply natural language processing (NLP) methods to extract information from unstructured clinical data e.g. narrative text, such as physical examination, clinical laboratory reports, operative notes and discharge summaries.

AI in healthcare for better prevention, detection, diagnosis and treatment of disease (31)

The authors emphasized the idea of embracing changes with the advancement in technology with the potential integration of AI into the field of healthcare in a way that is beneficial to each healthcare worker. They focused in utilizing AI to obviate repetitive tasks to enhance patient–physician relationship and increase practice of empathy and emotional intelligence. Authors focused on deep learning algorithm implementation with increased data flow that allows machines to self-develop a complex function with improved predictability, as long as a large amount of data is fed as input. They developed a deep convoluted neural network for skin cancer detection, image analysis for diabetic retinopathy evaluation, smartphone-based AI platform to measure adherence in patients on direct oral anticoagulants (64) and patient's visit length reduction (65). The strengths of this study include the potential to augment healthcare access in areas where specialists might not be physically available, and medication can still be prescribed utilizing the combined efforts of AI and a primary care physician, especially in developing countries. However, one missing aspect is to address the loss of privacy, as well as possibilities of patient data exploitation. The potential for Health Insurance Portability and Accountability Act (HIPAA) protection is likely feasible (1,9).

ML in medicine with better patient–provider interactions (32)

This approach is focused on examining the essential structural changes in the healthcare systems that are necessary to unleash the full potential of ML in medicine. It emphasizes developing the concepts of ML in medicine, which may be centralized around the idea of personalized diagnosis and treatment on the basis of all known information about the patient and collective experience. Giving rationale, authors highlighted proof-of-concept models that have been tested so far, e.g. difficulties in finding a relationship between current ML models and traditional statistical models, need for a tremendous amount of data to train ML classifiers for establishing general and complex associations and training clinicians in AI for accurate data interpretation. Furthermore, they effectively debated about consumption of physi-

Table 1. Feature and variability analysis of reviewed approaches, and real time implementation of ML algorithms

Approach	Objectives	Approach	AI & ML
AI power digital medicine (31)	Reduce repetitive tasks and burdens of electronic medical records through the utilization of AI and ML.	Increased task automation with improved image processing. Monitor medication adherence and detect any changes.	Deep convoluted neural network for skin cancer detection and reducing visit length. Deep neural network to evaluate images for diabetic retinopathy. Smartphone-based AI platform to measure adherence in patients on direct oral anticoagulants.
ML in medicine (32)	Examining the essential structural changes in the healthcare system that are necessary to unleash the full potential of machine learning in medicine.	Accumulation of large data set and implement ML to anticipate events, develop search engine, and monitor data flow.	Applied deep learning on the current EHRs data to generate associations and meaningful data for personalized diagnosis and treatment.
Precision medicine with electronic medical records (33)	Applying ML to the EHRs to generate personalized medicine by converting EHR into reliable risk predictors, and incorporating patient's variabilities for treatment and prevention of disease.	Analyze patterns within the subset of population who present similar clinical phenotypes of complex disease.	Supervised learning (support vector machine, discriminant analysis, naïve Bayes, nearest neighbor and neural network), unsupervised learning encompass (linear & logistical regression, decision tree, cluster analysis, and neural network).
AI, ML and the evolution of healthcare (34)	Examining AI integration in healthcare.	AI methods for the extraction of big data and aid clinicians in care delivery	SVM model development for physiological data segmentation and analysis, disease progression prediction, and diagnosis.
Solving healthcare problems with precision medicine (35)	Tailoring medical treatment with respect to the individualized characteristics of patients.	Use of information technology for multidisciplinary collaboration establishment between clinicians and researchers.	ML for the implementation of precision medicine, which includes data storage and analysis for determining the association between disease outcome (e.g. disease risk, prognosis, or treatment), identification of patient characteristics and optimal treatment.
Role of AI in precision medicine (36)	Examining role of AI in precision medicine implementation.	Analyzing large scaled clinical dataset.	Combining DL with human pathologist to improve success rate of diagnosis.
AI towards health in resource-poor settings (37)	Utilization of AI in poor settings, and improving health outcome in those areas.	Implementing NLP over EHR for surveillance and out breaking predictions.	Pattern identification and tracking disease transmission through ML.
Integrated precision medicine and role of EHR in personalized treatment (38)	Early diagnosis of chronic conditions through proper extraction of clinical insights.	Feature extraction from clinical data, and utilization of silico dataset.	Predictive, proactive intervention in healthcare through AI, and clinical decision support tool development.
AI in healthcare (39)	Analyzing AI applications in healthcare, and their potential outcome in future.	Precise analysis at the extracted useful information from a large patient population.	ML algorithms to extract and cluster data, and perform principal component analysis, SVM to determine model parameters, and identify imaging biomarkers, NLP for text processing and classification, and DL for diagnostic imaging and electro diagnosis.

(Continued)

Table 1. Continued

Approach	Objectives	Approach	AI & ML
ML Knowledge Base with ontology for pattern recognition in personalized medicine (40)	Examine three main pillars integrating personalized medicine into everyday clinical practice, which are phenotype categories, population size and statistical analysis.	Developing knowledgebase of existing phenotypes, patient enrollments, and data expansion.	ML approaches for pattern recognition and development of statistical models (sample size and effect size). Knowledge base of all existing phenotype categories and disease. Organized clinical dataset of population size. Software platform for statistical analysis of high-dimensional healthcare and multi-omics data.
Data science, AI, and ML for laboratory medicine (41)	Predictive modeling for better collaboration between hospitals without sharing data and complying privacy regulation.	Data Science (DS) and AI to mimics the human processes and improve the process of decision-making.	ML for healthcare data analysis and optimization, and reducing cost, improving efficiency of staff and resources.
AI to solve the human resource crisis in healthcare (42)	Solve the human resource crisis in healthcare with AI.	Implementation of AI techniques (43).	Artificial narrow intelligence for performing a single task. Artificial general intelligence for understanding and reasoning environment like humans. Artificial superintelligence for scientific creativity. Deep learning for image recognition, natural language processing and translation.
Data analytics and ML for disease identification in EHR (44)	Analyzing EHR for the identification of wide range of medical conditions and diagnosis.	Converting electronic healthcare record into reliable risk predictors.	ML algorithm for structured and unstructured big data analysis for the identification of wide range of medical conditions and diagnosis.
AI, Big Data and Cancer (45)	Application of AI and large scaled database for cancer diagnosis and treatment, worldwide.	Application of cognitive computer systems for approaching cancer diagnosis and treatment (read, remember, recommend, and remind).	Cognitive computer systems for providing rapid access to accurate information and treatment procedures, and assisting in decision-making.
Use of EHR in comparative effectiveness research (46)	Reporting caveats in existing healthcare systems.	Literature review and reporting caveats.	Implementing ML for overcoming existing big data limitations in healthcare systems.
Deep learning health care system (47)	Reporting unintended consequences due to the application of ML in existing healthcare systems.	Creating more precise analytics platform for prognosis modeling and pattern recognition.	ML for prognosis modeling in oncology, and pattern recognition in radiology and pathology.
DL to transform healthcare (48)	Transform healthcare by using ML.	Outperforming clinical systems and modeling complex relationships among active hidden factors of data	Implementation of DL for the digital image analysis.
High-performance medicine with AI (49)	Exploring importance and pitfalls of AI in medicine.	Literature review and field analysis.	Deep neural networks for pattern recognition and analysis medical images. NLP in drug discovery by analyzing biomedical literature.

(Continued)

Table 1. Continued

Approach	Objectives	Approach	AI & ML
Intelligent digital pathology (50)	Improving diagnostic accuracy and efficiency with the use of ML.	Implemented, examined and compared the performance of DL at test dataset.	DL for analyzing whole-slide pathology images.
ML for prediction in EHR (51)	Implementing ML for better understanding heterogeneous treatment effects to implement precision medicine.	Evaluating positives and negatives of ML algorithms.	ML algorithms for addressing different clinical questions by analyzing and finding nonlinear relationships in the EHR.
Unintended consequences of ML (52)	Safe, effective, efficient and humanistic care.	Implementing DL in healthcare analytic systems development and modeling tools.	DL for digital imaging, curating data sets, integrative heterogeneous data analysis, identifying novel associations, and remote monitoring and digital consultations.
Finding the missing link for big biomedical data (53)	Biomedical data integration and analysis located at heterogeneous sources.	Identify and discussed challenges in biomedical data linking.	AI and ML tools development to analyze biomedical data for better clinical decision-making.
ML classifies cancer (54)	Identification of novel tumor classes.	Application of ML for the identification of tumor by analyzing histology and genomics data.	ML for analyzing histological data. Supervised ML for analyzing CNS tumor type genome-wide methylation data to identify methylation patterns. Unsupervised ML to search patterns in the data sets to develop classification categories.
Analyzing and visualizing knowledge structures of health informatics (55)	Finding future strands of research, including new algorithms, tracking tools and Internet of Things-based decision support systems.	Quantitative review of the health informatics field, employing text mining and bibliometric research methods.	DL, new ML algorithms and advanced big data analytics for better-personalized treatment.
Big data and ML algorithms for healthcare delivery (56)	AI tools development based on incremental learning to refine the predictive accuracies.	Identification of clinical problems, annotation of extracted healthcare data, application of appropriate ML algorithms and its effect on decision-making, addressing legal and ethical implications, assessment of ML effect in trail, designing freeze and submission of dossier for medical devices, training clinicians of ML tool, and monitoring for adverse outcomes.	HCI-based AI and ML applications for different clinical developments in oncological.
Intelligent health data analytics (57)	AI for advanced health data analytics.	Health data analytics process involving a methodical order of data processing, modeling, and analysis steps.	Implementation of AI and ML based analysis with the inclusion of health data preprocessing, selecting algorithm based on expected outcome, developing analytical models, and interpreting results.

(Continued)

Table 1. Continued

Approach	Objectives	Approach	AI & ML
Ethical challenges of implementing ML in healthcare (58)	Challenges of implementing ML in healthcare	Literature review and field analysis.	Addressing current challenges in healthcare systems due to the implementation of ML.
Data science, AI, and ML for laboratory medicine (59)	Implementing Data science, AI, and ML for laboratory medicine.	Framework including defining tasks, metrics, models and datasets.	ML for finding patterns, discovering inefficiencies, predicting outcomes and taking factual decisions.
Causal inference and ML (60)	Examined the implications of progress in observational research design and healthcare databases.	RWE framework.	ML for data classification and prediction in RWE to support clinical and regulatory decision-making.
Big data analytics in healthcare (61)	Application of big data analytics in healthcare.	Conceptual architecture of big data analytics, which includes developing multi source data input, transformation, structure, management and analysis using traditional SQL, OLAP and mining.	ML for data mining and analysis.
ML and genomics in precision medicine (62)	Substantial improvements to address clinical and genomic data security problems.	Combining the latest computational data protection principles with legal and ethical perspectives to construct a secure framework for data sharing.	ML models to address the challenges of gene variations and similarities among patients.
ML in cancer prognosis and prediction (63)	ML to detect key features by predictive modeling of complex and heterogeneous datasets for progression and treatment of cancerous conditions, risks and outcomes.	Data preprocessing with focus on data modification via dimensionality reduction and feature detection.	ML (ANNs, BNs, SVMs, graph-based SSL and DT) to model the progression and treatment of cancerous conditions thru examining complex datasets and revealing their relevance.

cians' valuable time due to increased features (e.g. check boxes) in EHRs for administrative and billing purposes, which might prevent them from providing the best quality care for their patients. At the same time, they claimed that the integration of ML in EHRs may lead to potential fear of overreliance as well as decreased vigilance of errors and automation bias. Authors claimed a training ML classifier at EHRs for pattern detection to allow physicians to anticipate future events in high-risk patients, to obtain accurate and comprehensive diagnosis and provide with a quick search engine in locating the pertinent information within a patient's chart, less clicking, voice dictation and better predictive typing.

AI, ML and the evolution of healthcare (34)

Overview and implementation of AI and ML in healthcare for the extraction of big data, and aiding clinicians in providing better care delivery were recently highlighted

by authors. They discussed use of support vector machine (SVM)- and deep learning-based model development for physiological data segmentation and analysis, disease progression prediction and diagnosis in radiology. Authors discussed processes, which include designing effective models to aid in the diagnosis based on the information that resemble certain diseases, image analysis and interpretation to improve the decision-making performance of clinicians. Authors also raised ethical concerns in utilizing ML, primarily in governance and management of big data and future of employment.

Medical AI for increasing availability of healthcare data and rapid development of big data analytic methods (39)

A study analyzed different AI applications in healthcare and discussed their potential future outcomes. Authors argued that currently AI research in healthcare is mainly focused on

Table 2. Real time examples of AI and ML algorithms (support vector machine, deep learning, logistic regression, discriminant analysis, decision tree, Random forest, linear regression, naïve Bayes, K-nearest neighbor, hidden Markov, genetic algorithm) in healthcare

ML algorithms	Examples in healthcare
Support vector machine	Symptoms classification and analysis to improve diagnostic accuracy.
	Identifying imaging biomarkers of neurological and psychiatric disease.
	Diagnosing mental illness.
	SVM with leave-one-out cross-validation for multiple myeloma by analyzing SNPs.
	SVM with 20-fold cross-validation for breast cancer by analyzing SNPs.
	SVM with hold-out for breast cancer by analyzing clinical, pathologic and epidemiologic data.
	SVM with hold-out for cervical cancer by analyzing clinical and pathologic data.
	SVM with 10-fold cross-validation for breast cancer by analyzing clinical and population data.
	SVM with cross-validation for oral cancer by analyzing clinical and genomic data.
	SVM with leave-one-out cross-validation for breast cancer by analyzing genomic data.
	SVM with cross-validation for oral cancer by analyzing clinical, molecular data.
Deep learning	Evaluating images for diabetic retinopathy.
	Identification of type 2 diabetes (T2D) subgroups.
	Measure medication adherence via camera interface.
	Detection and segmentation of lung and liver tumors by analyzing CT scans.
	Diagnosing eye diseases (diabetic retinopathy) by analyzing retinal images.
	Diagnosing cardiac anomalies by analyzing images of MRI of heart ventricles.
	Detecting malignant lung nodules by analyzing radiographs.
	Producing glioma survival predictions by analyzing histological imaging and genomic marker data.
	Histological diagnoses prediction in women with cytological abnormalities.
	Oncology diagnosis: • **Thoracic (lung cancer)** • **Abdominal and pelvic (tomography and magnetic resonance imaging)** • **Colonoscopy (colonic polyps)** • **Mammography (microcalcifications)** • **Brain (brain tumors)** • **Radiation oncology (segmenting tumors for radiation, and quantifying specific radiographic characteristics by analyzing 3D shape of a tumor from)** • **Dermatology (skin cancer)** • **Pathology (digital whole-slide of biopsy samples)** • **Prostate (cancer tumors by analyzing ultrasound of biopsy cores)** • **DNA and RNA sequencing (RNA-binding and DNA-binding proteins)**
Logistic regression	Risk assessment of complex diseases (e.g. tuberculosis, breast cancer, coronary heart disease).
	Predicting patient survival rate.
	Diagnosing coronary heart disease (CHD).
	Non-Hodgkin's lymphoma diagnosis with multivariable logistic regression modeling.
	Identification of pulmonary thromboembolism by analyzing prognostic factors.
Discriminant analysis	Identify surgical and operative factors to classify patients for surgical procedure.
	Predict the clinical diagnosis of primary immunodeficiencies.
	Patient data satisfaction.
	Prediction of depression elements in cancer patients.
	Classification of BOLD fMRI response to naturalistic movie stimuli.
	Identify protein-coding regions of rice genes.
	Parkinson's disease symptoms recognition.
	Risk assessment of for chronic illnesses.
	Diagnosis of hypercalcemia.
	Predicting patient care visits by identifying discriminatory characteristics.

(Continued)

Table 2. Continued

ML algorithms	Examples in healthcare
Decision tree	Real-time healthcare monitoring medical decision support system, sensors anomaly detections, and data mining model for pollution prediction. Supporting clinical decisions. Strategies alternate therapies for oncology patients. Collaborative clinical decision-making in mental health care. Identify predictors of health outcomes. Find factors related to hypertension. Discover factors associated with pressure ulcers (PUs) among elderly people. Identify the potential recipients of telehealth services. Patient data stratification for interpretable decision-making for precision medicine. Content analysis for patient aids decision. Diabetic foot amputation risk analysis. Support understanding of antenatal lifestyle interventions.
Random Forest	Diagnosing mental illness. Detecting knee osteoarthritis. Monitoring medical wireless sensors. Diagnosing Alzheimer disease. Predicting metabolic pathways. Predicting outcomes of a patients encounter with behavioral health providers. Healthcare cost prediction. Mortality prediction for intensive care unit (ICU) patients. Classification of Alzheimer's disease. Identifying social and economic factors to study social determinants of health. Predicting disease risks from imbalanced data. Identify associates of diabetic peripheral neuropathy diagnosis. Predicting the risk of emergency admission. Detecting patients ready to discharge from intensive care. Nonparametric estimation of heterogeneous treatment effects. Diagnose sleep disorders. Predicting the depression in patients suffering with Alzheimer's disease. Predicting myopia by analyzing EHR.
Linear regression	Identification of prognostically relevant risk factors. Predict hand surgery. Monitor prescribing patterns and ensure treatment appropriateness. Mean on decision-making in health care. Reducing high costs of the health system. Analysis of skewed healthcare cost data. Understand HIV/AIDS prevalence patterns.
Naïve Bayes	Predictive modeling for different diseases (brain, asthma, prostate and breast cancer etc.). Risk prediction using censored and time-to-event data. Mucopolysaccharidosis type II detection. Predicting Alzheimer's disease from genome-wide data. Measuring quality healthcare services. Finding audit targets in performance-based financing in health. Modeling medical diagnosis for decision support. EHR classification. Classifier and genetic score for risk prediction. Decision support system for heart disease.

(Continued)

Table 2. Continued

ML algorithms	Examples in healthcare
K-nearest neighbor	Diagnostic performance of the model.
	Preserving privacy of medical diagnosis in e-Health cloud.
	Medical dataset classification.
	Classification of lymph node metastasis in gastric cancer.
	Pattern classification for breast cancer diagnosis.
	Pattern classification for health monitoring applications.
	Pancreatic cancer prediction combining published literature and EHR data.
	Disease diagnosis and detection of Parkinson.
	Hand-gesture recognition for elderly individuals.
	ECG features extraction for mobile healthcare applications.
Hidden Markov	Analyzing sequence data (predicting exons and introns, identifying ORFs, insertions, deletions, substitutions, functional motifs in proteins, aligning two sequences, and switching from exon to intron in a DNA sequence)
	Modeling 'Healthy' and 'Unhealthy' unobserved health states.
	Analyze time-series personal health checkup data.
	Improving length of stay prediction and reducing health care costs.
	Mining adverse drug reactions from online healthcare forums.
	Monitor, model and cluster medical inpatient journeys.
	Analyzing healthcare service utilization after transport-related injuries.
	Monitoring circadian in telemetric activity data.
	Predicting patients entering states with a high number of asynchronies.
	Analyzing subject-specific seizure, automatic segmentation of infant cry signals.
Genetic algorithm	Detecting microcalcifications in mammograms leading breast cancer.
	Developing non-invasive technique for cervical cancer detection.
	Analyzing microarray data from cancer cell lines.
	Investigating relationships between soil trace elements and cervical cancer mortality.
	Parameter estimation for determining tissue elasticity.
	Predicting risk of a major adverse cardiac event (MACE).
	Detect QRS complexes.
	Detecting hypoglycemia EEG signals.
	Predicting time to reach full cervical dilation.
	Selecting optimal features of cardiotocogram recordings.
	Identifying autism by analyzing gene expression microarray data.
	Predicting outcome of patients with non-small cell lung cancer (NSCLC).
	Diagnosing patients by classifying lung sounds into normal, wheeze, and crackle.
	Choosing appropriate highly active antiretroviral therapy (HAART) to control HIV.
	Improving the selection of gantry angles to optimize stereotactic radiotherapy.
	Training robot for physiotherapy of the lower limb.
	Estimating Cobb angle from torso asymmetry in scoliosis.
	Analyzing mutations in Parkinson's disease.
	Predicting of tacrolimus blood levels, scheduling patient admission in ophthalmic hospital.

only a few health conditions, e.g. cancer, neural disease and cardiovascular disease. They criticized that existing health-care systems do not provide incentives for data sharing and have no structure for the implementation of AI. Further classified applications of AI consist of ML algorithms to extract and cluster useful information from a large patient population to assist in making real-time inferences for health risk alerts and health outcome predictions, perform principal component analysis and reduce diagnostic and therapeutic errors; SVM to determine model parameters and identify imaging biomarkers; NLP for text processing and classification; and deep learning for diagnostic imaging and electronic diagnosis.

Data analytics and ML for disease identification in EHR (44)

Managing large volumes of data is a huge challenge for doing big data analytics. A group of researchers used an ML algorithm for structured and unstructured big data analysis to identify a wide range of medical conditions and diagnosis from the large-scale EHR database, inclu-

ding information about test results, historical information, management plans, and billing codes, etc. Authors initially applied their algorithm for the identification of potential predictors and keyword associations in the EHR to find presence of pseudoexfoliation syndrome (PXF) by using NLP, least absolute shrinkage and selection operator. The reported outcomes were based on positive and negative predictive values of the algorithm, which were also validated by the glaucoma specialists. Authors highlighted the potential of their algorithm with its application for data mining and predictive analytics in many other applications as well.

AI, big data and cancer (45)

The application of cognitive computer systems has been very successful for approaching cancer diagnosis and treatment (read, remember, recommend and remind). According to a team of researchers, cognitive computer systems can support physicians by providing rapid access to accurate information and treatment procedures, assisting in decision-making. Such systems have the potential to distribute cancer knowledge into clinical practice and remote areas worldwide, by optimizing clinical research and trials with reduced bureaucracy and cost. They presented the possibilities of AI in future cancer care and research, which includes developing international cancer networks, identifying beneficial therapies for rare and highly aggressive cancers, observing different therapeutic outcomes by different parameters, analyzing associations of cancer with other disease-specific attributes, discovering new cancer etiologies, incorporating pertinent patient and cancer characteristics into clinic-based uses, conducting economic broad-based cancer trials, uncovering genomic and molecular events sensitive to existing or new treatments and analyzing and developing new treatment pathways. Furthermore, authors predicted that large-scaled AI databases may benefit cancer programs and help cancer treatment and research.

Use of operational EHR in comparative effectiveness research (46)

EHR has a great role in improving the quality and cost of healthcare, advancing biomedical science and facilitating clinical research. Supporting this claim, this overview presented some examples of running and accomplished research projects, e.g. Electronic Medical Records and Genomics (eMERGE) Network, Strategic Health IT Advanced Research Projects (SHARP) Program and Health Maintenance Organization Research Network's Virtual Data Warehouse Project. The authors also discussed multiple caveats in existing healthcare systems e.g. inaccurate data entry in EHRs; incomplete patient information in EHRs; incorrect diagnosis and medication codes (e.g. ICD, National Drug Code (NDC)), and their conversion into research descriptions and vice versa; data extraction and integration from clinical notes in to EHR; multiple data sources in EHRs affecting data provenance; mismatches of data granularity with comparative research; and differences in research protocols and clinical care.

Deep learning in the healthcare system (47)

Deep learning is a dominant ML approach, which has been greatly augmented in healthcare, analytic systems development and modeling tools. This review presented potential healthcare applications based on multiple driving factors, including learning digital imaging (e.g. radiology, radiotherapy, pathology, ophthalmology and dermatology) with deep learning to facilitate effective decision-making and therapy, digitization of EHR and applying ML for curating and analyzing data sets, integrative heterogeneous data analysis using deep learning, applying deep learning for hypothesis generation by identifying novel associations to establish causation and causal pathways, appropriate deployment of AI and ML-based platforms for remote monitoring and digital consultations and improving performance of deep learning with exposure to larger datasets. Authors predicted safe, effective, efficient and humanistic care in the future with the successful application of ML.

Deep learning to transform healthcare (48)

Deep learning has the potential to transform healthcare by outperforming clinical systems and modeling complex relationships among active hidden factors of data. This study highlights the background, workflow and challenges of deep learning with its successful implementation of digital image analysis, and some examples like analyzing fundus images of the retina to predict cardiovascular disease and identifying images for melanoma, basal and squamous cell carcinoma by matching sensitivity criteria. They suggested developing and training classifiers based on the deep neural network to make predictions on big clinical datasets by using the nonlinear features for modeling regularities. Authors claimed that in the coming years, with the continuous increase in the volume of datasets and without any enhancements to the basic learning techniques, more promising results can be achieved by ML.

Intelligent digital pathology with deep learning (50)

The data efficiency of deep learning can be used to augment information by improving diagnostic accuracy and efficiency by analyzing whole-slide pathology images, which

cannot be easily observed by the human eye. Performance of deep learning at test sets of 129 whole-slide images was examined and compared for identifying metastases in hematoxylin and eosin–stained tissue sections of lymph nodes from women with breast cancer. Authors of the study optimized deep learning algorithms for obtaining impressive results, even better than the participating panel of 11 pathologists. They predicted a bright future of deep learning applications in digital pathology and discussed the need of new intelligent tool development for handling diagnostic sensitivity and specificity.

ML for predictions and finding nonlinear relationships in the EHR (51)

ML real-time applications have been used to address different clinical questions by analyzing and finding nonlinear relationships in the EHR. A research study highlighted one important aspect that could be one of the ML drawbacks and that is the utilization of most of the ML algorithms to solve clinical problems when they were originally proposed for other matrices. The authors of the study provided some examples to justify their claim and discussed that the primary purpose of operational EHR systems development is to not support any kind of ML implementation for predictions. Due to the lack of robust and structured clinical data, authors showed uncertainty in achieving high-quality results with the use of ML. They stressed the need for new tools based on the ML algorithms especially designed with prior thresholds for improvements. They also stated that the present discovery phase needs to implement ML for better understanding of heterogeneous treatment effects to implement precision medicine.

Analyzing and visualizing knowledge structures of health informatics to uncover the explicit and hidden patterns (55)

A quantitative review of the health informatics field uncovers the scientific growth in this field. As discussed, despite a long period of progress, there is still no proper characterization of the knowledge and no common language, which necessitates illumination of the knowledge structures, text mining methods, scientometric analysis, tracking tools, Internet of Things-based decision support systems and social network visualization to identify hidden patterns. The study conducted produced six clusters and nine research themes to support health information technology by improving patient safety by reducing medication errors, and associating decision support, knowledge representation, telehealth innovations and professional behavioral changes in medicine. For the provision of better-personalized treatment, authors suggested that the health

informatics field should involve DL, new ML algorithms and advanced big data analytics.

Intelligent health data analytics to improve health system management, health outcomes, knowledge discovery and healthcare innovation (57)

Various healthcare systems are generating heterogeneous data of significant volumes which demands exploitation of healthcare data for resource optimization, patient satisfaction, improved care quality and health outcomes. Recommended use of AI and ML for advanced health data analytics with the arrangement of active partners in the healthcare processes requires ubiquitous services, therapeutic decision support, ethnographic health surveillance, integration of health-related data sources, personalized and predictive medicine by learning non-linear associations and drawing causal relationships among inherent data elements. A team of scientists designed a health data analytics process involving a methodical order of data processing, modeling, and analysis steps categorized as data- and knowledge-driven methods for Decision, Predictive, Descriptive, Optimization, Comparative, Prescriptive and Semantic analysis. They advocated implementation of AI and ML-based analyses with the inclusion of health data pre-processing, selecting algorithm based on expected outcome, developing analytical models and interpreting results.

ML in cancer prognosis and prediction (63)

ML can be applied to detect key features by predictive modeling of complex and heterogeneous datasets for progression and treatment of cancerous conditions, risks and outcomes. Many research groups have highlighted the implementation of different ML algorithms in cancer research for estimating unknown dependencies to predict new outputs of the system, including artificial neural network, Bayesian network, SVMs, graph-based semi-supervised learning (SSL) and decision tree. There are numerous ML real-world applications. The artificial neural network has been established for breast cancer by analyzing mammographic and demographic data with k-fold cross-validation rate and lung cancer by analyzing clinical and gene expression data; SVM has been used for analyzing single-nucleotide polymorphisms (SNPs) with leave-one-out cross-validation for multiple myeloma, with k-fold cross-validation for breast cancer and with k-fold cross-validation for breast cancer by analyzing clinical and population data; the Bayesian network can efficiently analyze clinical and pathologic data with cross-validation for colon carcinomatosis and with k-fold cross-validation for oral cancer by analyzing clinical and imaging tissue

genomic; blood genomic data; SVM hold-out for breast cancer by analyzing clinical, pathologic and epidemiologic data and for cervical cancer by analyzing clinical and pathologic data; graph-based SSL algorithm with *k*-fold cross-validation for colon cancer; breast cancer by analyzing protein–protein interactions (PPIs) and gene expression data and *k*-fold cross-validation for breast cancer by analyzing Surveillance, Epidemiology, and End Results (SEER); SVM with cross-validation for oral cancer by analyzing clinical and genomic data; SVM with leave-one-out cross-validation for breast cancer by analyzing genomic data; Bayesian network with hold-out for breast cancer by analyzing clinical, microarray data; decision tree with cross-validation for breast cancer by analyzing SEER; SSL co-training algorithm with *k*-fold cross-validation for Breast cancer by analyzing SEER. Authors acknowledged the active contributions of ML in accurate cancer susceptibility, recurrence and survival predictions. However, some concerns exist due to lack of external validation regarding the predictive performance of models utilizing integrated clinical and genomic data. They underlined that application of ML methods in cancer requires significant validation in order to utilize them in everyday clinical practice. Authors suggested data pre-processing with focus on data modification via dimensionality reduction and feature detection.

Precision medicine approaches

The underlying assumption here is that precision medicine will provide tailored healthcare to patients and will yield lower rates of associated adverse outcomes. A classic example of precision medicine is the customization of disease treatment for a single individual which in the old paradigm was a one-size-fits-all medicine; an effective treatment is the treatment known to benefit most of the target population. However, a certain treatment may actually yield benefit to only a few individuals. The rest of the population will not benefit from the treatment and may even incur adverse effects. This exemplifies the need for AI and ML-based systems bridging multiple domains in a secure environment for heterogeneous healthcare data analysis and visualization.

Precision medicine with EMR analysis for prevention and treatment of diseases (33)

Different healthcare institutions do not necessarily utilize the same EMR system with the possibilities of effectively communicating with each other, which makes it difficult for physicians to track patients' overall medical history. A team of authors brought about the importance of population perspective and need of foundation platforms for precision medicine, which can support large-scaled clinical data integration, and communication between different EMRs accessible to patients through different health centers. Authors proposed a universal EMR platform development by integrating population perspectives for the establishment of protocols to identify subgroups that fit distinct clinical phenotypes of complex disease and provides avenue for different treatment methods. Authors suggested an additional fifth 'P'—Population-wise Perspective—to the existing 4P precision medicine concept. They justified this claim by examining a study about identification of a type 2 diabetes subgroup through a topological analysis of a population's EMR and determined three major distinct clusters (66). Authors also suggested that it might be difficult to implement ML divisions, which include supervised learning (SVM, discriminant analysis, Naïve Bayes, nearest neighbor and neural network), unsupervised learning (linear & logistical regression, decision tree, cluster analysis and neural network) and deep learning. Authors recommended multi-cluster environment implementation for analyzing patterns within the subset of populations, which might present similar clinical phenotypes of complex diseases with the assumption that the treatment based on one cluster might not be as effective for another subgroup.

Solving healthcare problems with precision medicine (35)

Precision medicine has been groundbreaking in tailoring individualized and effective medical treatments based on the characteristics of each patient, and different susceptibilities to a particular disease, e.g. trastuzumab for HER2-positive breast cancer. Many authors are convinced by the importance of information technology and ML for the implementation of precision medicine, which includes data storage and analysis for determining the association between disease outcome (e.g. disease risk, prognosis or treatment), identification of patient characteristics and optimal treatment. Furthermore, authors highlight the requirement for multidisciplinary collaborations between clinicians and researchers.

Role of AI in patients the point-of-care, advanced analytics and foundation of precision medicine (36)

The scientific community criticizes the current healthcare structure, to be based on one-size fits all, and not transitioning from trial and error to evidence-based medicine. The authors recognize the importance of innovative technologies (e.g. genome sequencing, health sensors, advanced biotech) and essential roles of AI in precision

medicine implementation. Authors proposed three-factor-based approaches, which include point of care, large-scaled clinical datasets for training classifier and analytics and finally building foundations for precision medicine. The author's combined deep learning prediction with human pathologist's diagnosis resulted in a success rate of 99.5% and reduced human error by 85% at an early stage. Authors also presented guidelines for implementing AI in precision medicine, which included creating ethical standards for AI applications, gradual and incremental development of AI and training medical professionals about AI.

Integrated precision medicine towards targeted and personalized treatment for a given patient (38)

There has been a lot of debate over the current state of clinical decision support and how it can be improved in providing precision medicine. Many scholars have discussed involving ML in healthcare settings for the diagnosis and better treatment of chronic disease in clinical, translational and public health. Furthermore, they have highlighted the significance of big data management, privacy, de-identification and data sharing. Focus on early diagnosis of chronic conditions through proper extraction of clinical insights and utilization of *in silico* datasets could allow the replacement of animal and human models when conducting clinical trials by generating virtual patients with specific characteristics that enhance the outcome of each study. According to them, 25% of drug discovery occurs by chance, which can be highly accidental. They signify using predictive, proactive intervention in healthcare through AI and clinical decision support system development for leveraging big data analysis to make better predictions on the potential outcome of patients, ultimately supporting better decision-making by the physicians.

Authors have reviewed and reported analysis of the different clinical decision support states, improvements in patient outcome and various limitations and challenges in Laboratory, Medication, Diagnosis, Radiology, Pathology, Clinical Evidence & Outcomes, and Procedures, based on the availability of different data structures and standards in healthcare systems. Challenges in the laboratory mainly involve incorporation of genetic results into EMR in a searchable way; tests conducted at external labs cannot be incorporated due to lack of standardization, not encoding lab tests with Logical Observation Identifiers Names and Codes (LOINC), missing genetic information in EHR. Challenges in Medication include over time and ineffective drug combination. Challenges in Diagnosis surround handling of International Classification of Diseases (ICD) codes, as not all codes are not billable and some diagnoses

are not even encoded. Challenges in Radiology are of metadata mostly not formatted according to Digital Imaging and Communications in Medicine (DICOM), and reports are unstructured. Challenges in Pathology hold generation of unstructured reports, and frequent utilization of unstandardized nomenclature. Challenges in Clinical Evidence & Outcomes consist of insufficient data types, and unavailability of universally adopted data models. Challenges in Procedures mainly cover the process for approving new procedural codes. Authors have discussed the evolution of EHR to effectively delivering personalized treatment. With discussion of two case studies, they emphasized the implementation of precision medicine for the personalized delivery of care based on patient-specific patterns of disease progression and determination of precise therapies.

ML knowledgebase with ontology for pattern recognition in personalized medicine (40)

Personalized medicine is a broad and rapidly advancing field in healthcare, primarily based on disease-related clinical, genomic, metabolomics and environmental information. However, failure to correctly identify disease is one of the major reasons for misleading diagnosis, treatment and prognosis for the patient. A research group presented essential components (pattern recognition, knowledge base, ontology and patient profile) for accurate disease examination needed for successfully integrating personalized medicine into everyday clinical practice. Utilizing ML approaches for pattern recognition and development of statistical models (sample size and effect size), creating a knowledgebase of all existing phenotype categories and disease, organization of clinical datasets of population size and open software platform development for statistical analysis of high-dimensional healthcare and multi-omics data are crucial for practical realization of precision medicine.

ML classifies cancer by visual assessment of tumor cells (54)

One of the key technological advancements in the diagnosis of brain tumors was microscope-based analysis. Overcoming the visual limitation of such techniques for leading different classifications of a given sample by different individuals, a group of authors discussed ML as a precise solution for accurate diagnosis by analyzing molecular data. A previous study trained an ML classifier at a maximum number of images of tumors that were classified by physicians, as it is not possible to get precise conclusions, especially when tumor is histologically indistinguishable. Authors discussed the application of supervised ML for analyzing central

nervous system (CNS) tumor-type genome-wide methylation data to identify methylation patterns. They mentioned application of unsupervised ML to search patterns in the data sets to develop classification categories. Authors suggested application of ML for molecular analysis and visual inspection, as the disease can be the manifestation of both molecular and cellular changes.

ML and genomics in precision medicine (62)

Authors studied major concerns regarding data security like breaches in patient privacy and also presented the latest developments in the field of computational data protection. They discussed advancements in the era of big data for computer-aided diagnosis, and a recent concept of precision medicine for providing care for clinical, environmental and genetic characteristics. Authors emphasized developing new ML algorithms for computational analysis of genomic data to systematize the process of finding genetic similarities among patients with similar prognosis, or response to a treatment. They also stressed the need of implementing intelligent procedures to ensure the privacy of data, as they believe that data anonymization is not enough to guarantee unidentifiability due to auxiliary information. They discussed some existing solutions, which includes de-identification by data suppression, k-anonymization, learning from noisy data, homomorphic encryption, multi-party computation, cryptographic hardware and protecting genomic databases. Authors acknowledged the current contributions of ML but with the expectations for substantial improvements to address clinical data security problems and development of new significant ML models to address the challenges of gene variations and similarities among patients. They suggested combining the latest computational data protection principles with legal and ethical perspectives to construct a secure framework for data sharing.

Healthcare resource management and ethical challenges

Resource management is very important in any field of life, especially in healthcare. Aligning people and technology with organizational goals can positively impact with efficient implementation of planned workflow in achieving on time high-quality results. However, inefficient resource management may lead to over-exaggeration of organizational resources, which includes, time, cost, manpower and computational, bench and infrastructure resources. Furthermore, it is important to address the ethical and data privacy challenges, when implementing traditional state-of-the-art and intelligent healthcare data analytics.

AI towards health in resource-poor settings (37)

Importance and utilization of AI to improve health outcomes in low-income settings and regions has always been a point of discussion. One research focus has been on clearly identifying problems and issues related to the integration of global healthcare, application and deployment of AI in real-time environments, high-quality healthcare data collection in under-developed areas, automatic data extraction from hand-written notes in local languages, availability of high-performance computing and cloud-based environments for data management and analysis, and construction of knowledge bases and expert systems (67). Furthermore, concerns about data security, consent, and ownership, and ethical, patient safety and privacy-related issues that accompanies the utilization of AI are also present. Many solutions have been presented to address the problems of implementation of NLP in EHR for surveillance and predictions, AI implementation at pre-existing systems, weather and land pattern identification and tracking disease transmission through ML algorithms, implementing advanced expert systems, utilizing NLP to translate hand-written notes according to WHO standardized medical terminologies and developing local dictionaries, establishment of environments capable of working offline and synchronized with the remote databases, adopting cloud computing for the implementation of public health without established IT structure in Low and Middle Income Countries (LMICs), and using 'blockchain' for cryptocurrency and addressing issues related to the privacy and transparency. The scientific fraternity has predicted tremendous cost-saving and improved care delivery in coming years with the implementation of such along with related tasks.

Data science, AI and ML for laboratory medicine (41)

Data science has demonstrated success in laboratory medicine and reinforcing its value in transforming the healthcare system. This is a review on the perspective applications of data science current problems in healthcare, which includes the need of significant computational power to contribute to solving data optimization problems, overfitting in experimental designs, lack of data standardization, large-scaled datasets for training ML classifiers and other concerns in clinical laboratories, e.g. protected health information challenges, financial limitations and ethical concerns. Authors discussed the concept that data science and AI mimics human processes and improves the process of decision-making. They reasoned the use of predictive modeling for better collaboration between hospitals without sharing data and complying privacy regulation

through ML (supervised and unsupervised learning) for healthcare data (clinical data, imaging, laboratory tests) mining, analysis and optimization, identification of large patient clusters with certain disease characteristics, reducing medical errors and cost and improving efficiency of staff and resources. Data science has its own applicability and ethical challenges especially related to AI integration. AI encompassing data science when adopted can even improve human thought processes in efficient decision-making with different ML algorithms, data mining, knowledge discovery, improving complex analytical tasks and calculating clinical pathways. Authors of a study acknowledged success stories of ML in healthcare (radiology, pathology, dermatology and genomics), weather forecasting, structural recognition, NLP, games, analyses of financial transactions and improvements in industrial processes by finding patterns, discovering inefficiencies, predicting outcomes and taking factual decisions. They provided examples of ML application in healthcare for the detection of cervical cancer (neural networks), prediction of histological diagnoses (artificial neural network-based decision-support scoring systems) and radical hysterectomies with gene expression analysis (neuronal network), DNA (epigenetic modifications) and RNA (messenger, long non-coding or double-stranded) data analysis. ML is contributing to better personalized treatment by monitoring patient activities (analyzing data received by sensors for continuous measurement, e.g. glucose); analyzing molecular biomarkers; predicting drug efficacies, treatment responses and disease pathways; and identifying molecular factors and genetic variants. The authors presented an ML framework, which includes defining tasks, matrices, models and datasets. They showed concerns regarding implementation of human rights by the Universal Declaration of Human Rights at the 1948 United Nations General Assembly for data privacy, protection, de-identification and encryption for data handling, collection and sharing. Authors emphasized the consideration of key ethical concerns, which include consenting patient, AI human warranty and regulation of healthcare data according to principles of bioethical law.

AI to solve the human resource crisis in healthcare (42)

The healthcare workforce crisis is widening across the globe. With the increase in the number of chronic and complex diseases, the demand towards efficient healthcare system is consistently growing. However, the lack of access to care, differing quality and doctor shortages are increasing worldwide. Some authors have presented AI as the solution in healthcare and discussed its progress in the direction of

predictive and proactive interventions in clinical decision support systems. They reviewed AI approaches, dividing them into three categories (43), artificial narrow intelligence (performing a single task), artificial general intelligence (agent-based system, known as human-level AI) and artificial superintelligence (agent-based system, but smarter than better human). Authors debated the potential role of AI in filling the human resource gaps by integrating it with physicians to improve diagnostics and help in better decision-making. At the same time, they also highlighted the ethical implication of using AI technology as integrating part of the healthcare system.

High-performance medicine with AI for improving workflow and reducing medical errors (49)

There are multiple challenges of implementing AI in medicine. A research group underlined the obstacles and caveats, especially when applied in radiology, pathology, dermatology, ophthalmology, cardiology, gastroenterology and mental health. They predicted adaptation of AI by almost every type of clinician, which mainly includes deep neural networks for pattern recognition and analysis of medical images (e.g. medical scans, pathology slides, skin lesions, retinal images, electrocardiograms, endoscopy, faces and vital signs); applying deep learning to EHR for estimating the risk of a patient's hospital readmission; supporting doctors in decision-making for resuscitation; determining patients at risk of developing sepsis and other diseases; and predicting biological age, and critical diseases leading to death. Authors highlighted current challenges to the field of life sciences that can be solved by ML algorithms, like identification and isolation of rare cells, multi-omics data analysis, classification of somatic and germline mutations and gene–gene interactions and prediction of protein structure and PPIs, the microbiome and single cells. They also discussed the use of NLP in drug discovery by analyzing biomedical literature; mining molecular structures; predicting off-target effects, toxicity and right dose for experimental drugs; developing cellular assays; and using AI cryptography for determining unidentified drug interactions. Together with these advantages, authors also discussed formidable obstacles and drawbacks due to the field of AI, which includes data privacy and security, data hacking and breaches, uncertainty in use of black boxes of algorithms to resolve output, state of AI hype in validation and readiness for implementing in patient care. Authors anticipated the bright future of AI applications by predicting useful clinical outcomes in health systems, algorithmic interpretation of images and data, reducing errors, inefficiencies and cost.

Big data and ML algorithms for better healthcare delivery (56)

ML is considered a branch of AI encompassing algorithmic methods to solve scientific, healthcare and various other problems without traditional computer programming; this involves interplay between large population-specific datasets and model building. Suggested AI tool development should be based on incremental learning (ability to continuously improve with the inclusion of new data) to refine predictive accuracies. One study proposed analytics exercise starting with the identification of clinical problems, annotation of extracted healthcare data, application of appropriate ML algorithms and its effect on decision-making, addressing legal and ethical implications, assessment of ML effect in trial, designing freeze and submission of dossier for medical devices, training clinicians on use of ML tools and monitoring for adverse outcomes. Authors of the study presented ML algorithms suitable for different clinical applications and reported real-time clinical developments in oncology with the implementation of different AI and ML applications and algorithms. Authors discussed the importance of human–computer interaction in AI and ML platform development to effectively support decision-making processes in healthcare. They presented different real-time examples, which include HealthSuite, GE Healthcare, Lumada, DISCOVERY and CURATE.AI. They recommended model linking EHR in future ML platform development for addressing healthcare data analytics concerns.

Causal inference with ML (60)

Another study examined the implications of progress of AI in observational research design and healthcare databases, and implementation of ML for data classification and prediction in Real-World Evidence (RWE) to support clinical and regulatory decision-making. They discussed American Recovery and Reinvestment Act (ARRA) explicitly prohibiting cost-effectiveness and focusing on broad interventions for better diagnoses, treatments, disease management programs and healthcare organization models by mainly analyzing EHR and establishing patient-centered Outcomes Research. Authors discussed a wide range of increased size and nature of biomedical operational and research data, which requires efficient utilization of ML approaches to restructure, dimensionality reduction, clustering, modeling, linking, classification, analysis and predictions. Meeting RWE objectives, authors acknowledged ML as one of the powerful tools today for bringing significant improvement to care service. They exemplified ML with the implementation of SVMs to predict hospitalization; regression-based methods for reducing the risk of overfitting; *k*-fold cross-validation for splitting one's sample into two models; and

deep learning models for feature extraction, development of graphical processing units, predicting in-hospital mortality, unplanned readmissions, prolonged stays, and discharge diagnoses. However, they showed lack of confidence in ML for shielding against the normal challenges in observational data analysis, as screening some concerns require evidence.

Promise and potential of big data analytics in healthcare (61)

The effective use of big data by digitizing, combining large hospital networks and implementing efficient analytical approaches has been widely successful. According to a team of scientists, there are tremendous benefits of such approach in detecting diseases at earlier stages; predicting risk for medical complications; managing individual and population health, avoiding frauds; and addressing numerous healthcare questions. Furthermore, it can reduce waste and inefficiency by determining clinically relevant and cost-effective ways to diagnose and treat patients, applying ML algorithms to predict models and analyzing EHR and disease patterns to discover adverse treatment effects. Authors also discussed the current challenges, covering unavailability of user-friendly and transparent real-time big data platforms; lags between data acquisition, collection, cleansing, processing and standardization; missing ability to manipulate data at different levels for granularity, privacy and security enablement and quality assurance; difficulties in the management of large, diverse and complex data with traditional approaches; changing healthcare reimbursement models; lack of professional tools, infrastructure and techniques to leverage big data effectively; and absence of dynamic analytics algorithms for efficient data modeling. Authors presented a conceptual architecture of big data analytics, from developing multi-source data input, transformation, structure, management and analysis using traditional SQL, OLAP, to data mining.

Unintended consequences of ML (52)

Authors reviewed some of the factors driving wide adoption of deep learning and other forms of ML in the health ecosystem. They stated that ML has the potential of promoting changes in specialty that requires accurate prognosis models (e.g. oncology) and pattern recognition (e.g. radiology, pathology). However, authors anticipated some unintended consequences due to the application of ML-based decision support systems in healthcare, which includes overreliance on automation, more potential for decision errors by physicians, misleading diagnosis, misinterpretation of data, intrinsic uncertainty in medicine, inevitable intrinsic uncertainties and rationale and inscrutable outcome of ML algorithms. Authors presented unintended consequences in

clinical and operational research support in reducing the odds and better implementing ML in medicine. Authors appealed for developing more precise analytics platforms for pathology images, next-generation radiology tools, converting EHR into reliable risk predictors and monitoring patients' health through wearables and personal devices.

Finding the missing link for big biomedical data (53)

The problem of integration and analysis of large-scale biomedical data located in heterogeneous data sources can be solved by intelligently conforming biomedical data to support physicians and researchers in conducting new studies and drawing new hypothesis, leading to novel interventions. This study lists challenges in biomedical data linking, which includes identification of potential sources, determination of linking values, lack of a national unique patient identifier and data privacy and security concerns. The authors also presented some potential solutions to these challenges and support personalized treatment, urging the identification of potential sources of health information, considering non-traditional data (e.g. social media, purchase history, census records, etc.) to assemble a universal view of a patient and probabilistic linkage to the unique patient identifier problem. They recommended future (AI and ML) tools development to analyze biomedical data for better clinical decision-making.

Ethical challenges of implementing ML in healthcare (58)

The current trend towards AI and ML algorithm development and application comes with its set of issues and challenges. Despite the helpful contributions of AI and ML in healthcare, they have some concerns, especially related to the ethical challenges, which include unexpected risks due to the utilization of algorithms in medicine originally proposed for some other development; likelihood of using ML as the source of the communal medical mind; mirroring human prejudices in decision-making by learning from different unforeseen biases; in extreme premature conditions relying on ML results can lead to fatal conclusions; mostly applied ML algorithms are designed to perform in unethical conditions; ethical strain of gaining industrial profits with clinical decision-support systems without informing its users; increase in ethically problematic outcomes due to constructed ML black boxes; and reimagining of confidentiality due to ML. A review justifies these limitations in AI and ML with some examples, e.g. Uber's software tool Greyball and Volkswagen's algorithm. Authors are hopeful that these issues will be addressed in future ML-based healthcare systems, with the involvement or professionals

and researchers from policy enactment, programming, task-forces (59), etc.

Theoretical background of AI, ML and examples in healthcare

AI aimed to improve the intellectual capabilities and performances of machines to solve complex and big data-oriented problems by classifying interaction patterns among variables, learning from experiences, strategizing and predicting better orientations. AI has been in business for almost over 50 years, and its applications are heavily demanded in numerous fields of life, science, technology and medicine (68,69). AI developments are based on supervised, unsupervised and reinforced learning principles, which include computational command line, desktop, web-based, robotics and smartphone applications with different analytics capabilities, e.g. machine translation, speech recognition, NLP, data mining, risk modeling, image recognition, machine vision, knowledge bases, expert systems and agent-based systems. AI is categorized in four main types for decision-making, which are reactive machines (based on the current situation without learning from experiences), limited memory (based on short memory, and learning from experiences), theory of mind (based on humanlike capabilities and abilities to attribute mental states) and self-awareness (based on human-level consciousness).

In the last few years, AI has become more popular and seriously considered for analyzing diverse clinical data (EHR, images, etc.) for accurate diagnosis and effective treatment in different practices (70), e.g. radiology (71) (early diagnosis, enhance visualization of pathologies and predicting emergency situations (64,73,71)), oncology (diagnosis of breast (75), skin (76), lung cancer (77)), cardiology (interpreting electrocardiogram readings, echocardiography with 3D cardiac imaging, cardiac CT angiography for calcification of the coronary vessels, cardiac MRI for measuring perfusion and blood flow and longitudinal evaluation to find predictors of heart failure (78,79,80)), gastroenterology (analyzing endoscopic images for screening regimens to abnormal findings (81)), ophthalmology (detection of diabetic retinopathy in retinal fundus photographs (64)), pediatric (augmenting diagnostic evaluations (82)) and surgery (robotic-assisted surgery (83,84)), but not limited to these (85). One of the most recent trends is the utilization of AI in Precision Medicine (86,87,88) with the application of ML algorithms for analyzing heterogeneous patient data, e.g. clinical, genomics, metabolomics, imaging, claims, labs, nutrients and life-style.

ML is a branch of AI that utilizes and proposes different algorithms for learning from numerous data variables and
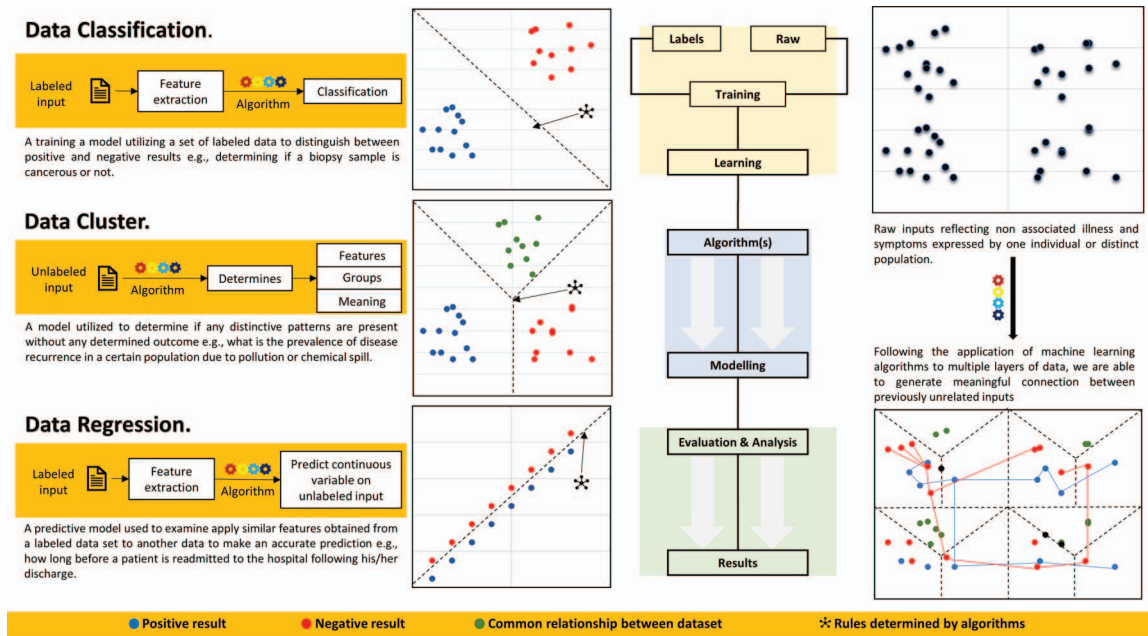
Figure 2. Data classification, clustering and regression for healthcare data analytics. ML application process includes creating and labeling of raw data, training classifier for data modeling using appropriate algorithm and analyzing and reporting results.

revealing multifaceted relationships among data features to predict accuracies in different contexts and support decision-making processes. The overall ML process starts with the manifestation of either or both data and labels, training classifier to learn and model using an algorithm and then performing data evaluation and analysis to estimate final results. ML is mainly categorized in three learning approaches: classification, cluster, and regression (Figure 2). Classification and regression are based on supervised learning, while clustering involves unsupervised learning. Classification predicts discrete, categorical response values by using labels and parameters, e.g. determining if a biopsy sample is cancerous or not (positive/negative). Cluster is to partition data into sub-groups, e.g. what is the prevalence of disease recurrence (positive/negative) in a certain population due to pollution or chemical spill (common relationship between dataset). Regression predicts continuous-response numeric values to identify distribution trends, e.g. how long before a patient is readmitted to the hospital following his/her discharge (positive/negative).

ML is becoming the transformational force in healthcare (89) for guiding individual and population health with several computational benefits, which includes real-time patient monitoring; disease patterns analysis (90); disease diagnosis and prescription of medicines; patient-centric care provision with enhanced treatment (91); clinical errors reduction (92); prognostic scoring (93); therapeutic decision-making (94); identification of sepsis and high risk for medical emergencies (95); identification of phenotypes; screening claims data (96); extraction of clinical codes

from death certificates and autopsy reports; identification of heart failure, cancer and other chronic disease causing symptoms; risk predictions, interventions, paneling and resourcing (97,98) and clinical decision-making (99). Most commonly used ML algorithms in medicine includes SVM, deep learning, logistic regression, DA, decision tree, random forest, linear regression, Naïve Bayes, K-nearest neighbor (KNN) and hidden Markov model (HMM) (Figure 3).

SVM (100) is one of the most widely used ML algorithms in bioinformatics and healthcare (101). It is considered a great supervised learning method (102,103) for accurately working in general practice. SVM was proposed in 1963 (100) to model diverse and high-dimensional data (101) with kernel methods to generate nonlinear decision boundaries and train classifier (104,105,106). It assists in the field of medicine with vast variety of contributions, e.g. symptoms classification and analysis to improve diagnostic accuracy, identifying imaging biomarkers of neurological and psychiatric disease, validation for multiple myeloma and breast cancer by analyzing SNPs, hold-out for breast and cervical cancer by analyzing clinical, pathologic and epidemiologic data, validation for oral cancer by analyzing clinical, molecular and genomic data, and diagnosing mental illness (106,107,108). Its ML benefits include modeling nonlinear class boundaries, unlikely overfitting, reduced computational complexity to quadratic optimization problems and controllable complexity of decision rule and frequency of error (109). However, the complex structure of this algorithm is its limitations leading to slow data processing speed, and difficulties in determining optimal
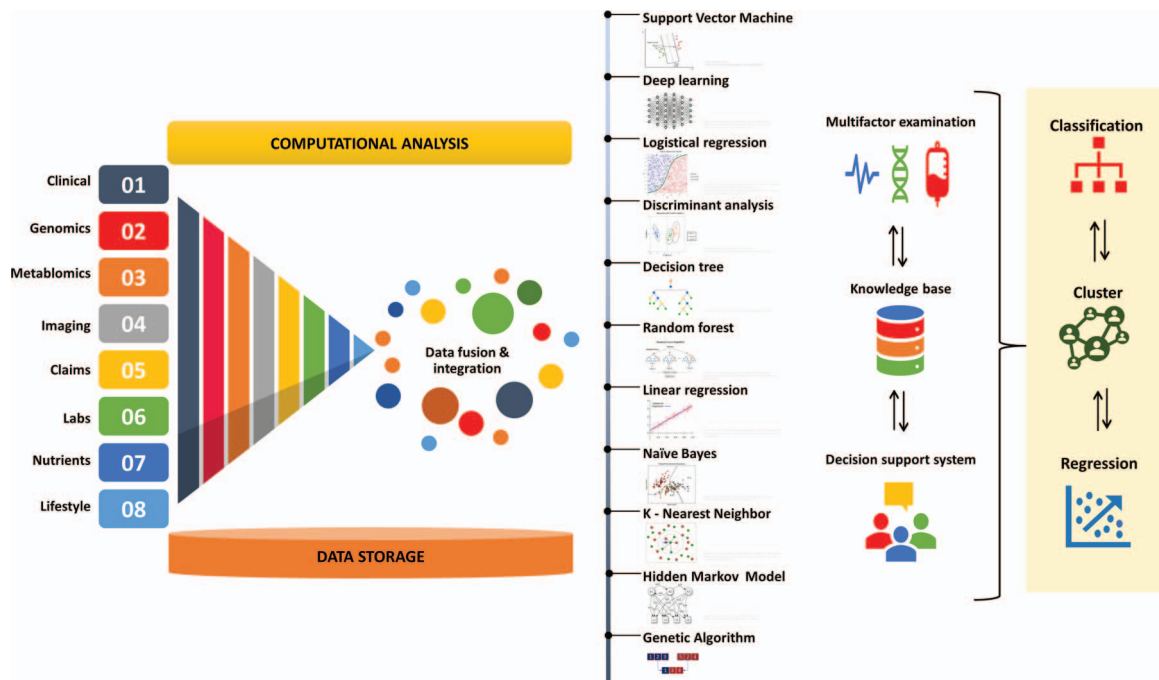
Figure 3. Applying machine learning algorithms for clinical, genomics, metabolomics, imaging, claims, labs, nutrients and life style data fusion, integration and analysis. Machine learning algorithms include, support vector machine, deep learning, logistic regression, discriminant analysis, decision tree, Random forest, linear regression, naïve Bayes, K-nearest neighbor, hidden Markov model and genetic algorithm.

parameters especially when training dataset is not linearly classified (108).

Deep learning is a dominant approach based on artificial neural network. It was inspired by the ability of the brain to learn complicated patterns, and one of the most popular ML algorithms in healthcare today, aiming to advance clinical medicine and care delivery (47). The artificial neural network base gives its several nested layers of neurons (110) to learn complex relationships between features and labels from heterogeneous clinical data (32). Deep learning has four categories: deep belief network, deep neural network, convolutional neural network and recurrent neural network (39) for implementing pattern recognition and predictive modeling at high-dimensional big data sets. It is fundamentally a uniquely different paradigm in ML (111,112) and its capabilities include multitasking, automatic construction of complex features, digitization of EHR and image-based data, integrating heterogeneous data sets assembled from diverse sources, combining with wearables for remote monitoring (47). Its ML benefits include application in classification or regression with ability to represent Boolean functions (AND, OR, NOT), handling noisy inputs and classifying instances for more than one output. However, its limitations include difficulties in understanding structures of algorithms, possibility of too many overfitting attributes and its optimal network structure which can only be created by experimentation (108). Deep learning is one of the most

widely used algorithms in medicine for analyzing different types of images from various healthcare disciplines but especially oncology, e.g. thoracic (lung cancer); abdominal and pelvic (computed tomography (CT) and magnetic resonance imaging (MRI)); colonoscopy (colonic polyps); mammography (microcalcifications); brain (brain tumors); radiation oncology (segmenting tumors for radiation, and quantifying specific radiographic characteristics by analyzing 3D shape of a tumor (113,114)); dermatology (skin cancer (115,76)); pathology (digital whole-slide of biopsy samples) (115); prostate (cancer tumors by analyzing ultrasound of biopsy cores); malignant lung nodules by analyzing radiographs; glioma by analyzing histological imaging and genomic marker data; and DNA and RNA sequencing (RNA-binding and DNA-binding proteins). Furthermore, deep learning has been applied for the diagnosis of several other diseases, e.g. nodular BCC, dermal nevus and seborrheic keratosis in dermatopathology (116); diabetic retinopathy (64); type 2 diabetes subgroups (66); diabetic retinopathy by analyzing retinal images; histological prediction in women with cytological abnormalities; measure medication adherence via camera interface (65); and cardiac anomalies and congestive heart failure by analyzing images of MRI of heart ventricles (117).

Logistic regression is a statistical method to assess the relationships between various predictor categorical and continuous variables and dichotomous binary outcome

(118). It has been applied in the fields of medicine (clinical practice, surgery), epidemiology and biochemistry to perform predictive and explanatory modeling by obtaining odds ratios and risk factors explaining variations in specific outcomes (119,120,121,122). Its examples of real-time implementation in the field of medicine include risk assessment of complex diseases (123) (e.g. tuberculosis (124), breast cancer (125)), predicting patient survival rate and diagnosing coronary heart disease (CHD) (123), non-Hodgkin's lymphoma diagnosis with multivariable logistic regression modeling (126), and identification of pulmonary thromboembolism (PTE) by analyzing prognostic factors (127). Its ML benefits include concurrent multiple explanatory variables analysis with reduced effect of confounding factors and modeling categorical dependent variables (128,129). However, its limitation lies in handling continuous explanatory variables with more than two levels as it is based on variables with a constant range of values (122), difficulties in understanding odds and probabilities (119), right predictor variable selection (118,122), reference group setup and managing relationships between input and output variables.

Discriminant analysis is a widely applied technique in medical studies for pattern recognition (130). Discriminant analysis is used to predict and classify group members by building one or multiple functions (131,132,133) based on normal distribution and equal variance–covariance of independent variables (134,135). Its real-time applications include identifying surgical and operative factors to accurately classify patients for surgical procedure (136), predict the clinical diagnosis of primary immunodeficiencies (137), patients' symptom-relief satisfaction data (138), prediction of depression elements in cancer patients (135), classification of BOLD fMRI response to naturalistic movie stimuli (139), identify protein coding regions of rice genes (140), Parkinson's disease symptoms recognition (141), risk assessment for chronic illnesses (142), diagnosis of hypercalcemia (143) and predicting patient care visits by identifying discriminatory characteristics (144). Its ML benefits include robustness, reduced dimensionality, easy implementation as it requires fewer parameters to be estimated (137). However, its limitations include over-fitting to dataset (131), limit performance at novel datasets (145) and lack of cross-validation (137).

The decision tree utilizes a tree structure modeling approach with conditional control statements for establishing an efficient decision-making process (146). The decision tree is based on the concept of classification rules, following paths from root to leaf. Its internal nodes represent 'test' on an attribute, branch represents the outcome of the test and leaf represents decision taken after computing all attributes. Its ML benefits include ease of understanding,

order of training algorithm instances with no effect, and no overfitting problem while pruning, allowing predictive model implementation with high precision, permanence and ease of clarification, well-mapped non-linear relationships and suitability for both classification or regression problems. However, its limitations include mutually exclusive classes, dependency on order of attribute selection, being error-prone when training on excessively complex decision tree, time-consumption and branching of missing values for an attribute (108). It has been well applied in the field of medicine for real-time healthcare monitoring, medical decision support system, anomaly detecting and sensor and a data mining model for pollution prediction. A few real-time examples include supporting clinical decisions (147), strategies for alternating therapies in oncology patients (148), collaborative clinical decision-making in mental health care (149), identifying predictors of health outcomes (150), finding factors related to hypertension (151), discovering factors associated with pressure ulcers (PUs) among elderly people (152), identifying the potential recipients of telehealth services (153), patient data stratification for interpretable decision-making for precision medicine (154), content analysis for patient aids decision (155), diabetic foot amputation risk analysis (156) and support understanding of antenatal lifestyle interventions (157).

Random forest is also known as the random decision forest, a combination of algorithms to build predictive models for classification and regression problems (158). Its classifier generates a set of decision trees based on training of randomly selected subsets to aggregate the elects from different decision trees to get final object. Its ML benefits include overcoming the problem of overfitting, less variance, not requiring input data preparation, flexibility and high accuracy even with missing large portions of the data. However, its limitations include its complexity, difficulty in implementing, requirements for additional computational resources, less intuitiveness and more time consumption than most other algorithms. Nevertheless, it has been applied in the field medicine for data mining, real-time patient monitoring, disease classification, implementation in wearables and personal devices and modeling big data based on engine recommendations. Some of its reported contributions include diagnosing mental illness (106), detecting knee osteoarthritis (159), monitoring medical wireless sensors (160), diagnosing Alzheimer's disease (161), predicting metabolic pathways (162), predicting outcomes of a patient's encounter with behavioral health providers (163), healthcare cost prediction (164), mortality prediction for intensive care unit (ICU) patients (165), classification of Alzheimer's disease (166), identifying social and economic factors to study social determinants

of health (167), predicting disease risks from imbalanced data (168), identifying associates of diabetic peripheral neuropathy diagnosis (169), predicting the risk of emergency admission (170), detecting patients ready to discharge from intensive care (171), nonparametric estimation of heterogeneous treatment effects (172), diagnosing sleep disorders (173) and predicting depression in patients suffering from Alzheimer's disease (174).

Linear regression is an ML approach to model relationships between dependent and independent variables using linear predictor functions to identify errors of prediction in a scatter plot and characterize relationships among multiple factors (175). Linear regression has been applied in the field of medicine for many computational analyses and predictions, from identification of prognostically relevant risk factors (175), predicting hand surgery (176), monitoring treatment prescribing patterns and ensuring its appropriateness (177), averaging decision-making in healthcare (178), reducing high costs of the health system (179), analyzing skewed healthcare cost data (180) and understanding human immunodeficiency virus (HIV) prevalence patterns (181).

Naïve Bayes is a supervised ML technique based on Bayes' theorem for data mining, classification, and predictive modeling (182) to find a maximum probability value from a conditional probability chain (183). Naïve Bayes is well applied in the field of health performing predictive modelling for different diseases (brain, asthma, prostate, and breast cancer etc.) (184), predicting risk using censored and time-to-event data (185), detecting Mucopolysaccharidosis type II (186), predicting Alzheimer's disease from genome-wide data (187), measuring quality healthcare services (188), finding audit targets in performance-based financing in health (189), modeling medical diagnosis for decision support (190), classifying EHR (191), classifying and genetic scoring for risk prediction (192), and designing a decision support system for heart disease (193). Its ML benefits include modelling based on statistical foundation, ease to understand and train algorithm, and usefulness across multiple domains. However, its limitations include difficulties in handling redundant attributes, distribution of statistically independent attributes, and management of class frequencies affecting accuracy (108).

KNN is also a supervised learning-based algorithm used for classification and regression in pattern recognition, data mining and intrusion detection by classifying points to given categories from a training dataset (194). Its ML benefits include implementation of instances for fast non-linear data classifications, robustness in addressing irrelevant or novel attributes, well-handled instances with noise and missing attribute values and applicability for both regression and classification. However, its limitations include languidity in

updating description concept, expecting similar classifications and relevancy from instances with similar attributes and increase in computational complexity with the number of attributes (108). KNN has been used in many scientific fields but with limited applications in the field of medicine (195). Some of real-time examples include modeling diagnostic performance (195), preserving privacy of medical diagnosis in e-Health cloud (196), medical dataset classification (197), classification of lymph node metastasis in gastric cancer (198), pattern classification for breast cancer diagnosis (199), pattern classification for health monitoring applications (200) and pancreatic cancer prediction combining published literature and EHR data (201).

HMM was originally proposed to solve speech problems by making complex and instinctive probabilistic models for finding and processing hidden states and paths (202,203, 204). Since the late 1980s, it has been effectively applied in the field of life sciences, especially in biology for analyzing sequence data (e.g. predicting exons and introns, identifying ORFs, insertions, deletions, substitutions, functional motifs in proteins, aligning two sequences and switching from exon to intron in a DNA sequence (205)) by capturing hidden information from observable sequential symbols. Later, it was well adapted in the field of medicine and its real-time contributions include modeling 'Healthy' and 'Unhealthy' unobserved health states (206); analyzing time-series data on personal health check-up (206); improving length of hospital stay prediction and reducing health care costs (207); mining adverse drug reactions from online healthcare forums (208); monitoring, modeling and clustering medical inpatient journeys (209); analyzing healthcare service utilization after transport-related injuries (210); monitoring circadian in telemetric activity data (211); predicting patients entering states with a high number of asynchronies (212); and analyzing subject-specific seizure and automatic segmentation of infant cry signals (213). However, its modeling limitations include computing probability of sequence observation, choosing an accurate corresponding state, adjusting parameters (214) and dealing with correlations between residues due to the underlying decency assumption problem (215).

The genetic algorithm was inspired by Charles Darwin's theory of natural evolution to solve constrained and unconstrained data optimization and standardization problems by repeatedly modifying a population of individual solutions (215). Primarily, it is based on heuristic search with three active rules: selection of data elements (parents), crossover rules for two parents from children and mutation for random change (216). Its ML benefits include easier implementation than other algorithms; application for feature classification, selection and optimization; and relative success. However, its computational limitations include

development of non-trivial scoring function, complications in training classifiers of given data, and its being not the best method to find optima. It has been vigorously involved in the fields of life and medical sciences for benefitting analytics, development in radiology, oncology, cardiology, endocrinology, pediatrics, surgery, pulmonology, infectious diseases, radiotherapy, rehabilitation medicine, orthopedics, neurology, pharmacotherapy, health care management, obstetrics and gynecology (217). It has been successful in detecting microcalcifications in mammograms of breast cancer (218,219), developing a non-invasive technique for cervical cancer detection (220), analyzing microarray data from cancer cell lines (221), investigating relationships between soil trace elements and cervical cancer mortality (222), parameter estimation for determining tissue elasticity (223), predicting risk of a major adverse cardiac event (MACE) (224), detecting QRS complexes (225), detecting hypoglycemia EEG signals (226), predicting time to reach full cervical dilation (227), selecting optimal features for cardiotocogram recordings (228), identifying autism by analyzing gene expression microarray data (229), predicting outcomes of patients with non-small cell lung cancer (NSCLC) (230), diagnosing patients by classifying lung sounds into normal, wheeze and crackle (231), choosing appropriate highly active antiretroviral therapy (HAART) to control HIV (232), improving the selection of gantry angles to optimize stereotactic radiotherapy (233), training robots for physiotherapy of the lower limb (234), estimating Cobb angle from torso asymmetry in scoliosis (235), analyzing mutations in Parkinson's disease (236), predicting tacrolimus blood levels and scheduling patient admission in an ophthalmic hospital (237).

Discussion

AI and its application in healthcare could be another great leap in medicine and a transformational force for guiding personalized and population medicine with several computational benefits. The extent of its popularity in healthcare can be easily determined by the number of AI related published work in medicine. At the time of the study (09 April 2019), a total of 16 166 AI and ML papers were available through PubMed. The growth of scientific literature in AI increased in the last 10 years with 14 469 papers (2009–2019), and more than 70% work was published in the last five years. While the growing importance and relevance of AI in healthcare is indisputable, to improve public sector clinical practice, there is a critical need for development of intelligent frameworks to connect operational and analytical healthcare systems in a way that experts from multiple domains can perform measurement and predictive analysis. AI has the potential to play a vital role at various levels

of clinical operations, research and analytics to achieve significant improvements in providing better individualized and population healthcare at lower costs, and together with better work-life style for clinicians and staff. The major barriers in successfully implementing AI in healthcare include gaps in AI building blocks and infrastructures; low quality data availability; business model sustainability; regulations and policies for data collection, usage and sharing; disruption in the physician–patient relationships; integration to operational health systems; reduced evidence and reproducibility; selection of most appropriate ML algorithm; lack of understanding of AI and ML processes to predict; hazard of dehumanization of healthcare data and job insecurity threat; conflicts of interest and impartial access; accountability exploitation of AI; 'sanity' check to minimize any bias; handling of misleading and erroneous results; and data privacy, ethics, consent and ownership (108,238,239,240).

Determining which AI approach to use for which task is a challenge in itself. Classifying tasks based on available predictor variables is a key step to correctly addressing the problem. Traditional AI models can be used for simple prediction tasks while complex tasks require more complex models. The next step is to consider how the model will be used in practice. Setting rules to identify if a task needs to pre-process data before execution or if a complex task can be broken down into simple tasks is amenable to a traditional model. Creating and updating these rules is very time-consuming but is very useful. Availability of training data is another key factor in intelligent automation. Simple models can work with little data and few variables, but complex models require huge amount of data with multiple examples and scenarios as it is expected to remove noisy data and learn to identify complex statistical patterns (32). Numerous approaches have been proposed recently to successfully target current challenges of implementing AI and ML. We studied and reviewed contributions and variability analysis of various approaches in healthcare. We defined 15 different features to assess the potential of discussed approaches (Table 3). These features are (i) intelligent interface development; (ii) next-gen radiology and imaging tools development; (iii) global expansion of medical resources; (iv) automated ETL, linkage and data mining in HER; (v) risk prediction and containment of antibiotics resistance; (vi) pathology image analysis; (vii) AI in machines and medical devices; (viii) smart solutions and methods for cancer treatment; (vix) EMR analysis for accurate risk predictors; (x) wearable devices for monitoring patients health; (xi) smartphone applications as diagnostic tools; (xii) AI-based clinician decision-making; (xiii) search engine for healthcare data flow; (xiv) data privacy and security; and (xiv) personalized treatments.

Table 3. Variability analysis of reviewed approaches.

Literature/features	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Fogel *et al.* 2018 (31)															
Rajkomar *et al.* 2019 (32)															
Nayak *et al.* 2016 (33)															
Jones *et al.* 2018 (34)															
Redekop *et al.* 2018 (35)															
Mesko *et al.* 2017 (36)															
Wahl *et al.* 2018 (37)															
Sitapati *et al.* 2017 (38)															
Jiang *et al.* 2017 (39)															
Emmert-Streib *et al.* 2019 (40)															
Grusona *et al.* 2019 (41)															
Mesko *et al.* 2018 (42)															
Stein *et al.* 2019 (44)															
Kantarjian *et al.* 2015 (45)															
Hersh *et al.* 2013 (46)															
Naylor, 2018 (47)															
Hinton, 2018 (48)															
Topol, 2019 (49)															
Acs *et al.* 2018 (50)															
Rose, 2018 (51)															
Cabitza *et al.* 2017 (52)															
Weber *et al.* 2018 (53)															
Wong *et al.* 2018 (54)															
Saheb *et al.* 2019 (55)															
Ngiam *et al.* 2019 (56)															
Abidi *et al.* 2019 (57)															
Char *et al.* 2018 (58)															
Gruson *et al.* 2019 (59)															
Crown, 2019 (60)															
Raghupathi *et al.* 2014 (61)															
Azencott, 2018 (62)															
Kourou *et al.* 2015 (63)															

It is based on 15 different features, which includes 1: intelligent interface development; 2: next-gen radiology and imaging tools development; 3: global expansion of medical resources; 4: automated ETL, linkage and data mining in HER; 5: risk prediction and containment of antibiotics resistance; 6: pathology images analysis; 7: AI in machines and medical devices; 8: smart solutions and methods for cancer treatment; 9: EMR analysis for accurate risk predictors; 10: wearable devices for monitoring patients health; 11: smartphone applications as diagnostic tools; 12: AI-based clinician decision-making; 13: search engine for healthcare data flow; 14: data privacy and security; 15: personalized treatments. Color coding, red represents 'absence', and blue represents 'presence' of respective feature discussion

Although many AI and ML-based approaches have embraced the data gold rush in biomedicine, many concerns remain, which include integration of biomedical data located in heterogeneous data sources; handling extensively available irrelevant, error-prone and missing data; unnecessary follow-up diagnosis and treatment as a result of overloaded health information; loss of data privacy and reimagining of confidentiality; ethical strain of gaining industrial profits with clinical decision-support systems without informing the users; possibilities of patient data exploitation, employment insecurity due to governance and management of big data; regulation of healthcare data analysis according to principles of bioethical law; evidence-based observational data analysis and screening, slow data processing speed and difficulties in determining optimal parameters; difficulties in understanding structures of algorithms; possibility of too many overfitting attributes and its optimal network structure; handling continuous explanatory variables with more than two levels; difficulties in understanding odds and probabilities (119); correct predictor variable selection, reference group setup, and managing relationships between input and output variables (118,122); error-prone branching of missing values for an attribute (108); lack of data availability on social determinants of health; difficulties in handling redundant attributes, distribution of statistically independent attributes and management of class frequencies affecting accuracy (108); increased computational complexity with increasing number of attributes; computing probability of sequence observation; choosing accurate corresponding states and adjusting parameters (214); dealing with correlations between residues due to underlying decency assumption problem (215); and prediction modeling with biases in confounding, causal inference, complexity-based model selection, benchmark development and pragmatic interoperability including reproducibility and generalizability.

Potential solutions to these concerns are based on design and implementation of healthcare IT infrastructure (35,49); EMR documentation and automation of repetitive tasks with ML algorithms (31); deep learning to mine (41), train (39) and learn complex relationships between features and labels (32); ML algorithms to effectively link data between different platforms (33); development of effective clinical decision support systems (CDS) through the utilization of ML algorithms (38); translation of local languages into EHRs and cloud-based data sharing (37); deep learning for integrative EMR analysis from diverse sources (47); computer vision algorithms to identify accurate indigestion of medications for patient (31); deep learning for speech recognition, image interpretation and language translation (48); AI algorithms for diagnostic disease modeling and

computer-aided design (CAD) (34); ML models to create knowledge base systems of phenotypes (40); ML algorithms to predict outbreak patterns and surveillance for new emergence (37); integrative approach to help decision-making processes between physicians and AI (89); ML algorithms to enhance and optimize cancer treatment and development of new drug treatments (45); ML algorithms to perform longitudinal population studies for analyzing the effects of treatments (56); ML-based hybrid model classifier to enhance overall healthcare predictability (72); and ML algorithms to transform clinical research into a much higher capacity and lower cost information processing care service (74).

We also considered most adaptable traditional (statistical) AI and ML algorithms (SVM, deep learning, logistic regression, discriminant analysis, decision tree, random forest, linear regression, naïve Bayes, KNN, HMM and genetic algorithm) in healthcare and justified their contributions with living examples for different clinical applications, e.g. deep neural network for diagnosing eye diseases (diabetic retinopathy) by analyzing retinal images; diagnosing cardiac anomalies by analyzing MRI images of heart ventricles; detecting malignant lung nodules by analyzing radiographs; grading prostate cancer tumors by analyzing ultrasound of biopsy cores; and producing glioma survival predictions by analyzing histological imaging and genomic marker data. Some other non-traditional algorithms were used, e.g., 'Watson for Oncology' (241,242) and 'CURATE.AI' (243,244), for analyzing EHR of cancer patients and recommend treatments and drug doses; TREWScore (245,246,247,248) for analyzing EHR for predicting septic shock risk and analyze CT brain scans for three-dimensional convolutional neural network neurological disorder classification; 'PowerLook Density Assessment 3.4' (249) for analyzing mammogram images for breast density assessment; 'OsteoDetect' (84,250,251) for detecting distal radius fractures in the wrist; cloud-based deep neural network algorithm for diagnosing cardiac anomalies and segmentation of lung and liver tumors by analyzing CT scans (56); and AI system for breast cancer prediction (252). Furthermore, major clinical oncological developments with the implementation of different AI and ML applications and algorithms include Chatbot; breast MRI interpretation; breast lesion classification; colorectal polyp classification; identification of colorectal cancer biomarkers in cell-free blood assays; gastric mucosal disease classification; detection of lung nodules in low-dose lung CT screening; prediction and evaluation of risk and malignancy and classification of dysplastic nevi, spitz nevi and basal and squamous cell carcinoma; detection of esophageal cancer; and identification of pancreatic cancer biomarkers (56).

The goal of this study is to highlight recent contributions and effectiveness of AI and ML in the development of computational systems towards better healthcare and precision medicine. We have reviewed and discussed multiple AI and ML-based approaches (Tables 1, 2 and 3) and algorithms with variable colors. Deep learning (253,254) has proven to be one of the most trending algorithms today, but this does not undermine the importance of other machine learning algorithms. We believe that a right approach and algorithm should be chosen for the development of the most effective solutions to the targeted problems. In spite of various traditional and AI-based solutions, current limitations and challenges by the healthcare community include uneven distribution of resources towards the future of digital healthcare; business model unsustainability; disruption in the physician–patient relationships; hazards of dehumanization of healthcare data and job insecurity threat; prevention of early adopters for dragging down to the lowest common technology denominator; inconvenient adoption of digital processes for healthcare interoperability; shortages in development of customized IT infrastructure for data science; global unavailability of de-identified healthcare data for research; understanding of how healthcare data of different people relates to one another; need to harmonize big data with the definitions of the clinical phenotypes and diagnosis; limited EHR systems for assimilating operational and analytics interests; EHR offering limited patient participation; unavailability of high-quality open-source EHR systems inviting third-party extensions and assertions; and inflexible EHR database schemas not geared for precision medicine.

To effectively meet the goals of healthcare data analytics, while dealing with the aforementioned community, and traditional and AI-based challenges, significant efforts are required from experts in multidisciplinary sciences. To facilitate and improve public sector clinical research and practice, there is a critical need for academic frameworks that can connect operational and analytical systems in a way that experts from multiple domains can perform measurement and descriptive analysis, even without strong computational background. There is a need to develop standalone, user-friendly, standardized, open-source, and comprehensive solutions, which implement healthcare data analysis by connecting all kinds of patient data generated from any of the existing commercial EHR systems at any level, which includes patient's demographic information, personal life style, medical history, recent visits to the practices, providers attended, diagnosis performed, lab tests conducted, longitudinal images, medications and procedures, samples taken for wet and dry lab experimentations for research and treatment of disease with no cure. Furthermore, the ideal system should be capable of automatically linking and communicating with similar systems.

Unmet clinical research and operational data analytics needs development of intelligent and secure systems to support practice transformation for implementing precision medicine at a global level. Overarching goals include new multi-functional platforms founded on the clinical, AI and scientific premise that integrates and analyzes heterogeneous clinical data received from multiple platforms. We need to implement and train AI classifiers at available structured clinical dataset of over a million subjects to help in early detection and diagnosis of common, frequently occurring and rare diseases and predicting the performance of provided treatments. We need to further explore AI methodologies to design models segregating disorders, identify causative medical conditions and determine the best drug therapies, especially when adjusted for age, race and gender. The goal should be to accelerate clinical care and discovery by satisfying research aims, improving quality and transition of care, obtaining actionable care gap-based information about patients and developing communication and coordination across hospitals, specialists, community-based providers, sub-acute care, nurses, quality inspectors, management, researchers and analysts. We need to strive to address possible challenges that continue to slow the advancements of this breakthrough treatment approach.

Along with AI- and ML-based methodological developments, it is important to address the issues related to healthcare data privacy and security. In most of the cases, academic and applied research environments do not have access to the healthcare data. The major reason is secure handling of protected health information (PHI) and lack of trust by the healthcare institutions in providing access to their medical records. We need to develop secure research-based HIPAA-compliant frameworks for efficient PHI storage, pre-processing, de-identification and integration to serve a large community of users, support organizational policies and provide efficient access and connectivity. It is mandatory to implement HIPAA rules for system user data security, which includes application and data criticality, risk management and analysis, information system activity review, contingency plan, device and media controls, disaster recovery plan, data backup plan, emergency mode operations plan, device and media controls, access controls, security incident procedures, vulnerability assessment, penetration testing, physical security, business associate agreements, polices and procedures, evaluation, audit, assignment of responsibility sanctions, workstation use and security. Furthermore, it is important to provide an external layer to the overall healthcare data analytics systems for placing clinical data in a distributed centralized structure; administering security by encrypting data as well

as offering a multi-user-based graphical interface with controlled access; and supporting data backup with redundant data management plans.

Conclusion

Precision medicine is progressing but with many challenges lying ahead (255), which require addition of useful analytic tools, technologies, databases and approaches (4,6) to efficiently augment networking and interoperability of clinical, laboratory and public health systems, as well as address ethical and social issues related to the privacy and protection of healthcare and omics data with effective balance. This will also require more efficient management of massive amounts of generated data, as well as earlier mined consensus and actionable data. Most efforts involved currently are manual and time-consuming, whether it is extraction of healthcare data from operational clinical systems, identification of common and rare functional variants, metabolite penetrance using listed features and abnormalities, examining relations between genomic variations and metabolite levels, analyzing biochemical pathways in metabolites with patterns of multimodal distributions for candidate genes and management and assimilation of healthcare, along with epidemiological and omics data generated at each step of entry, production and analysis. Cutting-edge, new AI and ML-based big data platform development has the potential to revolutionize the field of medicine and improve the quality and transition of healthcare by intelligently analyzing structured clinical data available in great count and volume, posing unprecedented challenges in data storage, processing, exchange and curation, and developing a better understanding of biology.

Supplementary data

Supplementary data is available at *Database* online.

Author contributions

ZA proposed and lead the study. ZA, KM, and SZ conducted review research, and evaluated reported findings. XD guided the study. ZA drafted the manuscript, and all authors participated in writing and review.

Acknowledgements

References

1. Ahmed,Z., Kim,J. and Liang,B.T. (2019) MAV-clic: framework towards management. Analysis and visualization of clinical big data. *J. Am. Med. Inf. Assoc. Open*, **2**, 23–28.

2. Makary,M.A. and Daniel,M. (2016) Medical error—the third leading cause of death in the US. *BMJ.*, **353**, i2139.

3. Ritchie, M. D., de Andrade, M. & Kuivaniemi, H. The foundation of precision medicine: integration of electronic health records with genomics through basic, clinical, and translational research. *Front. Genet.*, **6**, 104 (2015).

4. Sboner,A. and Elemento,O. (2016) A primer on precision medicine informatics. *Brief. Bioinform.*, **17**, 145–153.

5. Zeeshan,S., Ruoyun,X., Liang,B.T. and Ahmed,Z. (2019) 100 years of evolving gene-disease complexities and scientific debutants. *Brief. Bioinform.*, bbz038.

6. Karczewski,K.J. and Snyder,M.P. (2018) Integrative omics for health and disease. *Nat. Rev. Genet.*, **108**, 1111.

7. Marx,V. (2013) Biology: the big challenges of big data. *Nature*, **498**, 255–260.

8. Combi,C. (2017) Editorial from the new editor-in-chief: artificial intelligence in medicine and the forthcoming challenges. *Artif. Intell. Med.*, **76**, 37–39.

9. Ahmed,Z. and Liang,B.T. (2019) Systematically dealing practical issues associated to healthcare data analytics. In: *Lecture Notes in Networks and Systems 69 Springer Nature*.

10. Beam,A.L. and Kohane,I.S. (2018) Big data and machine learning in health care. *JAMA*, **319**, 1317–1318.

11. Raghupathi,W. and Raghupathi,V. (2014) Big data analytics in healthcare: promise and potential. *Health Inform. Sci. Syst.*, **2**, 3.

12. Alyass,A., Turcotte,M. and Meyre,D. (2015) From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Med. Genom.*, **8**, 33.

13. McShane,L.M., Cavenagh,M.M., Lively,T.G. *et al.* (2013) Criteria for the use of omics-based predictors in clinical trials. *Nature*, **502**, 317–320.

14. Berger,B., Peng,J. and Singh,M. (2013) Computational solutions for omics data. *Nat. Rev. Genet.*, **14**, 333–346.

15. Kim,M.O., Coiera,E. and Magrabi,F. (2017) Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. *J Am. Med. Inform. Assoc.*, **7**, 246–260.

16. Sligo,J., Gauld,R., Roberts,V. and Villa,L. (2017) A literature review for large-scale health information system project planning, implementation and evaluation. *Int. J. Med. Inf.*, **97**, 86–97.

17. Lu,Z. and Su,J. (2010) Clinical data management: current status, challenges, and future directions from industry perspectives. *Open Access J. Clin. Trials*, **2**, 93–105.

18. Haux,R., Knaup,P. and Leiner,F. (2007) On educating about medical data management the other side of the electronic health record. *Methods Inf. Med.*, **46**, 74–79.

19. Rumsfeld,J.S., Joynt,K.E. and Maddox,T.M. (2016) Big data analytics to im-prove cardiovascular care: promise and challenges. *Nat. Rev. Cardiol.*, **13**, 350–359.

20. Van Panhuis,W.G., Paul,P., Emerson,C. *et al.* (2014) A systematic review of barriers to data sharing in public health. *BMC Public Health*, **14**, 1144.

21. Fegan,G.W. and Lang,T.A. (2008) Could an open-source clinical trial data-management system be what we have all been looking for? *PLoS Med.*, **5**, e6.

22. Wang,X., Williams,C., Liu,Z.H. and Croghan,J. (2017) Big data management challenges in health research—a literature review. *Brief Bioinform*, bbx086.

23. Duffy,D.J. (2016) Problems, challenges and promises: perspectives on precision medicine. *Brief. Bioinform.*, **17**, 494–504.

24. Xu,R., Li,L. and Wang,Q. (2014) dRiskKB: a large-scale disease-disease risk relationship knowledge base constructed from biomedical text. *BMC Bioinform.*, **15**, 105.

25. Xu,R. and Wang,Q. (2015) PhenoPredict: a disease phenome-wide drug repositioning approach towards schizophrenia drug discovery. *J. Biomed. Inform.*, **56**, 348–355.

26. Brown,A.S. and Patel,C.J.M.S.H.D.D. (2017) Literature-based drug-drug similarity for drug repositioning. *JAMIA*, **24**, 614–618.

27. Luo,Y., Szolovits,P., Dighe,A.S. and Baron,J.M. (2017) 3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data. *J. Am. Med. Inform. Assoc*, ocx133.

28. Walker,J.G., Bickerstaffe,A., Hewabandu,N. *et al.* (2017) The CRISP colorectal cancer risk prediction tool: an exploratory study using simulated consultations in Australian primary care. *BMC Med. Inform. Decis. Mak.*, **17**, 13.

29. Liu,L., Liu,L., Fu,X. *et al.* (2017) A cloud-based framework for large-scale traditional Chinese medical record retrieval. *J. Biomed. Inform.*, **77**, 21–33.

30. Ahmed,Z., Zeeshan,S. and Dandekar,T. (2016) Mining biomedical images towards valuable information retrieval in biomedical and life sciences, Database baw118.

31. Fogel,A.L. and Kvedar,J.C. (2018) Artificial intelligence powers digital medicine. *NPJ Digit. Med.*, **1**, 5.

32. Rajkomar,A., Dean,J. and Kohane,I. (2019) Machine learning in medicine. *N. Engl. J. Med.*, **380**, 1347–1358.

33. Nayak, L., Ray, I. & De, R. K. Precision medicine with electronic medical records: from the patients and for the patients. *Ann. Transl. Med.* **4**(Suppl 1), S61 (2016).

34. Jones,L.D., Golan,D., Hanna,S.A. and Ramachandran,M. (2018) Artificial intelligence, machine learning and the evolution of healthcare: a bright future or cause for concern? *Bone Joint Res.*, **7**, 223–225.

35. Redekop,W.K., Bakker,L.J. and Aarts,J. (2018) Healthcare problems cannot be solved using health technologies alone: the example of precision medicine. *Health Policy Technol.*, **7**, 3–4.

36. Mesko,B. (2017) The role of artificial intelligence in precision medicine. *Expert Rev. Precis. Med. Drug Dev.*, **2**, 239–241.

37. Wahl,B., Cossy-Gantner,A., Germann,S. and Schwalbe,N.R. (2018) Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob. Health*, **3**, e000798.

38. Sitapati, A., Kim, H., Berkovich, B., Marmor, R., Singh, S., El-Kareh, R., Clay, B. & Ohno-Machado, L. Integrated precision medicine: the role of electronic health records in delivering personalized treatment. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **9**, 3 (2017).

39. Jiang,F., Jiang,Y., Zhi,H. *et al.* (2017) Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. Neurol.*, **2**, 230–243.

40. Emmert-Streib,F. and Dehmer,M. (2019) A machine learning perspective on personalized medicine: An automized comprehensive knowledge base with ontology for pattern recognition. *Mach. Learn Knowl Extr.*, **1**, 149–156.

41. Gruson,D., Helleputte,T., Rousseau,P. and Gruson,D. (2019) Data science, artificial intelligence, and machine learning: opportunities for laboratory medicine and the value of positive regulation. *Clin. Biochem.*, **69**, 1–7.

42. Meskó,B., Hetényi,G. and Győrffy,Z. (2018) Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Serv. Res.*, **18**, 545.

43. Bostrom,N. (2014) *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

44. Stein,J.D., Rahman,M., Andrews,C. *et al.* (2019) Evaluation of an algorithm for identifying ocular conditions in electronic health record data. *JAMA Ophthalmol.*, **137**, 491–497.

45. Kantarjian,H. and Yu,P.P. (2015) Artificial intelligence, big data, and cancer. *JAMA Oncol.*, **1**, 573–574.

46. Hersh,W.R., Weiner,M.G., Embi,P.J. *et al.* (2013) Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med. Care*, **51**, S30–S37.

47. Naylor,C.D. (2018) On the prospects for a (deep) learning health care system. *JAMA*, **320**, 1099–1100.

48. Hinton,G. (2018) Deep learning-a technology with the potential to transform health care. *JAMA*, **320**, 1101–1102.

49. Topol,E.J. (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.*, **25**, 44–56.

50. Acs,B. and Rimm,D.L. (2018) Not just digital pathology, intelligent digital pathology. *JAMA Oncol.*, **4**, 403–404.

51. Rose,S. (2018) Machine learning for prediction in electronic health data. *JAMA Netw Open.*, **1**, e181404.

52. Cabitza,F., Rasoini,R. and Gensini,G.F. (2017) Unintended consequences of machine learning in medicine. *JAMA*, **318**, 517–518.

53. Weber,G.M., Mandl,K.D. and Kohane,I.S. (2014) Finding the missing link for big biomedical data. *JAMA*, **311**, 2479–2480.

54. Wong,D. and Yip,S. (2018) Machine learning classifies cancer. *Nature*, **555**, 446–447.

55. Saheb,T. and Saheb,M. (2019) Analyzing and visualizing knowledge structures of health informatics from 1974 to 2018: a bibliometric and social network analysis. *Healthc. Inform. Res.*, **25**, 61–72.

56. Ngiam,K.Y. and Khor,I.W. (2019) Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.*, **20**, e262–e273.

57. Abidi,S.S.R. and Abidi,S.R. (2019) Intelligent health data analytics: a convergence of artificial intelligence and big data. *Healthc. Manage. Forum.*, **32**, 178–182.

58. Char,D.S., Shah,N.H. and Magnus,D. (2018) Implementing machine learning in health care - addressing ethical challenges. *N. Engl. J. Med.*, **378**, 981–983.

59. He,J., Baxter,S.L., Xu,J. *et al.* (2019) The practical implementation of artificial intelligence technologies in medicine. *Nat Med.*, **25**, 30–36.

60. Crown,W.H. (2019) Real-world evidence, causal inference, and machine learning. *Value Health*, **22**, 587–592.

61. Raghupathi,W. and Raghupathi,V. (2014) Big data analytics in healthcare: promise and potential. *Health Inform. Sci. Syst.*, **2**, 3.

62. Azencott,C.A. (2018) Machine learning and genomics: precision medicine versus patient privacy. *Philos. Trans. A Math. Phys. Eng. Sci.*, **13**, 376.

63. Kourou,K., Exarchos,T.P., Exarchos,K.P. *et al.* (2014) Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.*, **13**, 8–17.

64. Gulshan,V. *et al.* (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, **316**, 2402–2410.

65. Labovitz,D.L., Shafner,L., Reyes Gil,M. *et al.* (2017) Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. *Stroke*, **48**, 1416–1419.

66. Li,L. *et al.* (2015) Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.*, **7**, 311ra174.

67. Sheikhtaheri,A., Sadoughi,F. and Hashemi Dehaghi,Z. (2014) Developing and using expert systems and neural networks in medicine: a review on benefits and challenges. *J. Med. Syst.*, **38**, 110.

68. Buch,V.H., Ahmed,I. and Maruthappu,M. (2018) Artificial intelligence in medicine: current trends and future possibilities. *Br. J. Gen. Pract.*, **68**, 143–144.

69. Bali,J., Garg,R. and Bali,R.T. (2019) Artificial intelligence (AI) in healthcare and biomedical research: why a strong computational/AI bioethics framework is required? *Indian J. Ophthalmol.*, **67**, 3–6.

70. Mintz,Y. and Brodie,R. (2019) Introduction to artificial intelligence in medicine. *Minim. Invasive. Ther. Allied Technol.*, **28**, 73–81.

71. Prevedello,L.M. *et al.* (2017) Automated critical test findings identification and online notification system using artifical intelligence. *Radiol.*, **285**, 923–931.

72. Neesha Jothia,N.A. (2015) Data Mining in Healthcare - a review. *Procedia Comput. Sci.*, **72**, 306–313.

73. Bejnordi,B.E. *et al.* (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastasis in women with breast cancer. *JAMA*, **318**, 2199–2210.

74. Beam,A.L. and I,K. (2016) Translating Artificial intelligence into clinical care. *JAMA*, **316**, 2368–2369.

75. Somashekhar, S. P. *et al.* Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. *Ann. Oncol.* **29**, 1–6 (2018).

76. Esteva,A. *et al.* (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, **542**, 115–118.

77. Yu,K.H. *et al.* (2016) Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.*, **7**, 12474.

78. Slomka,P.J. *et al.* (2017) Cardiac imaging: working towards fully-automated machine analysis & interpretation. *Expert Rev. Med. Dev.*, **14**, 197–212.

79. Choi,E. *et al.* (2017) Using recurrent neural network models for early detection of heart failure. *J. Am. Med. Inform. Assoc.*, **24**, 351–370.

80. Seetharam,K., Shrestha,S. and Sengupta,P.P. (2019) Artificial intelligence in cardiovascular medicine. *Curr. Treat Options Cardio. Med.*, **21**, 25.

81. Karkanis,S., Magoulas,G. and Theofanous,N. (2000) Image recognition and neuronal networks: intelligent systems for the improvement of imaging information. *Min. Invas. Ther. Allied Technol.*, **9**, 225–230.

82. Liang,H. *et al.* (2019) Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat. Med.*, **25**, 433–438.

83. Herron,D.M., Marohn,M. and SAGES-MIRA Robotic Surgery Consensus Group (2008) A consensus document on robotic surgery. *Surg. Endosc.*, **22**, 313–315.

84. Craft,J.A. (2018) Artificial intelligence and the softer side of medicine. *Missouri Med.*, **115**, 406–409.

85. Galimova,R.M., Buzaev,I.V., Ramilevich,K.A. *et al.* (2019) Artificial intelligence-developments in medicine in the last two years. *Chronic Dis. Transl. Med.*, **5**, 64–68.

86. Mesko,B. (2017) The role of artificial intelligence in precision medicine. *Expert Rev. Precis. Med. Drug Dev.*, **2**, 239–241.

87. van Hartskamp,M., Consoli,S., Verhaegh,W. *et al.* (2019) Artificial intelligence in clinical health care applications. *Interact. J. Med. Res.*, e12100, 8.

88. Schork,N. and Artificial,J. (2019) Intelligence and personalized medicine. *Cancer Treat. Res.*, **178**, 265–283.

89. Mesko,B. (2019) Artificial intelligence is the stethoscope of the 21st century. *The Medical Futurist*. https://medicalfuturist.com/ibm-watson-is-the-stethoscope-of-the-21st-century. Updated July 18,2017. Accessed 23 August 2019.

90. Challen,R., Denny,J., Pitt,M. *et al.* (2019) Artificial intelligence, bias and clinical safety. *BMJ Qual. Saf.*, **28**, 231–237.

91. Kaura,P., Sharmab,M. and Mittalc,M. (2018) Big data and machine learning based secure healthcare framework. *Procedia Comput. Sci.*, **132**, 1049–1059.

92. Kaushal,R., Shojania,K.G. and Bates,D.W. (2003) Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Arch. Intern. Med.*, **163**, 1409–1416.

93. Bouch,D.C. and Thompson,J.P. (2008) Severity scoring systems in the critically ill. *Cont. Edu. Anaesth. Crit. Care Pain.*, **8**, 181–185.

94. Gianfrancesco,M.A., Tamang,S., Yazdany,J. and Schmajuk,G. (2018) Potential biases in machine learning algorithms using electronic health record data. *JAMA Int. Med.*, **178**, 1544–1547.

95. Sidey-Gibbons,J. and Sidey-Gibbons,C.J. (2019) Machine learning in medicine: a practical introduction. *BMC Med. Res. Methodol.*, **19**, 64.

96. Panch, T., Szolovits, P. & Atun, R. Artificial intelligence, machine learning and health systems. *J. Global Health* **8**, 020303 (2018).

97. Hippisley-Cox,J., Coupland,C., Vinogradova,Y. *et al.* (2008) Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ Clin. Res. Ed.*, **336**, 1475–1482.

98. Rajkomar,A., Yim,J.W., Grumbach,K. and Parekh,A. (2016) Weighting primary care patient panel size: a novel electronic

health record-derived measure using machine learning. *JMIR Med. Inform.*, **4**, e29.

99. Sullivan, T. Next up for EHRs: Vendors adding artificial intelligence into the workflow. Healthcare IT News. https://www.healthcareitnews.com/news/next-ehrs-vendors-adding-artificial-intelligence-workflow. Updated March 13, 2018. Accessed 23 August 2019. (2018).

100. Vapnik,V. (1963) Pattern recognition using generalized portrait method. *Autom. Remote Control.*, **24**, 774–780.

101. Noble, W. S. What is a support vector machine? *Nat. Biotech.* **24**, 1564–1567 (2006).

102. Kaura,P., Sharmab,M. and Mittalc,M. (2018) Big data and machine learning based secure healthcare framework. *Procedia Comput. Sci.*, **132**, 1049–1059.

103. Hearst,M.A., Dumais,S.T., Osuna,E. *et al.* (1998) Support vector machines. *IEEE Intell. Syst. Appl.*, **13**, 18–28.

104. Boser,B.E., Guyon,I.M. and Vapnik,V.N. (1992) A training algorithm for optimal margin classifiers. In: *5th Annual ACM Workshop on COLT*. ACM Press, pp. 144–152.

105. Shawe-Taylor,J. and Cristianini,N. (2004) *Kernel Methods for Pattern Analysis*. Cambridge University Press.

106. Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y. & Xu, W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genom. Proteom.* **15**, 41–51 (2018).

107. Cho,G., Yim,J., Choi,Y. *et al.* (2019) Review of machine learning algorithms for diagnosing mental illness. *Psychiatry Investig.*, **16**, 262–269.

108. Cruz,J.A. and Wishart,D.S. (2007) Applications of machine learning in cancer prediction and prognosis. *Cancer Inform.*, **2**, 59–77.

109. Vapnik,V. (1982) *Estimation of Dependences Based on Empirical Data*. Springer Verlag.

110. Chang,H.Y., Jung,C.K., Woo,J.I. *et al.* (2019) Artificial intelligence in pathology. *J. Pathol. Transl. Med.*, **53**, 1–12.

111. LeCun,Y., Bengio,Y. and Hinton,G. (2015) Deep learning. *Nature*, **521**, 436–444.

112. Kevin Zhou,S., Greenspan,H. and Shen,D. (2017) *Deep Learning for Medical Image Analysis*. Academic Press.

113. Hosny,A., Parmar,C., Quackenbush,J. *et al.* (2018) Artificial intelligence in radiology. *Nat. Rev. Cancer*, **18**, 500–510.

114. Langlotz,C.P. *et al.* (2019) A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/the Academy Workshop. *Radiology*, **291**, 781–791.

115. Haenssle,H.A. *et al.* (2018) Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.*, **29**, 1836–1842.

116. Olsen,T.G. *et al.* (2018) Diagnostic performance of deep learning algorithms applied to three common diagnoses in dermatopathology. *J. Pathol. Inform.*, **9**, 32.

117. Rajkomar,A. *et al.* (2018) Scalable and accurate deep learning with electronic health records. *Nat. Digital Med.*, **1**, 18.

118. Ranganathan,P., Pramesh,C.S. and Aggarwal,R. (2017) Common pitfalls in statistical analysis: Logistic regression. *Perspect. Clin. Res.*, **8**, 148–151.

119. Sperandei,S. (2014) Understanding logistic regression analysis. *Biochem. Med.*, **24**, 12–18.

120. McHugh,M.L. (2009) The odds ratio: calculation, usage, and interpretation. *Biochem. Med.*, **19**, 120–126.

121. Kalil,A.C., Mattei,J., Florescu,D.F. *et al.* (2010) Recommendations for the assessment and reporting of multivariable logistic regression in transplantation literature. *Am. J. Transplant.*, **10**, 1686–1694.

122. Tolles,J. and Meurer,W.J. (2016) Logistic regression: relating patient characteristics to outcomes. *JAMA*, **316**, 533–534.

123. Xu,W., Zhao,Y., Nian,S. *et al.* (2018) Differential analysis of disease risk assessment using binary logistic regression with different analysis strategies. *J. Int. Med. Res.*, **46**, 3656–3664.

124. Mamiya,H. *et al.* (2015) Towards probabilistic decision support in public health practice: predicting recent transmission of tuberculosis from patient attributes. *J Biomed. Inform.*, **53**, 237–242.

125. García-Laencina,P.J. *et al.* (2015) Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. *Comput. Biol. Med.*, **59**, 125–133.

126. Nick,T.G. and Logistic Regression,C.K.M. (2007) Topics in biostatistics. *Methods Mol. Biol.*, **404**.

127. Yoo,H.H. *et al.* (2003) Logistic regression analysis of potential prognostic factors for pulmonary thromboembolism. *Chest.*, **123**, 813–821.

128. Zhang,W.T. and Kuang,C.W. (2011) SPSS statistical analysis-based tutorial, 2nd edn. Higher Education Press, Beijing.

129. Hosmer,D.W., Lemeshow,S. and Sturdivant,R.X. (2013) Applied Logistic Regression, 3rd edn. Hoboken Wiley.

130. Solberg,H.E. (1978) Discriminant analysis. *CRC Crit. Rev. Clin. Lab. Sci.*, **9**, 209–242.

131. Davis,J.C. (1986) *Statistics and Data Analysis in Geology*. Wiley, New York.

132. Afifi,A.A. and Clark,V. (1996) Computer-aided multivariate analysis, 3rd edn. Chapman & Hall, London.

133. Brown,C.A. (1998) *Applied Multivariate Statistics in Geohydrology and Related sciences*. Springer, New York.

134. Panagopoulos,G.P. *et al.* (2016) The contribution of cluster and discriminant analysis to the classification of complex aquifer systems. *Environ. Monit. Assess.*, **188**, 591.

135. Shayan,Z., Mohammad Gholi Mezerji,N., Shayan,L. and Naseri,P. (2015) Prediction of depression in cancer patients with different classification criteria, linear discriminant analysis versus logistic regression. *Glob. J. Health Sci.*, **8**, 41–46.

136. Winkler-Schwartz,A. *et al.* (2019) Machine learning identification of surgical and operative factors associated with surgical expertise in virtual reality simulation. *JAMA Netw. Open.*, 20192, e198363.

137. Murata,C. *et al.* (2019) Discriminant analysis to predict the clinical diagnosis of primary immunodeficiencies: a preliminary report. *Rev. Alerg. Mex.*, **62**, 125–133.

138. Dansky,K.H. and Brannon,D. (1996) Discriminant analysis: a technique for adding value to patient satisfaction surveys. *Hosp. Health Services Admin.*, **41**, 503–513.

139. Mandelkow, H., de Zwart, J. A. & Duyn, J. H. Linear discriminant analysis achieves high classification accuracy for the BOLD fMRI response to naturalistic movie stimuli. *Front. Hum. Neurosci.* **10**, 128 (2016).

140. Jin,J. and An,J. (2011) Robust discriminant analysis and its application to identify protein coding regions of rice genes. *Math Biosci.*, **232**, 96–100.

141. Armañanzas,R., Bielza,C., Chaudhuri,K.R. *et al.* (2013) Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach. *Artif. Intell. Med.*, **58**, 195–202.

142. C.-H. Jen, C.-C. Wang, B. C. Jiang, Y.-H. Chu & M.-S. Chen. Application of classification techniques on development an early-warning system for chronic illnesses. *Expert Syst. Appl.* **39**, 10, 8852–8858 (2012).

143. Johnson,K.R., Mascall,G.C. and Howarth,A.T. (1984) Differential laboratory diagnosis of hypercalcemia. *Crit. Rev. Clin. Lab. Sci.*, **21**, 51–97.

144. Lee,E.K., Yuan,F., Hirsh,D.A. *et al.* (2012) A clinical decision tool for predicting patient care characteristics: patients returning within 72 hours in the emergency department. AMIA ... Annual Symposium proceedings. *AMIA Symposium*, **2012**, 495–504.

145. Deo,R.C. (2015) Machine learning in medicine. *Circulation*, **132**, 1920–1930.

146. Quinlan,J.R. (1986) Induction of decision trees. *Machine Learning*, **1**, 81–106.

147. Bae,J.M.T. (2014) Clinical decision analysis using decision tree. *Epidemiol. Health*, **36**, e2014025.

148. Montbriand,M.J. (1995) Decision tree model describing alternate health care choices made by oncology patients. *Cancer Nurs.*, **18**, 104–117.

149. Bonner,G. (2001) Decision making for health care professionals: use of decision trees within the community mental health setting. *J. Adv. Nursing*, **35**, 349–356.

150. Tsai,W.M., Zhang,H., Buta,E. *et al.* (2016) A modified classification tree method for personalized medicine decisions. *Stat. Interface*, **9**, 239–253.

151. Tayefi,M. *et al.* (2017) The application of a decision tree to establish the parameters associated with hypertension. *Comput. Methods Programs Biomed.*, **139**, 83–91.

152. Moon,M. and Lee,S.K. (2017) Applying of decision tree analysis to risk factors associated with pressure ulcers in long-term care facilities. *Healthc. Inform. Res.*, **23**, 43–52.

153. Chern,C.C., Chen,Y.J. and Siao,B. (2019) Decision tree-based classifier in providing telehealth service. *BMC Med. Inform. Decis. Mak.*, **19**, 104.

154. Valdes,G., Luna,J.M., Eaton,E. *et al.* (2016) MediBoost: a patient stratification tool for interpretable decision making in the era of precision medicine. *Sci. Rep.*, **6**, 37854.

155. Gheondea-Eladi,A. (2019) Patient decision aids: a content analysis based on a decision tree structure. *BMC Med. Inform. Dec. Mak.*, **19**, 137.

156. Kasbekar,P.U., Goel,P. and Jadhav,S.P. (2017) A decision tree analysis of diabetic foot amputation risk in Indian patients. *Front. Endocrinol.*, **8**, 25.

157. Ainscough,K.M., Lindsay,K.L., O'Sullivan,E.J. *et al.* (2017) Behaviour change in overweight and obese pregnancy: a decision tree to support the development of antenatal lifestyle interventions. *Public Health Nutri.*, **20**, 2642–2648.

158. Ho,T.K. (1998) The random subspace method for constructing decision forests. *IEEE T. Pattern Anal.*, **20**, 832–844.

159. Kotti,M., Duffell,L.D., Faisal,A.A. and McGregor,A.H. (2017) Detecting knee osteoarthritis and its discriminating parameters using random forests. *Med. Eng. Phys.*, **43**, 19–29.

160. Pachauri,I. and Sharma,S. (2015) Anomaly detection in medical wireless sensor networks using machine learning algorithms. *Procedia Comput. Sci.*, **70**, 325–333.

161. O'Bryant,S.E. *et al.* (2010) A serum protein–based algorithm for the detection of Alzheimer disease. *Arch. Neurol.*, **67**, 1077–1081.

162. Toubiana,D. *et al.* (2019) Combined network analysis and machine learning allows the prediction of metabolic pathways from tomato metabolomics data. *Commun. Biol.*, **2**, 214.

163. Roysden,N. and Wright,A. (2015) Predicting health care utilization after behavioral health referral using natural language processing and machine learning. *AMIA*, **2015**, 2063–2072.

164. Morid,M.A., Kawamoto,K., Ault,T. *et al.* (2018) Supervised learning methods for predicting healthcare costs: systematic literature review and empirical evaluation. AMIA ... Annual symposium proceedings. *AMIA Symposium*, **2017**, 1312–1321.

165. Lee,J. (2017) Patient-specific predictive modeling using random forests: An observational study for the critically ill. *JMIR Med. Inform.*, **5**, e3.

166. Sarica,A., Cerasa,A. and Quattrone,A. (2017) Random Forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. *Front. Aging Neurosci.*, **9**, 329.

167. Seligman,B., Tuljapurkar,S. and Rehkopf,D. (2017) Machine learning approaches to the social determinants of health in the health and retirement study. *SSM Popul. Health*, **4**, 95–99.

168. Khalilia,M., Chakraborty,S. and Popescu,M. (2011) Predicting disease risks from highly imbalanced data using random forest. *BMC Med. Inform. Dec. Mak.*, **11**, 51.

169. DuBrava,S. *et al.* (2017) Using random Forest models to identify correlates of a diabetic peripheral neuropathy diagnosis from electronic health record data. *Pain Med.*, **18**, 107–115.

170. Rahimian,F. *et al.* (2018) Predicting the risk of emergency admission with machine learning: development and validation using linked electronic health records. *PLoS Med.*, **15**, e1002695.

171. McWilliams,C.J. *et al.* (2019) Towards a decision support tool for intensive care discharge: machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK. *BMJ Open*, **9**, e025925.

172. Wager,S. and Athey,S. (2018) Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.*, **113**, 1228–1242.

173. Yulita,I.N., Fanany,M.I. and Arymurthy,A.M. (2018) Fast convolutional method for automatic sleep stage classification. *Healthc. Inform. Res.*, **24**, 170–178.

174. Byeon,H. (2019) Developing a random forest classifier for predicting the depression and managing the health of caregivers supporting patients with Alzheimer's disease. *Technol. Health Care*, 1–14.

175. Schneider,A., Hommel,G. and Blettner,M. (2010) Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Deutsches Arzteblatt international*, **107**, 776–782.

176. Bebbington,E. and Furniss,D. (2015) Linear regression analysis of hospital episode statistics predicts a large increase in demand for elective hand surgery in England. *JPRAS.*, **68**, 243–251.

177. Backenroth,D., Chase,H.S., Wei,Y. and Friedman,C. (2017) Monitoring prescribing patterns using regression and electronic health records. *BMC Medical Inform. Dec. Mak.*, **17**, 175.

178. Morton,V. and Torgerson,D.J. (2003) Effect of regression to the mean on decision making in health care. *BMJ Clin. Res. Ed.*, **326**, 1083–1084.

179. Madadizadeh,F., Ezati Asar,M. and Bahrampour,A. (2016) Quantile regression and its key role in promoting medical research. *Iran. J. Public Health*, **45**, 116–117.

180. Malehi,A.S., Pourmotahari,F. and Angali,K.A. (2015) Statistical models for the analysis of skewed healthcare cost data: a simulation study. *Health Econ. Rev.*, **5**, 11.

181. Madigan,E.A., Curet,O.L. and Zrinyi,M. (2008) Workforce analysis using data mining and linear regression to understand HIV/AIDS prevalence patterns. *Hum. Res. Health*, **6**, 2.

182. Langley,P., Iba,W. and Thompson,K. (1992) An analysis of Bayesian classifiers. Proceedings of the Tenth National Conference on. *Artif. Intell.*, 223–228.

183. Rish,I. (2001) An empirical study of the naive Bayes classifier. *IJCAI 2001 Work Empir Methods Artif Intell.*, **3**, 41–46.

184. Langarizadeh,M. and Moghbeli,F. (2016) Applying naive Bayesian networks to disease prediction: a systematic review. *AIM: journal of the Society for Medical Informatics of Bosnia & Herzegovina*, **24**, 364–369.

185. Wolfson,J. *et al.* (2015) A naive Bayes machine learning approach to risk prediction using censored, time-to-event data. *Stat. Med.*, **34**, 2941–2957.

186. Ehsani-Moghaddam,B., Queenan,J.A., MacKenzie,J. and Birtwhistle,R.V. (2018) Mucopolysaccharidosis type II detection by Naïve Bayes classifier: An example of patient classification for a rare disease using electronic medical records from the Canadian Primary Care Sentinel Surveillance Network. *PloS One*, **13**, e0209018.

187. Wei,W., Visweswaran,S. and Cooper,G.F. (2011) The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data. *JAMIA*, **18**, 370–375.

188. Doing-Harris,K., Mowery,D.L., Daniels,C. *et al.* (2017) Understanding patient satisfaction with received healthcare services: a natural language processing approach. *AMIA Symposium*, **2016**, 524–533.

189. Grover,D., Bauhoff,S. and Friedman,J. (2019) Using supervised learning to select audit targets in performance-based financing in health: An example from Zambia. *PloS One*, **14**, e0211262.

190. Wagholikar,K.B., Vijayraghavan,S. and Deshpande,A.W. (2009) Fuzzy naive Bayesian model for medical diagnostic decision support. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 3409–3412 2009.

191. Al-Aidaroos,K.M., Bakar,A.A. and Othman,Z. (2012) Medical data classification with naive Bayes approach. *Inform. Technol. J.*, **11**, 1166–1174.

192. Sebastiani,P., Solovieff,N. and Sun,J.X. (2012) Naïve Bayesian classifier and genetic risk score for genetic risk prediction of a categorical trait: not so different after all! *Front. Genet.*, **3**, 26.

193. Srinivas,K. *et al.* (2010) Applications of data mining techniques in healthcare and prediction of heart attacks. *Int. J. Comput. Sci. Eng.*, **2**, 250–255.

194. Altman,N.S. (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.*, **46**, 175–185.

195. Zhang,Z. (2016) Introduction to machine learning: k-nearest neighbors. *Ann. Translat. Med.*, **4**, 218.

196. Park,J. and Lee,D.H. (2018) Privacy preserving k-nearest neighbor for medical diagnosis in e-health cloud. *J. Healthc. Eng.*, **2018**, 4073103.

197. Hu,L.Y., Huang,M.W., Ke,S.W. and Tsai,C.F. (2016) The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus*, **5**, 1304.

198. Li,C., Zhang,S., Zhang,H. *et al.* (2012) Using the K-nearest neighbor algorithm for the classification of lymph node metastasis in gastric cancer. *Comput. Math. Methods Med.*, **876545**, **2012**.

199. Sarkar,M. and Leong,T.Y. (2000) Application of K-nearest neighbors algorithm on breast cancer diagnosis problem. *AMIA Symposium*, **2000**, 759–763.

200. Vitola,J., Pozo,F., Tibaduiza,D.A. and Anaya,M.A. (2017) Sensor data fusion system based on k-nearest neighbor pattern classification for structural health monitoring applications. *Sensors Basel. Switzerland*, **17**, 417.

201. Zhao,D. and Weng,C. (2011) Combining PubMed knowledge and EHR data to develop a weighted bayesian network for pancreatic cancer prediction. *J. Biomed. Inform.*, **44**, 859–868.

202. Baum,L.E. and Petrie,T. (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.*, **37**, 1554–1563.

203. Baum,L.E. and Egon,J.A. (1967) An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology. *Bull. Amer. Meteorol. Soc.*, **73**, 360–363.

204. Rabiner,L.R. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE.*, **77**, 257–286.

205. Eddy,S.R. (2004) What is a hidden Markov model? *Nat. Biotech.*, **22**, 1315–1316.

206. Wall,M.M. and Li,R. (2009) Multiple indicator hidden Markov model with an application to medical utilization data. *Stat. Med.*, **28**, 293–310.

207. Kawamoto,R. *et al.* (2013) Hidden markov model for analyzing time-series health checkup data. *Stud. Health Technol. Inform.*, **192**, 491–495.

208. Sotoodeh,M. and Ho,J.C. (2019) Improving length of stay prediction using a hidden Markov model. *AMIA Jt. Summits Transl. Sci.*, **2019**, 425–434.

209. Sampathkumar,H., Chen,X.W. and Luo,B. (2014) Mining adverse drug reactions from online healthcare forums using hidden Markov model. *BMC Med. Inform. Dec. Mak.*, **14**, 91.

210. Huang,Z., Dong,W., Wang,F. and Duan,H. (2015) Medical inpatient journey modeling and clustering: a Bayesian hidden Markov model based approach. *AMIA Symposium*, **2015**, 649–658.

211. Esmaili,N., Piccardi,M., Kruger,B. and Girosi,F. (2019) Correction: analysis of healthcare service utilization after transport-

related injuries by a mixture of hidden Markov models. *PloS One*, **14**, e0214973.

212. Huang,Q., Cohen,D., Komarzynski,S. *et al.* (2018) Hidden Markov models for monitoring circadian rhythmicity in telemetric activity data. *J. R. Soc. Interface*, **15**, 20170885.

213. Marchuk,Y. *et al.* (2018) Predicting patient-ventilator asynchronies with hidden Markov models. *Sci. Rep.*, **8**, 17614.

214. Naithani,G. *et al.* (2018) Automatic segmentation of infant cry signals using hidden Markov models. *EURASIP J. Audio Speech Music Process*, **1**, 1.

215. Reeves,C. (2003) Genetic algorithms. *Handbook of metaheuristics: Springer*, **3**, 55–82.

216. Hasancebi,O. and Erbatur,F. (2000) Evaluation of crossover techniques in genetic algorithm based optimum structural design. *Comput. Struc.*, **78**, 435–448.

217. Ghaheri,A., Shoar,S., Naderan,M. and Hoseini,S.S. (2015) The applications of genetic algorithms in medicine. *Oman Med. J.*, **30**, 406–416.

218. Karnan,M. and Thangavel,K. (2007) Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications. *Comput. Methods Programs Biomed.*, **87**, 12–20.

219. Pereira,D.C., Ramos,R.P. and do Nascimento,M.Z. (2014) Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm. *Comput. Methods Programs Biomed.*, **114**, 88–101.

220. Baum,K.G., Schmidt,E., Rafferty,K. *et al.* (2011) Evaluation of novel genetic algorithm generated schemes for positron emission tomography (PET)/magnetic resonance imaging (MRI) image fusion. *J. Digit Imaging*, **24**, 1031–1043.

221. Wold,S., Sjostrom,M. and Eriksson,L. (2001) PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.*, **58**, 109–130.

222. Tan,C., Chen,H., Wu,T. and Xia,C. (2011) Modeling the relationship between cervical cancer mortality and trace elements based on genetic algorithm-partial least squares and support vector machines. *Biol. Trace Elem. Res.*, **140**, 24–34.

223. Khalil,A.S., Bouma,B.E. and Kaazempur Mofrad,M.R. (2006) A combined FEM/genetic algorithm for vascular soft tissue elasticity estimation. *Cardiovasc. Eng.*, **6**, 93–102.

224. Zhou,X. *et al.* (2009) Identification of biomarkers for risk stratification of cardiovascular events using genetic algorithm with recursive local floating search. *Proteomics*, **9**, 2286–2294.

225. Tu,C., Zeng,Y. and Yang,X. (2005) A new approach to detect QRS complexes based on a histogram and genetic algorithm. *J. Med. Eng. Technol.*, **29**, 176–180.

226. Nguyen,L.B., Nguyen,A.V., Ling,S.H. and Nguyen,H.T. (2013) Combining genetic algorithm and Levenberg-Marquardt algorithm in training neural network for hypoglycemia detection using EEG signals. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 5386–5389 2013.

227. Hoh,J.K., Cha,K.J., Park,M.I. *et al.* (2012) Estimating time to full uterine cervical dilation using genetic algorithm. *Kaohsiung J. Med. Sci.*, **28**, 423–428.

228. Ocak,H. (2013) A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. *J. Med. Syst.*, **37**, 9913.

229. Latkowski,T. and Osowski,S. (2015) Computerized system for recognition of autism on the basis of gene expression microarray data. *Comput. Biol. Med.*, **56**, 82–88.

230. Jefferson,M.F., Pendleton,N., Lucas,S.B. and Horan,M.A. (1997) Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with nonsmall cell lung carcinoma. *Cancer*, **79**, 1338–1342.

231. Güler,I., Polat,H. and Ergün,U. (2005) Combining neural network and genetic algorithm for prediction of lung sounds. *J. Med. Syst.*, **29**, 217–231.

232. Castiglione,F., Poccia,F., D'Offizi,G. and Bernaschi,M. (2013) Mutation, fitness, viral diversity, and predictive markers of disease progression in a computational model of HIV type 1 infection. *AIDS Res. Hum. Retroviruses*, **20**, 1314–1323.

233. Yu,Y., Schell,M.C. and Zhang,J.B. (1997) Decision theoretic steering and genetic algorithm optimization: application to stereotactic radiosurgery treatment planning. *Med. Phys.*, **24**, 1742–1750.

234. Pei,Y., Kim,Y., Obinata,G. *et al.* (2011) Trajectory planning of a robot for lower limb rehabilitation. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 1259–1263 2011.

235. Jaremko,J.L. *et al.* (2002) Genetic algorithm-neural network estimation of cobb angle from torso asymmetry in scoliosis. *J. Biomech. Eng.*, **124**, 496–503.

236. Smigrodzki,R. *et al.* (2005) Genetic algorithm for analysis of mutations in Parkinson's disease. *Artif. Intell. Med.*, **35**, 227–241.

237. Chen,H.Y., Chen,T.C., Min,D.I. *et al.* (1999) Prediction of tacrolimus blood levels by using the neural network with genetic algorithm in liver transplantation patients. *Ther. Drug Monit.*, **21**, 50–56.

238. Prabhu,S.P. (2019) Ethical challenges of machine learning and deep learning algorithms. *Lancet Oncol.*, **20**, 621–622.

239. Park,S.H. *et al.* (2019) Ethical challenges regarding artificial intelligence in medicine from the perspective of scientific editing and peer review. *Sci. Ed.*, **6**, 91–98.

240. Park,S.H., Do,K.H., Kim,S. *et al.* (2019) What should medical students know about artificial intelligence in medicine? *J. Educ. Eval. Health Prof.*, **16**, 18.

241. Somashekhar,S.P. *et al.* (2018) Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. *Ann. Oncol.*, **29**, 418–423.

242. Liu,C., Liu,X., Wu,F. *et al.* (2018) Using artificial intelligence (Watson for oncology) for treatment recommendations amongst Chinese patients with lung cancer: feasibility study. *J. Med. Internet Res.*, **20**, e11087.

243. Theodore,K. *et al.* (2019) Harnessing CURATE.AI as a digital therapeutics platform by identifying N-of-1 learning trajectory profiles. In: *Advanced Therapeutics*.

244. Londhe,V.Y. and Bhasin,B. (2019) Artificial intelligence and its potential in oncology. *Drug Discov. Today.*, **24**, 228–232.

245. Henry,K., Hager,D., Pronovost,P. and Saria,S. (2015) A targeted real-time early warning score (TREWScore) for septic shock. *Sci. Transl. Med.*, **7**, 299ra122.

246. Amorim,F.F. and Santana,A.N. (2017) Automated early warning system for septic shock: the new way to achieve intensive care unit quality improvement? *Ann. Transl. Med.*, **5**, 17.

247. Khurana,H.S. *et al.* (2016) Real-time automated sampling of electronic medical records predicts hospital mortality. *Am. J. Med.*, **129**, 688–698.e2.

248. Saria,S., Butte,A. and Sheikh,A. (2018) Better medicine through machine learning: What's real, and what's artificial? *PLoS Med.*, **15**, e1002721.

249. Kerlikowske,K. *et al.* (2018) Automated and clinical breast imaging reporting and data system density measures predict risk for screen-detected and interval cancers: a case-control study. *Ann. Int. Med.*, **168**, 757–765.

250. Drug and Device News (2018) *P & T : a peer-reviewed journal for formulary management*, **43**, 382–429.

251. Badgeley,M.A. *et al.* (2019) Deep learning predicts hip fracture using confounding patient and healthcare variables. *NPJ Digit. Med.*, **2**, 31.

252. McKinney,S.M. *et al.* (2020) International evaluation of an AI system for breast cancer screening. *Nature*, **577**, 89–94.

253. Papadakis,G.Z. *et al.* (2019) Deep learning opens new horizons in personalized medicine. *Biomed. Rep.*, **10**, 215–217.

254. Parekh, V. S., & Jacobs, M. A. Deep learning and radiomics in precision medicine. *Expert Rev. Precis. Med. Drug Dev.*, **4**(2), 59–72 (2019).

255. Khoury,M.J. *et al.* (2016) Precision public health for the era of precision medicine. *Am. J. Prev. Med.*, **50**, 398–401.