

REVIEW

Open Access



# Introducing AI to the molecular tumor board: one direction toward the establishment of precision medicine using large-scale cancer clinical and biological information

Ryuji Hamamoto<sup>1,2\*</sup>, Takafumi Koyama<sup>3</sup>, Nobuji Kouno<sup>1,4</sup>, Tomohiro Yasuda<sup>1,5</sup>, Shuntaro Yui<sup>1,5</sup>, Kazuki Sudo<sup>3,6</sup>, Makoto Hirata<sup>7</sup>, Kuniko Sunami<sup>8</sup>, Takashi Kubo<sup>8</sup>, Ken Takasawa<sup>1,2</sup>, Satoshi Takahashi<sup>1,2</sup>, Hidenori Machino<sup>1,2</sup>, Kazuma Kobayashi<sup>1,2</sup>, Ken Asada<sup>1,2</sup>, Masaaki Komatsu<sup>1,2</sup>, Syuzo Kaneko<sup>1,2</sup>, Yasushi Yatabe<sup>9,10</sup> and Noboru Yamamoto<sup>3</sup>

## Abstract

Since U.S. President Barack Obama announced the Precision Medicine Initiative in his New Year's State of the Union address in 2015, the establishment of a precision medicine system has been emphasized worldwide, particularly in the field of oncology. With the advent of next-generation sequencers specifically, genome analysis technology has made remarkable progress, and there are active efforts to apply genome information to diagnosis and treatment. Generally, in the process of feeding back the results of next-generation sequencing analysis to patients, a molecular tumor board (MTB), consisting of experts in clinical oncology, genetic medicine, etc., is established to discuss the results. On the other hand, an MTB currently involves a large amount of work, with humans searching through vast databases and literature, selecting the best drug candidates, and manually confirming the status of available clinical trials. In addition, as personalized medicine advances, the burden on MTB members is expected to increase in the future. Under these circumstances, introducing cutting-edge artificial intelligence (AI) technology and information and communication technology to MTBs while reducing the burden on MTB members and building a platform that enables more accurate and personalized medical care would be of great benefit to patients. In this review, we introduced the latest status of elemental technologies that have potential for AI utilization in MTB, and discussed issues that may arise in the future as we progress with AI implementation.

**Keywords:** Molecular tumor board, Precision medicine, Artificial intelligence, Next-generation sequencing, Natural language processing

## Background

The human genome project, which began in 1990 to analyze the entire human sequence, was declared complete in April 2003 after 13 years and a budget of approximately

US\$3 billion [1–3]. The world then entered the post-genomic era, and expectations grew for the development of “personalized medicine,” in which genomic information is applied to medical treatment [4–6]. When the 454 Genome Sequencer 20 (GS20), the first next-generation sequencing (NGS) technology, was introduced in 2005, genetic analysis using NGS became actively pursued. Further, research in fields such as genetic medicine and pharmacogenomics became more active toward

\*Correspondence: rhamamot@ncc.go.jp

<sup>1</sup> Division of Medical AI Research and Development, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan  
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the realization of personalized medicine, which aims to provide medical care based on an individual's genetic information [7–9]. In 2014, Illumina made genetic analysis increasingly accessible with the announcement of a \$1,000 genome offer upon the sale of the HiSeq X Ten [10, 11]. Under these circumstances, the precision medicine initiative was announced by U.S. President Barack Obama in his New Year's State of the Union address on January 20, 2015. The presentation declared that the study of how genomic information, environmental factors, and lifestyle affect health maintenance and disease development using large clinical samples will divide patients/potential patients into subgroups with respect to disease susceptibility and develop appropriate treatment and disease prevention methods for each group [12–14]. President Obama's announcement impacted global healthcare policy, and the establishment of precision medicine systems was prioritized in countries globally. In particular, the U.S. FDA's approval of MSK-IMPACT™ and FoundationOne® CDx, tumor profiling tests for solid tumors based on NGS-based genetic mutation analysis, in late 2017 increased momentum for optimal treatment based on genetic information in actual clinical practice [15–17]. In Japan, the OncoGuide™ NCC Oncopanel system and FoundationOne® CDx were approved by the Ministry of Health, Labour and Welfare in 2018 and covered by insurance in 2019, making cancer genomic medicine available under insurance reimbursement [18–20]. On the other hand, the molecular tumor board (MTB), which is composed of experts in various fields such as clinical oncology and genetic medicine, discusses the results of genetic mutation analysis; however, it is a complex, time-consuming, and labor-intensive process [21–23]. With the growing expectations for precision medicine and the need to make MTBs more efficient and effective, several MTBs have been reported that utilize a virtual environment [24–26]. The expectation and burden on MTB members are predicted to increase in the future, and the introduction of artificial intelligence (AI) and the latest information and communication technology (ICT) to improve the efficiency and automation of MTB is important when establishing a precision medicine system.

Along with recent advances in machine learning technology, particularly deep learning, AI technology has been attracting attention, and social implementation of AI is progressing in a variety of fields [27–29]. The medical field is no exception, with a succession of AI-based medical device programs being approved in countries globally, and their use in clinical settings is progressing [30–35]. Because deep learning technology is particularly strong in image analysis, AI research and development using medical images, such as radiological, endoscopic,

ultrasound, and skin, has been actively conducted, and many important findings have been obtained [36–45]. In addition to medical image analysis, AI is used for omics data analysis and single cell analysis, and natural language processing (NLP), an AI technology, is now being used to analyze electronic medical records and medical papers, with research aimed at clinical applications [46–49].

This review focuses on the introduction of AI into the MTB and discusses the history of AI and the introduction of computers into diagnosis, the current status of AI utilization in the promotion of precision medicine, and future directions. Importantly, machine learning techniques, including NLP, are categorized as AI techniques because, as noted above, current AI techniques are based on machine learning methods.

### **Machine learning, the technological foundation of current AI research and development, and its application to medical research**

Current AI has become vastly popular due to its basis on machine learning with deep learning as its technological foundation [30, 50]. Based on its analytical characteristics, machine learning can be broadly classified into four categories: supervised, unsupervised, semi-supervised, and reinforcement learning (Fig. 1) [51, 52]. Supervised learning is then broadly classified into regression and classification problems, and regression models are used as models for disease onset prediction and prognostic prediction [53–55]. Classification by supervised learning is currently the most widely studied method, especially in medical image analysis, and is practically applied in clinical practice [33, 56–63]. This is because deep learning technology is particularly strong in image analysis, and when medical AI is practically applied to clinical practice, it is important to use data obtained from physicians' diagnoses (especially specialists) as training data. Unsupervised learning is subsequently used for clustering and dimensionality compression, and its applications in medical research include patient stratification, medical image categorization, and disease subdivision [64–68]. Semi-supervised learning combines supervised and unsupervised learning. This technology has attracted attention in the medical field because it enables highly accurate learning even with limited amount of training data available. After supervised learning with limited labeled data using this method, it becomes possible to perform analysis based on unsupervised learning with a large amount of unlabeled data and generate medical images using the generated model [69–73]. Lastly, reinforcement learning is a technique that continuously learns from each experience to optimize subsequent behavior and maximize the final outcome. The algorithm of AlphaGo, the AI that defeated the world's top Go players, uses a learning

method called Q-learning, which is a type of reinforcement learning with the optimal action value defined as Q-value. The algorithm selects the action that maximizes the Q-value from a large number of trials and their results [74, 75]. In medical research, it is used for disease detection, optimization of treatment strategies, and prediction of efficacy and side effects of anticancer drugs [76–80].

### Medical applications of computer-aided diagnosis support and NLP

Computer-aided detection/diagnosis (CAD) is important in describing AI-based diagnostic assistance. CADE (computer-aided detection) is a device that incorporates a software that allows a computer to automatically detect and mark the location of candidate lesions on an image, and the computer processes medical images and inspection data if possible to assist in the detection of lesions or abnormal values [81]. CADx (computer-aided diagnosis) is a stand-alone software or device with a software that detects suspected lesion sites, outputs quantitative data as numerical values and graphs, such as discrimination of lesion candidates as good or bad and the degree of disease progression [82], including those that provide diagnostic support by providing candidate diagnostic results, information on risk assessment, etc.

In the late 1950s, during the dawn of modern computers, studies examined the development of CAD [83–85]. In the 1970s, there was substantial worldwide interest in expert systems, and the MYCIN expert system was developed as an early CAD system. MYCIN is generally recognized as one of the most important early applications of AI in the medical field [86]. It uses a simple inference engine and has a knowledge base of ~500 rules. Several simple “yes/no” or written response questions are asked by the physician, and a final list of possible causative bacterial names (in order of probability), the level of confidence in each, its reason, and the recommended course of drug therapy are determined. Since it showed a relatively good diagnostic accuracy of ~65% in clinical cases [87], further refinement of the algorithm was expected to lead to more accurate diagnoses in various areas, leading to many research and development efforts aimed at applying the expert system to medicine. However, MYCIN was never applied clinically because computer performance was poor at that time, and the social infrastructure, such as laws and bioethics, had not been developed. The development of an expert system requires a substantial amount of time and effort; the expert system cannot convert ambiguous human expressions into rules, resulting in inconsistencies, and other fundamental shortcomings [30, 88].

Subsequently, in the late 1980s and early 1990s, the focus shifted to the use of data mining approaches for more sophisticated and flexible CAD systems. A significant milestone was achieved in 1998 when the first commercial CAD system for mammography, the Image-Checker M1000 system, was approved by the U.S. Food and Drug Administration (FDA) [89, 90]. Following this, CAD for detecting lung cancer (nodule shadows) in plain chest radiographs and chest CT images, and for polyp detection in colon CT examinations received FDA approval in rapid succession [91–93].

In the 21st century, with the development of deep learning technology using autoencoder by Hinton et al., image analysis using AI technology began to attract attention [94]. This trend has also been observed in the medical field, and the focus is on the development of CAD systems that take advantage of AI. Particularly in 2015, in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), an object recognition rate competition, a group from Microsoft demonstrated recognition accuracy with lesser than the average human error rate of 4.9% by using deep learning technology. Furthermore, in 2017, AlphaGo was named the world’s top performer in the game of Go, and there were a number of reports of AI surpassing human capabilities [95–97]. These facts spurred the development of the AI-based software as a medical device (SaMD). More than 300 AI-SaMDs have been approved by the U.S. FDA so far, and clinical applications are being explored, focused on medical image (radiological, endoscopic, ultrasound, etc.) analysis [98, 99]. Additionally, AI is also being introduced for analysis of omics, and medical information and research papers [100–105].

NLP, a branch of artificial intelligence, is a series of techniques that allows computers to process everyday human language [106]. With the recent developments in deep learning technology, it is becoming possible for machines to understand and translate natural language [107]. NLP is also being actively studied in the medical field, as a significant portion of its diverse data generated contains such sentences [108–112]. Until recently, secondary use of medical data had been primarily based on relatively structured data, such as health checkup and medical fee data, but recently it is being developed for handling larger-scale, unstructured data. We categorize the utilization of natural language data in the medical field into three major trends. The first is toward the utilization of data from physicians’ daily practice, represented by medical records (electronic) [113–115]. For example, automatic extraction of adverse drug reaction signals from electronic medical records is being attempted [116–118]. The second major trend is the use of NLP to analyze published data, such as medical articles and case reports,

**Table 1** Example of MTB workflow

Work item	Work details
I Assign biological significance to the genetic abnormality (eg., whether it contributes to the acquisition of a particular trait, such as oncogenic potential).	Focusing on variants whose pathological significance is judged differently in laboratory company reports and C-CAT survey results, the registration statuses of the Gene Polymorphism Database (gnomAD), Somatic Mutation Database (COSMIC), and ClinVar will be checked to determine their pathological significance.
II Interpretation of genetic evidence for diagnosis and prognosis.	Search public databases (CIVIC, OncoKB, etc.) and literature for information on diagnosis and prognosis.
III Attach a specific candidate drug and level of evidence corresponding to the genetic abnormality after considering basic patient information (cancer type)	Focus on drugs listed in laboratory company reports and C-CAT survey results, and search public databases (CIVIC, OncoKB, etc.) and literature to confirm the level of evidence based on the latest findings.
IV When germline gene abnormalities are recognized (or suspected), the significance and response should be discussed in accordance with relevant guidelines, guidance, and recommendations.	Consider significance and response based on guidelines, guidance, and recommendations related to secondary findings.
V After considering basic patient information including previous history of chemotherapy, specific candidate drugs and evidence levels corresponding to the genetic abnormality, as well as their approval status and clinical trial status, should be investigated.	Review the list of clinical trials being conducted at the hospital and consider whether to enroll the patient in the relevant clinical trial.
VI If necessary, consider whether any of the candidate drugs listed in IV can be recommended, considering the patient's condition, availability, etc.	Check the pathological significance of each gene, level of evidence for the drug linked to the genetic mutation, and availability of the drug.

to extract important information for clinical applications [119–121]. In particular, a vast number of medical papers and case reports are published daily, and it is physically impossible for a clinician to comprehensively examine them [110]. Therefore, we believe that the extraction of important information using NLP is rather essential. The third major trend, which has been gaining attention over the past few years or so, is toward the private data that patients exchange through social media and patient associations [122–126]. The NLP system extracts episodes related to patients' treatment, problems, and practical knowledge obtained from texts via social media. By creating and providing content with appropriate medical information to the extracted episodes, and building a mechanism for sharing similar episodes and practical knowledge among patients, an attempt is being made to create a foundation for patients and medical professionals to learn and utilize this information to enhance patient care.

Therefore, it is important to use NLP appropriately when considering MTB efficiency.

### General MTB tasks and workflow

MTB, a meeting held to medically interpret the clinical implications of the results of genetic analysis obtained using NGS with the aim of proposing appropriate treatments for each individual patient, is critically important in promoting precision medicine [127–132]. Table 1 describes the general workflow of the MTB, showing an example of the MTB conducted at the National Cancer Center (NCC) Japan. Although the databases used and other details vary by country and institution, the basic work performed is common. In addition, Table 2 introduces databases that are important for MTB.

The first important step in obtaining data from patient-derived samples analyzed using NGS is to assign biological significance to the genetic abnormality (e.g., whether it contributes to the acquisition of a particular trait, such as oncogenic potential). In the NCC Japan, focusing on variants whose pathological significance is judged differently in laboratory company reports and survey results of Center for Cancer Genomics and Advanced Therapeutics (C-CAT) [133], the registration status of gene polymorphism database (gnomAD), somatic mutation database (COSMIC), and ClinVar, a public database of variant interpretations, will be checked to determine the pathological significance of the final judgment.

This is followed by interpretation of genetic evidence for diagnosis and prognosis. Here, public databases (CIViC, OncoKB, etc.) and literature are searched to determine if there are any findings regarding diagnosis and prognosis.

The next step is to attach specific candidate drugs and evidence corresponding to the genetic abnormality, considering basic patient information (cancer type). Here, the level of evidence based on the latest findings was confirmed by searching public databases (CIViC, OncoKB, etc.) and literature, focusing on drugs listed in laboratory company reports and C-CAT survey results. When germline gene abnormalities are present (or suspected), the significance and response should be based on guidelines, guidance, and recommendations related to secondary findings. The list of clinical trials being conducted at the hospital will be reviewed and the possibility of enrollment in the relevant clinical trials will be considered.

If necessary, the specific candidate drugs listed will be reviewed to see if any of them are recommended for the patient's condition and availability. Here, the pathological significance of each gene, level of evidence for the drug linked to the genetic variant, and availability of the drug are identified.

The OncoKB database is used as an example for further details (as of September 2022) [134] because interpretation of evidence for genetic abnormalities is important [135]. First, FDA-approved biomarkers that predict response to FDA-approved drugs are level 1, with 44 registered genes. Then the FDA-approved standard of care biomarker that predicts response to the drug is at level 2, with 23 registered genes. There is strong clinical evidence showing that the biomarkers are predictive of response to drugs, but neither the biomarkers nor the drugs are standard of care at Level 3, with 33 registered genes. In addition, there is strong biological evidence showing that a biomarker predicts response to drugs, but neither the biomarker nor the drug is a standard treatment, which is level 4, with 25 genes registered. Furthermore, eight genes are registered at level R1, the FDA-approved standard of care biomarker for predicting resistance to drugs.

These are the series of MTB flows; however, with the increasing expectations for precision medicine, increasing burden on MTB members, and aim to conduct more efficient and comprehensive surveys, attempts are being made to introduce state-of-the-art AI and ICT technologies into MTB. The next section introduces research results using AI technology that could be applied to each task in MTB.

### AI-based prediction of biological significance for genetic abnormalities and its application to diagnosis and proposal of candidate drugs for treatment

Despite the existence of an excellent database on oncogenes, it is difficult to determine the significance of most of the mutations identified in oncogenes for tumorigenesis, regardless of tumor type. To address this challenge,

**Table 2** Databases commonly used in MTB

Database name	Operating organization [URL]*	Database contents	References
ClinVar	NCBI [ <a href="https://www.ncbi.nlm.nih.gov/clinvar">https://www.ncbi.nlm.nih.gov/clinvar</a> ]	This database collects information on the diversity of the human genome and related diseases and provides it as a freely available archive. It contains polymorphism locations, gene names, and their relationship to diseases. Gene information is linked to dbSNP and dbVar at NCBI, and phenotypes are linked to MedGen. Information is also collected from the NIH Genetic Testing Registry (GTR), OMIM, and PubMed. Each data can be viewed in HTML format or downloaded in XML or tab-delimited format, and some data in VCF files.	[209]
COSMIC	Wellcome Trust Sanger Institute [ <a href="https://cancer.sanger.ac.uk/cosmic/">https://cancer.sanger.ac.uk/cosmic/</a> ]	COSMIC is the most detailed and comprehensive resource for studying the effects of somatic mutations in human cancer. In addition to coding mutations, COSMIC covers all genetic mechanisms by which somatic mutations promote cancer, including non-coding mutations, gene fusions, copy number mutations, and drug resistance mutations. The core COSMIC database is complemented by additional datasets that allow users to contextualize the biomarkers they detect. The Cancer Gene Census provides a detailed catalog of > 700 genes involved in cancer, their biological functions, and descriptions of the genetic mechanisms that cause cancer. The Cancer Mutation Census provides information on the significance of all coding mutations based on biological and biochemical information from multiple sources. Mutation Actionability in Precision Oncology (Actionability) provides updates on drugs targeting specific somatic mutations at all stages of development. The Cell Lines Project advocates cell line omics data through systematic characterization of the genetics and genomics (variation and gene expression) of over 1,000 cancer cell lines.	[210]
gnomAD	Broad Institute [ <a href="https://gnomad.broadinstitute.org">https://gnomad.broadinstitute.org</a> ]	gnomAD is a resource developed by an international coalition of researchers with the goal of aggregating and harmonizing exome and genome data from various large-scale sequencing projects and making them available to a wider audience of scientists. The v2.1.1 dataset (GRCh37/hg19) provided by this site covers 125,748 exome sequences and 15,708 whole genome sequences of unrelated individuals sequenced as part of various disease-specific and population genetic studies (as of March 2022). v3.1.2 dataset (GRCh38) has 76,156 genomes selected (as of March 2022), similar to v2.	[211]
OncoKB	Memorial Sloan Kettering Cancer Center [ <a href="https://www.oncokb.org">https://www.oncokb.org</a> ]	OncoKB is a groundbreaking precision oncology knowledge base that leverages the clinical expertise of Memorial Sloan Kettering (MSK) to provide accurate and up-to-date information on the biological and clinical significance of > 5,000 cancer gene mutations. On October 7, 2021, the U.S. Food and Drug Administration approved OncoKB for partial listing as the first tumor mutation database to be included in the Public Human Gene Mutation Database. This is the first tumor mutation database to be recognized by the FDA. Treatment information is categorized by the OncoKB Levels of Evidence system, which assigns clinical utility (from standard to investigational treatment) to individual mutational events.	[212]
CIViC	The McDonnell Genome Institute at Washington University School of Medicine [ <a href="https://civicdb.org/home">https://civicdb.org/home</a> ]	CIViC is an expert knowledge base for the clinical interpretation of variants in cancer, describing the therapeutic, prognostic, diagnostic, and predispositional relevance of all types of genetic and somatic mutations. CIViC is committed to open source code, open access content, public application programming interfaces (APIs), and proof of supporting evidence, enabling the transparent creation of current and accurate variant interpretations for use in precision medicine for cancer.	[213]

**Table 2** (continued)

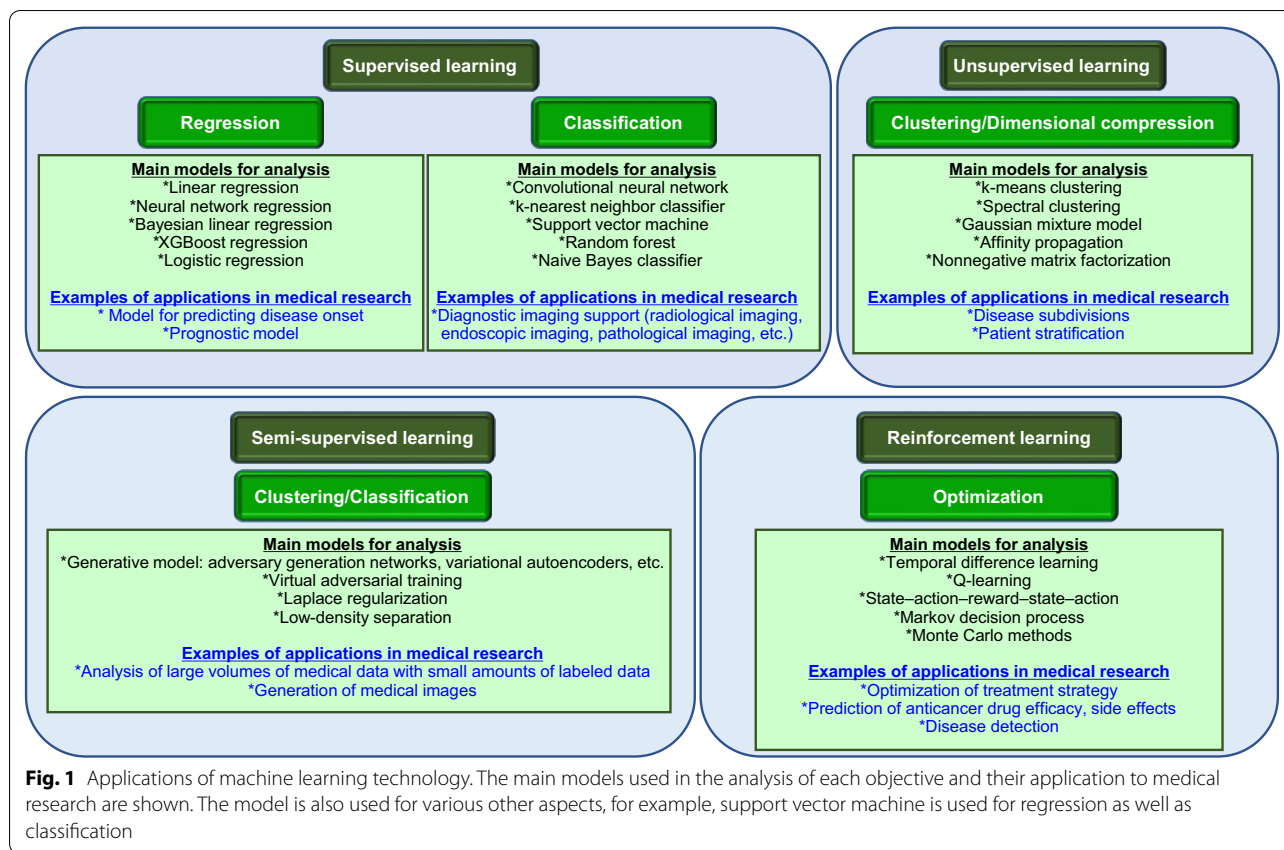
Database name	Operating organization [URL]*	Database contents	References
ClinicalTrials.gov	U.S. National Library of Medicine [ <a href="https://www.clinicaltrials.gov">https://www.clinicaltrials.gov</a> ]	ClinicalTrials.gov is a registry of clinical trials. Operated by the National Library of Medicine (NLM) of the U.S. National Institutes of Health, it is the largest clinical trials database with > 408,000 registered clinical trials in 220 countries. ClinicalTrials.gov was created as a result of the Food and Drug Administration Modernization Act of 1997 (FDAMA), which required the U.S. Department of Health and Human Services (HHS), through the NIH, to verify the effectiveness of experimental drugs for serious or life-threatening diseases and conditions by using the Clinical Trials. The NIH and the FDA jointly developed the site, which became available to the public in February 2000, to establish a registry of information on federal and private clinical trials conducted pursuant to an application.	[214]
BRCA Exchange	Global Alliance for Genomics and Health [ <a href="https://brcaexchange.org/">https://brcaexchange.org/</a> ]	This database aims to improve our understanding of the genetic basis of breast, ovarian, pancreatic, and other cancers by compiling BRCA1/2 gene mutations and corresponding clinical data from around the world. It is possible to search for BRCA1 or BRCA2 variants online.	[215]
LOVD	Molecular Health [ <a href="https://www.lovd.nl/">https://www.lovd.nl/</a> ]	The database provides a flexible and freely available tool to display gene-centric collections and DNA mutations. It also provides storage for patient-centric and NGS data, and extragenic mutations. LOVD is open source, released under the GPL license, and is being actively improved.	[216]

\*This information is current as of September 2022

Muñoz et al. developed BoostDM, a machine learning-based methodology for in silico saturation mutagenesis of cancer genes to assess the carcinogenicity of mutations in human tissues (Fig. 2A) [136]. In silico saturation mutagenesis is a term that generally refers to the computational assessment of all possible changes in a gene or protein. BoostDM defines a supervised learning strategy based on mutations observed in sequenced tumors and their annotation by site-specific mutation features, comparing mutations observed in genes with sufficiently high result type-specific excess (by dNdScv) with randomly selected mutations according to three-base mutation probabilities. This method examines the protein-coding sequence of the genome, and all considered mutations are mapped to the canonical transcript of the protein-coding gene according to Ensembl Variant Effect Predictor (VEP.92) [137]. Gene-tumor type-specific BoostDM models can be complemented with other models trained on pooled mutations from relevant tumor types and used to classify mutations observed in a patient's tumor into drivers and passengers, an important step toward precision cancer medicine. According to the ClinVar and OncoKB databases, only 6,886 and 5,136 (12% and 9%) of the 55,729 coding variants in 568 cancer genes in 28,076 tumor samples are considered drivers (pathogenic or potentially pathogenic) or passengers (benign or potentially benign), respectively. In contrast, more than half of the mutations can be interpreted using the BoostDM

model (26% via gene-tumor type-specific models). The BoostDM model is incorporated into the cancer genome interpreter (CGI; <https://www.cancergenomeinterpreter.org/>), a system designed to assist in the interpretation of newly sequenced tumor genomes.

Motzer et al. performed an integrative, multi-omics analysis of 823 tumors from patients with advanced renal cell carcinoma (RCC) and identified molecular subsets associated with differences in clinical outcomes with angiogenesis inhibitors alone or in combination with immune checkpoint inhibitors (Fig. 2B) [138]. In this study, to better understand the biology of RCC, an RNA-seq dataset of 823 tumor samples from patients with advanced RCC, including 134 tumor samples with sarcoma-like features, obtained from a randomized international phase III trial (IMmotion151) [139], was used to classify patients into seven clusters by utilizing non-negative matrix factorization (NMF). NMF is a machine learning method [140], and unsupervised clustering was used to identify subtypes with different angiogenesis, immunity, cell cycle, metabolism, and stroma programs. Results showed that VEGF receptor tyrosine kinase (sunitinib) and angiogenesis inhibitor (bevacizumab, anti-VEGF) + immune checkpoint inhibitor (atezolizumab) were effective in subsets with high angiogenesis, and bevacizumab + atezolizumab had improved clinical efficacy in tumors with high T effectors and cell cycle transcription. Somatic mutations in the PBRM1



(See figure on next page.)

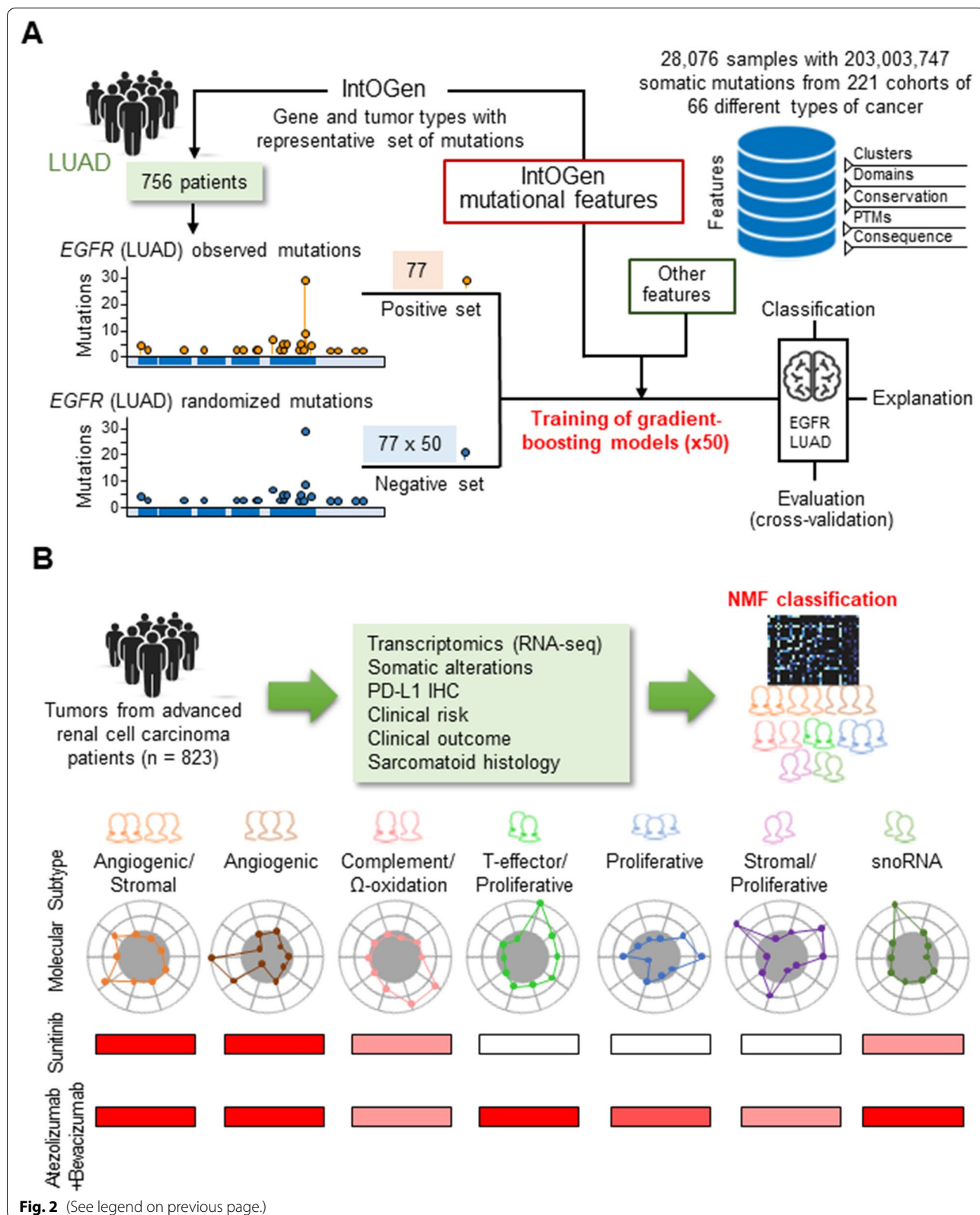
**Fig. 2** Examples of platforms for the prediction of biological significance for genetic abnormalities and its application to diagnosis. **A** Schematic diagram showing an overview of BoostDM modified from Ref. [136]. A specific model (gradient boosted tree) was constructed for each of the 282 gene-tissue combinations based on 18 features that characterize the mechanism of tumorigenesis of oncogenes. Specifically, 50 basic classifiers were trained on random subsets with equal numbers of positive and negative mutations to adequately represent the diversity of passenger mutations and prevent overfitting. **B** Schematic diagram of the identification of seven molecular subtypes of RCC tumors utilizing machine learning, modified from Ref. [138]. An integrative, multi-omics analysis of 823 tumors from RCC patients identified molecular subsets associated with differences in clinical outcomes with angiogenesis inhibitors alone or in combination with checkpoint inhibitors. Unsupervised transcriptome analysis using NMF revealed seven molecular subsets with different angiogenic, immune, cell cycle, metabolic, and stromal programs

and KDM5C genes were associated with high angiogenesis and AMPK/fatty acid participation gene expression, while changes in CDKN2A/B and TP53 were also associated with an increased cell cycle and anabolic metabolism. Sarcomas have a lower prevalence of PBRM1 mutations and angiogenic markers, higher frequency of CDKN2A/B mutations, and increased PD-L1 expression. These findings can be applied to the molecular stratification of patients, improving the prognosis of sarcomas by combining checkpoint inhibitors with angiogenesis inhibitors, and developing personalized medicine in RCC and other indications.

Substitutional mutations in tumors have been reported to account for 95% of somatic mutations, 90% of which

are missense mutations [141]. Substitutional mutations are further classified into driver mutations that favor cancer cell growth and passenger mutations that do not contribute to growth. Since the emergence of driver mutations and cancer heterogeneity are key factors in overcoming treatment resistance and treatment failure, distinguishing whether a substitution mutation is a driver or passenger mutation is an important challenge. Therefore, Dragomir et al. developed and reported a new method (DRIVE) that utilizes machine learning techniques to identify driver and passenger mutations (Fig. 3 A) [142]. Mutation-level characteristics are based on pathogenicity scores, while gene-level characteristics include the maximum number of protein-protein





interactions, biological processes, and types of post-translational modifications. To validate the ability of the proposed method, it was evaluated on a benchmark dataset, which showed that both gene- and mutation-level features were representative of driver mutations, and the proposed method was > 80% accurate in finding the true mutation type. The results suggest that machine learning methods can be used to gain knowledge from mutation data to achieve more targeted cancer treatments [142].

Cancer immunotherapy, represented by immune checkpoint inhibitors, is a treatment that can induce the immune system to effectively recognize and attack tumors. The main approved drugs are antibodies that target CTLA-4 and PD-1/PD-L1 and can induce sustained responses in patients with advanced cancer [143, 144]. However, clinical benefit has not been achieved in many patients, highlighting the need to identify patients who will respond to immunotherapy [145, 146]. Chowell et al. integrated genomic, molecular, demographic, and clinical data from a comprehensive curated cohort (MSK-IMPACT) of 1,479 patients treated with immune checkpoint inhibitors in 16 different cancer types to develop a machine learning platform (Fig. 3B) [147]. Using random forests as the machine learning technique, the platform achieved high sensitivity and specificity in predicting clinical response to immunotherapy in a retrospective analysis, predicting both overall survival (OS) and progression-free survival in test data across different cancer types. The analysis platform also significantly outperforms the tumor mutation burden-based predictions recently approved by the U.S. FDA for predicting immune checkpoint inhibitor responses and can quantitatively assess the most salient model features for prediction.

### Integrated analysis of EHR data using AI and its application to diagnosis and treatment

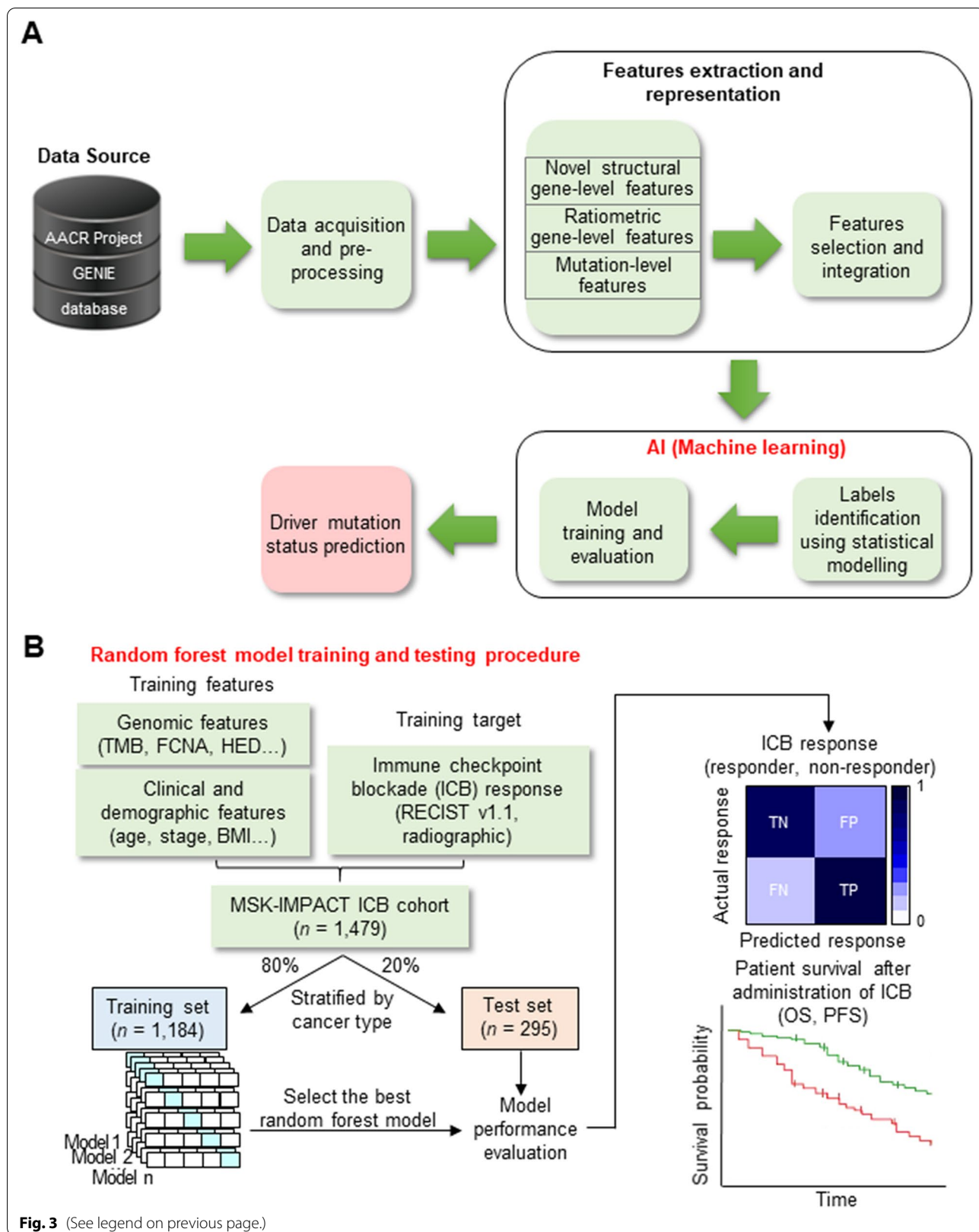
Modern healthcare systems generate and store vast amounts of digital information and have great potential for personalizing and improving healthcare delivery [148, 149]. Morin et al. developed a secure, comprehensive, dynamic, and scalable infrastructure called MEDomics designed to continuously capture multimodal electronic medical information across large, complex healthcare

networks (Fig. 4) [150]. MEDomics maintains structural data that encapsulates the entire timeline of a particular individual's medical care, and this cross-sectional profile can be used to develop a variety of AI applications aimed at practical interventions that can be returned to the healthcare system. Utilizing the MEDomics profile, an institution-wide mortality study in breast and lung cancer patients revealed correlations of mortality by stage and other factors consistent with the published literature. The impact of targeted and immunologic therapies on survival in metastatic breast and lung cancer patients was also investigated. In addition, this infrastructure allowed us to investigate the impact of previously reported non-oncologic risk factors, such as the Framingham cardiovascular risk score, on mortality in cancer patients. This indicates that MEDomics is not only useful for continuous learning, but also for generating and testing clinical hypotheses. Importantly, the study also used statistical learning to create a prognostic model to predict mortality with a high degree of accuracy. Furthermore, utilizing a chronological natural language processing approach, more electronic medical records were incorporated as the course of an individual's illness progressed, and accuracy was found to improve over time. Based on these results, we believe that an approach that combines structured and unstructured multimodal health information in a longitudinal context has the potential to facilitate the development of predictive and dynamic AI applications in oncology that improve the quality and duration of life for individuals.

Peterson et al. proposed a model to predict the risk of preventable acute care unit (ACU) after chemotherapy initiation using a machine learning algorithm trained on comprehensive electronic health record (EHR) data (Fig. 5 A) [151]. ACU, including emergency department visits and hospitalizations, accounts for approximately half of all cancer care-related costs in the United States [152, 153]. Not only is ACU costly, but unscheduled ACU negatively impacts a patient's quality of life and result in poor quality care [154, 155]. To improve quality of care, increase transparency, and reduce costs, the Centers for Medicare & Medicaid Services (CMS) introduced the chemotherapy measure (OP-35) [156, 157]. Peterson et al. successfully identified patients at high risk

(See figure on next page.)

**Fig. 3** Introduction of a machine learning-based genetic mutation analysis platform. **A** Schematic diagram showing an overview of DRIVE, a feature-based machine learning platform for pan-cancer assessment of somatic missense mutations, modified from Ref. [142]. This approach uses a total of 51 features spanning the gene and mutation levels. Several state-of-the-art supervised machine learning algorithms were applied to the final dataset, with results presented for the highest performing algorithms, including random forests, logistic regression, extreme gradient boosting, k-nearest neighbors, support vector machines, and multilayer perceptron. **B** Figure outlining a machine learning model for predicting response to immune checkpoint inhibitors, modified from Ref. [147]. Sixteen cancers were individually divided into training (80%) and testing (20%) subsets. To predict (responder and non-responder) immune checkpoint inhibitors, random forest models were trained on multiple genomic, molecular, demographic, and clinical characteristics on the training data using fivefold cross-validation. Consequently, trained models with optimal hyperparameters were evaluated on various performance measures using the test set



**Fig. 3** (See legend on previous page.)

for preventable acute care, the target of the CMS' OP-35 measure, using machine learning models trained on routinely collected medical information, which showed strong predictive performance. After obtaining structured EHR data generated prior to chemotherapy treatment, 80% of the data in the cohort was used to train a machine learning model to predict the risk of ACU after chemotherapy initiation. The remaining 20% of data were used to test the performance of the model by area under the receiver operating characteristics (AUROC) curve. The study included 8,439 patients, 35% of whom developed preventable ACU within 180 days of starting chemotherapy. In the proposed model, patients at risk of preventable ACU were classified by an AUROC of 0.783 (95% CI, 0.761–0.806) [151]. Patients who were hospitalized were identified better than those who visited the emergency room, and key variables included previous hospitalizations, cancer stage, race, laboratory values, and a diagnosis of depression. The analysis showed limited benefit from the inclusion of patient-reported outcome data and demonstrated inequities in outcome and risk models for Black and Medicaid patients. These results indicate that detailed EHR data can be used to identify patients at risk of ACU using machine learning, and the model proposed in this study has the potential to improve cancer treatment outcomes, patient experience, and costs by enabling targeted preventive interventions [151].

In a cohort study of 42,069 lung cancer patients, Yuan et al. extracted key cancer characteristics from structured data and narrative notes by developing a customized NLP tool using EHRs (Fig. 5B) [158]. Predictive analytics research solution and execution (PheCAP) [159] version 1.2.1 was used as the phenotyping program in this study to develop and evaluate an algorithm to classify lung cancer status. PheCAP consists of three main steps: feature extraction based on the Surrogate-Assisted Feature Extraction (SAFE) algorithm, algorithm development based on penalized regression, and algorithm validation to evaluate the accuracy of the algorithm. The initial PheCAP feature data also consisted of coded features identified by domain experts, NLP features identified from online knowledge source articles proposed in SAFE, and medical utilization features measured by total

counts of medical notes. After extracting eastern cooperative oncology group (ECOG) performance status and body mass index information using an electronic medical record numerical data extraction tool, the NLP interpreter for cancer extraction (NICE) tool was developed to infer cancer characteristics, such as stage, histology, diagnosis date, and somatic mutation information, from clinical records including pathology reports, discharge summaries, and progress notes (Fig. 5B). Smoking status is predicted using a classification algorithm. Importantly, the prognostic ability of the final model proposed in this study was statistically significantly superior to the base model AUROC, including gender, age, histology, and stage (1-year prediction: 0.774 [95% CI, 0.758–0.789];  $P < 0.001$ ; 2-year prediction: 0.779 [95% CI, 0.765–0.793];  $P = 0.002$ ; 3-year prediction: 0.780 [95% CI, 0.766–0.796];  $P = 0.002$ ; 4-year prediction: 0.782 [95% CI, 0.767–0.797];  $P = 0.001$ ; 5-year prediction: 0.782 [95% CI, 0.768–0.798];  $P < 0.001$ ). In the test set, the final and basic models had C-indexes of 0.726 and 0.697, respectively. On the calibration plots, the measured probability of OS was generally within 95% CI of the predicted probability of OS. EHRs provide a low-cost means of accessing detailed longitudinal clinical data from large populations, and lung cancer cohorts constructed from EHR data have shown the potential to be a powerful platform for clinical outcomes research.

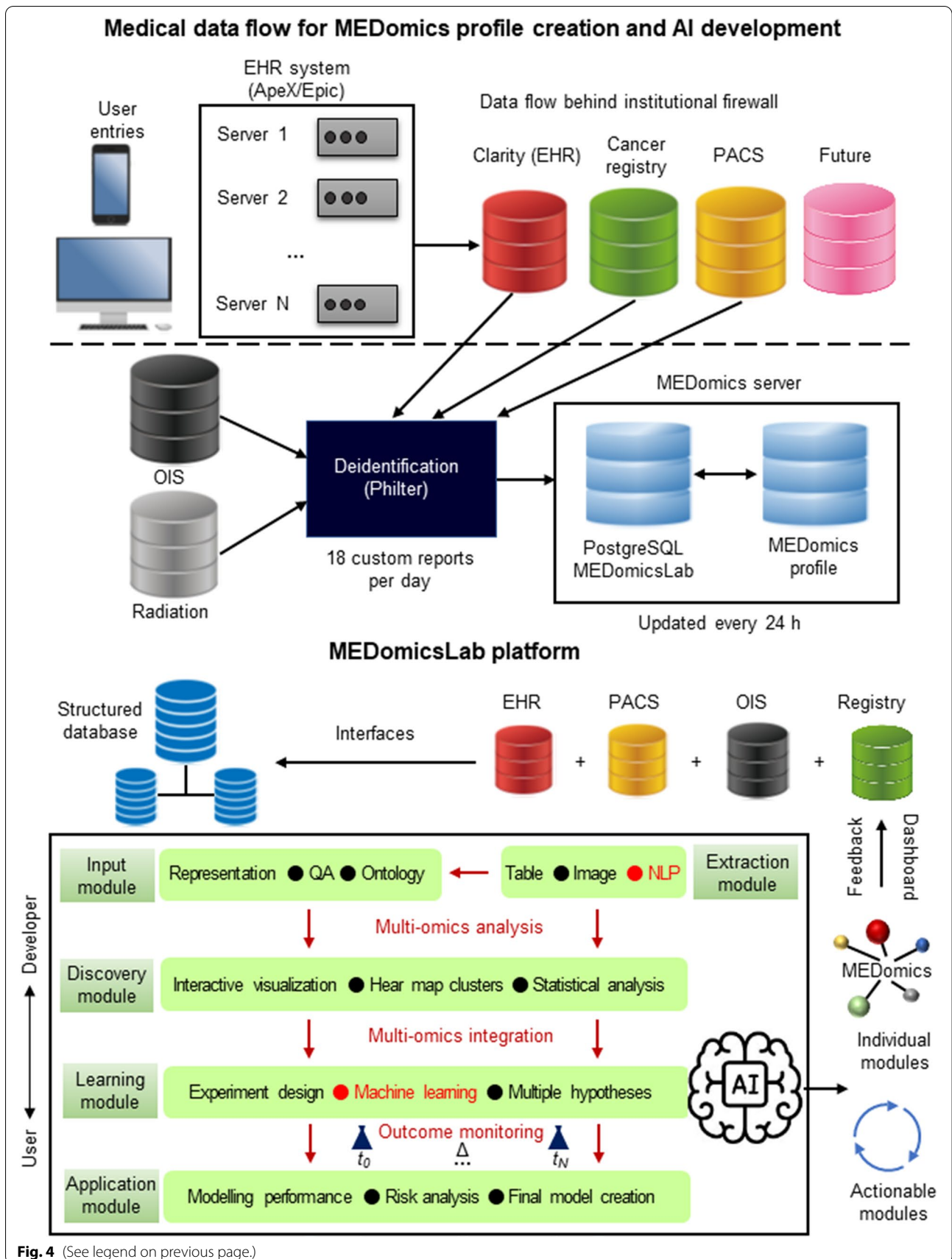
#### AI-based medical article retrieval

In the process of an MTB, it is necessary to refer to the literature when interpreting genetic evidence regarding diagnosis and prognosis, considering basic patient information (age, gender, cancer type, etc.) to address genetic abnormalities and attaching specific candidate drugs and evidence levels [160–162]. On the other hand, the number of medical papers published to date is substantial, and it is a difficult task for humans to extract important information from them. Therefore, research has been conducted to efficiently extract useful information from medical papers using NLP [163–165], one of the AI techniques, and we introduce some recent representative results.

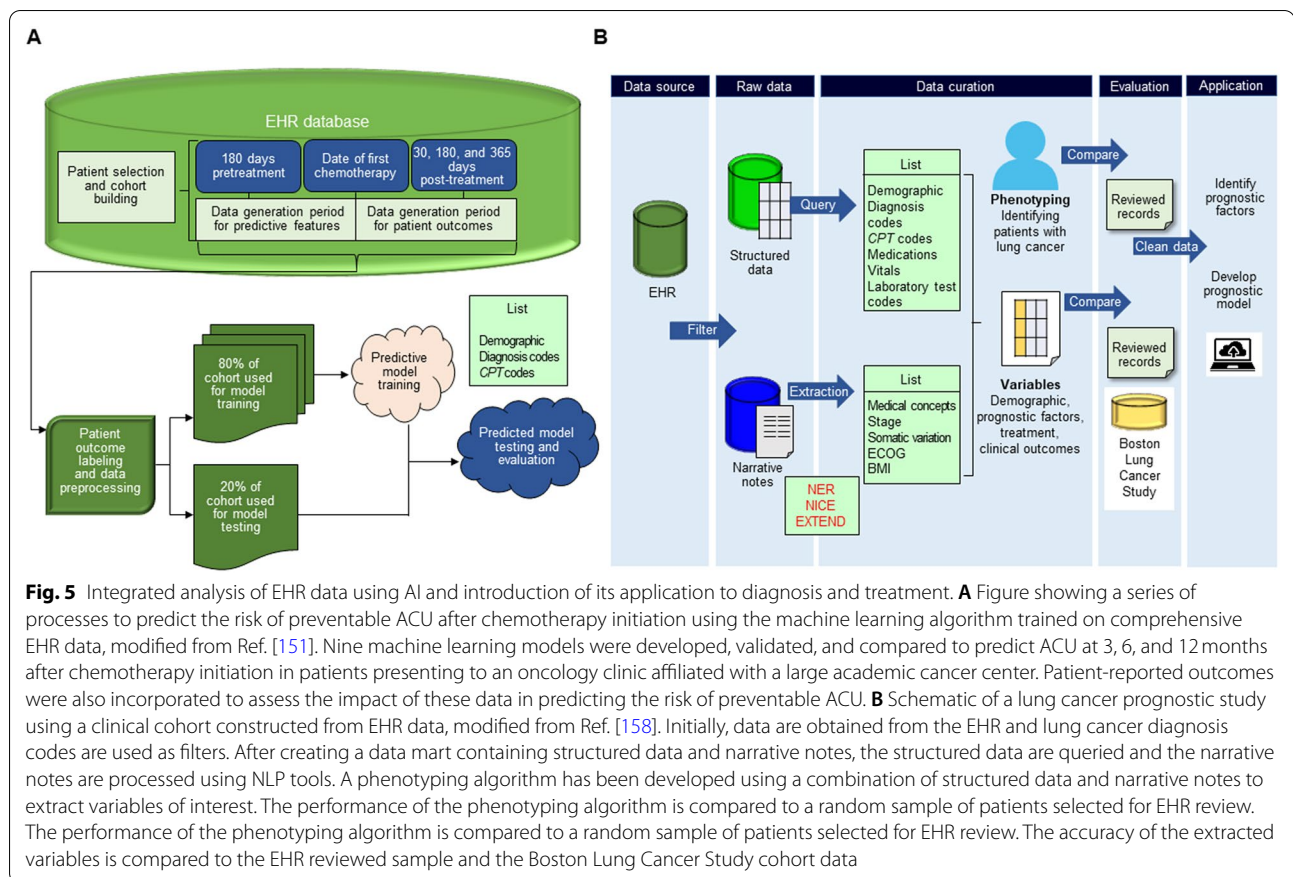
Zeng et al. developed RetriLite, an information retrieval and extraction framework that leverages NLP

(See figure on next page.)

**Fig. 4** Schematic diagram of the flow of medical data for the creation of MEDomics profiles and AI development, modified from Ref. [150]. Medical information generated by medical activities is continuously recorded in the Clarity (EHR) relational database. A custom report is generated from Clarity and sent to the MEDomics server for personal identification, data formatting, and feature extraction and calculation. Various other databases are processed by MEDomics, including tumor information systems (radiation), treatment planning systems, and imaging. Data are ultimately recorded and updated daily on a MEDomics server located behind the facility's firewall with dual authentication access. Medical information flowing from hospital databases is collected in a structured database and becomes input to the MEDomicsLab platform. This information is integrated and processed by the MEDomicsLab engine and used in five computational modules: input, extraction, discovery, learning, and application. This creates a statistical model for Precision Oncology that can be returned to the hospital's database to aid in clinical decision making

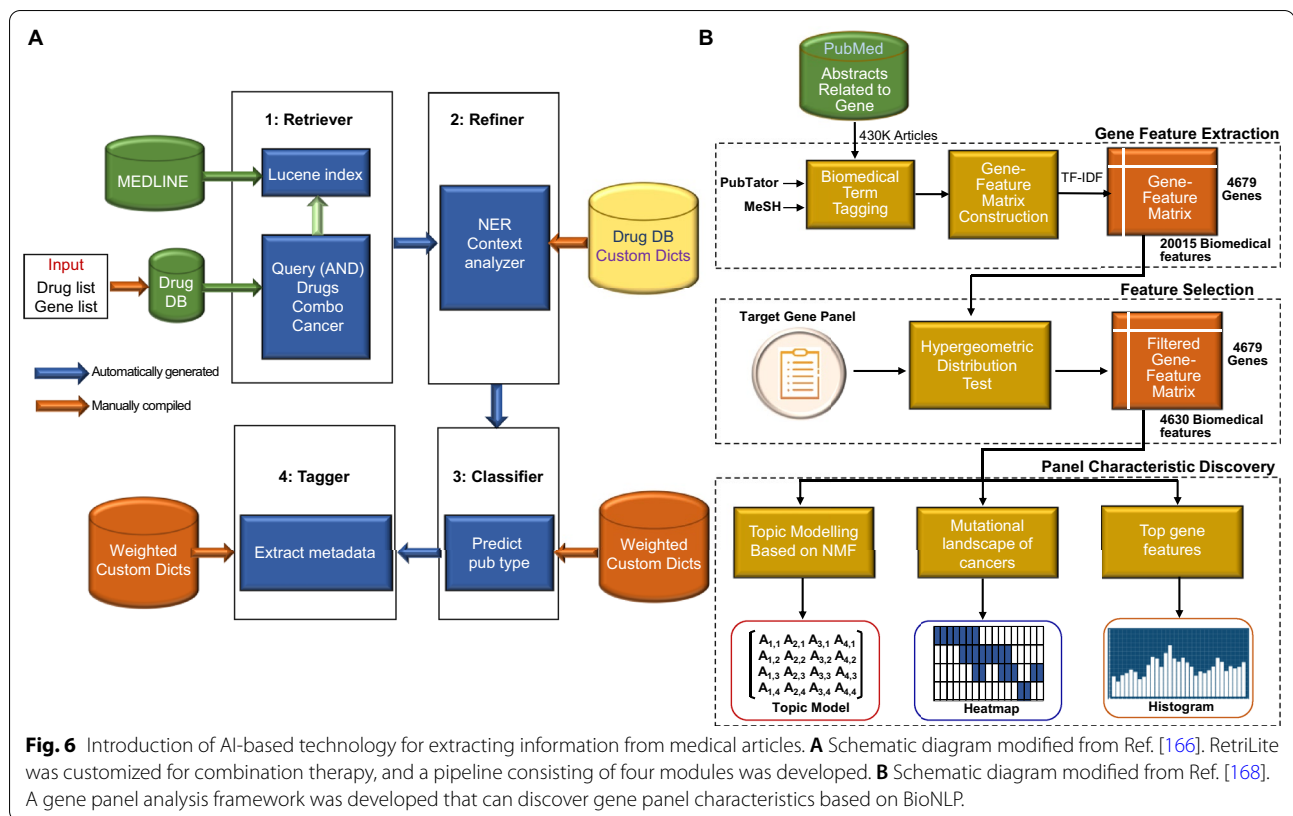


**Fig. 4** (See legend on previous page.)



and domain-specific knowledge bases to computationally identify relevant articles and extract important information [166]. RetriLite features systematically developed automatic query expansion, utilizing domain-specific dictionaries. The National Center for Biotechnology Information (NCBI) Entrez gene database uses gene symbols and aliases, the NCI Thesaurus uses drug names and aliases, and the glossaries developed by major cancer centers use cancer disease dictionaries created by their own institution. It also uses Lucene [167], a state-of-the-art information retrieval library, as the backbone of the application, rendering basic relevance-based ranking, and uses a term frequency, inverse document frequency weighting scheme as the default ranking, where terms matching search terms contribute to a document's relevance score. In addition, RetriLite has a keyword highlighting feature, which conveniently conveys the hidden knowledge used in creating the extended query and may aid in knowledge discovery. A general named entity recognition mechanism has been developed that uses a dictionary for input, recognizes the relevant entities in the text, and normalizes them using canonical terms. Regarding contextual analysis, text segmentation

was applied, and articles where the matched keywords did not appear in the same context were eliminated. Importantly, Zeng et al. customized RetriLite for combination therapy and developed a pipeline consisting of four modules (Fig. 6A) [166]. The "Retriever" is a module that uses gene and drug lists as inputs. For the gene list, it cross-references the institutional drug database to identify clinically available drugs that directly target the gene, along with their aliases. For the drug list, all names are searched, including the aliases associated with each drug. Next, a conjunctive Boolean search query is created, in which three elements coexist: the target drug, concept of cancer, and concept of combination therapy. Keywords related to combination therapy were created by two domain experts. In the second "Refiner" module, the article is considered qualified if it contains at least one sentence in which two drug entries co-occur with the concept of combination therapy, using named entity recognition and contextual analysis to refine the search function. In the third "Classifier" module, a customized weighted terminology dictionary created by the institution is used to classify the main themes of the article as either clinical or preclinical. The fourth "Tagger" module generates relevant metadata tags to facilitate expert



review and help navigate the large corpus. For example, tags have been created for general categories of cancer types related to solid tumors and/or hematological malignancies, drugs matching the search query (anchor drugs), other drugs not included in the original search query but recognized in the context of the combination, types of studies (clinical and/or preclinical), and specific safety-related concepts, such as side effects described in the abstract. Regarding the results, RetriLite achieved an F1 score of 0.93 after more extensive validation experiments to identify drugs with enhanced antitumor effects in vitro or in vivo using poly(ADP-ribose) polymerase inhibitors [166]. Of the articles determined to be relevant by this framework, 95.5% are true positives, achieving an accuracy rate of 97.6% with respect to distinguishing between clinical and preclinical articles. It is also worth mentioning that the inter-observer evaluation achieved a 100% agreement rate [166]. These results indicate that RetriLite is an applicable framework for building domain-specific information retrieval and extraction systems, and its extensive and high-quality metadata tags and keyword highlighting may allow more effective and efficient access to combination therapy information.

Chen et al. applied Biomedical Natural Language Processing (BioNLP) techniques to literature mining of cancer gene panels aimed at creating a pipeline that can

contextualize genes using text-mined co-occurrence features (Fig. 6B) [168]. The gene panel analysis framework was developed in this study. First, PubMed abstracts that mention genes relevant to humans were extracted. This step filters ~430,000 PubMed abstracts on genes from the current full PubMed corpus, which contains ~30 million articles. Second, biomedical named entity recognition is performed on the extracted PubMed abstracts using PubTator and Medical Subject Headings. Third, a genetic term-feature matrix was constructed using biomedical terms, with concepts similar to the document-term matrix. Fourth, to ensure the term features generated in the previous step correspond more strongly to the target gene panel, term feature selection is performed for each individual gene panel. An important aspect of this study is the exploration of hypergeometric distribution. By comparing the frequency distribution of each term feature in the target gene set and the total gene set, it is expected that term features that are more correlated with the target gene panel will be enriched. This approach allows for flexibility with respect to different target gene sets, such as the OncoPrint Cancer Research Panel (OCP) [169] and cardiovascular gene panels [170]. For results using this framework, the cosine similarity of gene frequencies between text mining and statistical results from clinical sequencing data was 80.8%. In the different

machine learning models, the peak accuracies for the prediction of MSK-IMPACT and OCP were 0.959 and 0.989, respectively. Receiver operating characteristic curve analysis also confirmed that the neural network model had better predictive performance (AUROC=0.992) [168]. By using text-mined post-occurrence features, the literature for each gene can be ascertained, and this approach could be used to evaluate several existing gene panels and predict the remaining genes using a portion of the gene panel set, leading to cancer detection.

### Attempts to support MTB by using AIs

Recent successes of AIs in various fields motivated researchers to develop AIs that support MTBs. It is well known that IBM won the game of Jeopardy! in 2011 with a strong NLP technology. The technology enabled the development of remarkable services called Watson for Oncology and Watson for Genomics, where the former is based on disease history and the later on genomic sequencing. They achieved excellent overall consistencies with human experts while reducing doctors' efforts [171–173]. In spite of these achievements, they faced difficulties in the messy reality of healthcare system and its performance depends on race, age, and cancer type [174]. Such difficulties might be universal obstacles for any AIs that aim to support MTBs. Another attempt to support MTBs is a cloud-based virtual molecular tumor board (VMTB) that includes a knowledge base, scoring model, rules engine with >51,000 rules, an asynchronous virtual chat room and a reporting tool [160]. VMTB also reduced time from data receipt to report delivery. In addition, biomarker-driven clinical-trial opportunities were identified for more patients from personalized treatment plans by VMTB than from a commercial lab test alone. However, variability in duration of response to targeted therapy was observed, which might be mitigated with more-explicit consideration of the extent of intra-patient tumor heterogeneity and evolution [175].

### Current challenges and possible future AI-based MTBs

In this review, we introduced the potential of AI implementation in MTBs with a particular focus on the following three areas: (1) AI-based prediction of biological significance for genetic abnormalities and its application to diagnosis and the proposal of therapeutic candidates; (2) AI-based integrated analysis of EHR and omics data and its application to diagnosis and treatment; and (3) AI-based medical article retrieval. Considering the current situation, the use of AI technologies, including machine learning and NLP, is essential for MTBs to proceed smoothly and efficiently, and the active introduction of AI is desirable in the future. On the other hand, there

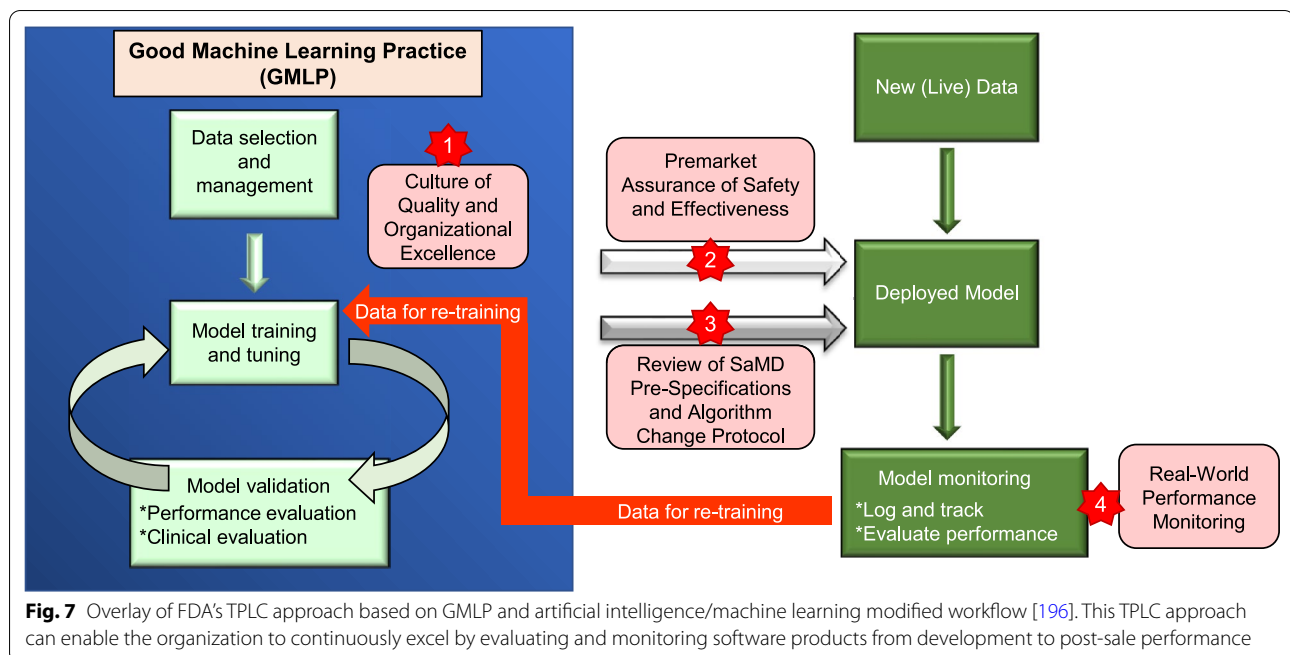
are several issues that must be resolved in the future. The challenges to be addressed are discussed here.

An issue with current cancer genome medicine is the number of patients who can be offered appropriate treatments as a result of genetic testing is limited [176–178]. This is mainly because current cancer genomic medicine is based primarily on a targeted-gene panels coupled with next-generation sequencing (only a limited number of major driver genes are tested). In the future, it is necessary to build a platform for precision medicine based on more omics information, such as whole genome analysis and epigenome information [50, 179]. These are still at the basic research level, and future research aimed at clinical application is desirable. In particular, because a strength of machine learning is its ability to perform multimodal analysis, it is also important to establish a method to integrate and analyze multiple omics information [50].

Second, to date, medical image analysis has been the leading medical AI research and development method, and most AI-based medical device programs approved by the FDA are also targeted at medical image analysis [30, 31, 99, 180, 181]. Compared to medical image analysis, the introduction of AI into omics analysis has not progressed. This is because the nature of omics data itself is difficult to handle, judging from the characteristics of machine learning technology. For example, samples in the medical field are difficult to obtain, and there are limitations on the number of cases that can be analyzed. On the other hand, there are ~30,000 genes, which are the central target of omics analysis, and the target of whole genome analysis consists of three billion base pairs. The number of parameters ( $p$ ) is overwhelmingly large compared to the number of cases, which makes machine learning difficult (called the small  $n$ , large  $p$  problem) [182–184]. In addition, since neighboring pixels tend to have similar information with respect to images, models such as convolutional neural networks, one of the deep learning techniques, are useful [185, 186]. On the other hand, with respect to genomic information, there is often a divergence between chromosomal location (proximity) and functional relatedness (close proximity between genes does not mean that they are functionally related). Consequently, the usefulness of machine learning in medical image analysis is often not observed in omics analysis. New models and analysis platforms for these problems are being developed by our research group and others [100, 102, 187–190], and it is hoped that robust systems that can be applied clinically will be developed in the future.

Third, regarding cancer genome medicine, the volume of data is increasing daily, and the types of anti-cancer drugs are also increasing; therefore, it would be





**Fig. 7** Overlay of FDA's TPLC approach based on GMLP and artificial intelligence/machine learning modified workflow [196]. This TPLC approach can enable the organization to continuously excel by evaluating and monitoring software products from development to post-sale performance

ideal for AI to continuously learn. On the other hand, the approved AI-SaMDs are basically locked AIs that were approved once they stopped learning, therefore they are not adaptive AIs that can continuously learn [191–193]. Various efforts are currently underway worldwide to address this issue. In 2019 in the United States, the FDA published a discussion paper on the regulation and framework for AI-SaMDs, where it proposed “SaMD Pre-Specifications (SPS)”, which describe the expected or planned changes to the device, and “Algorithm Change Protocol (ACP)”, a specific proposal on the methods companies should use to manage the risk of change [194]. A concept for the quality control of programmed medical devices called “Good Machine Learning Practice” (GMLP) was also proposed [195]. To limit degradation while allowing machine learning algorithms to completely leverage their power and continuously improve their performance, the total product lifecycle (TPLC) approach proposed by the FDA and based on GMLP is expected to balance benefits and risks, and enable safe and effective AI-SaMD delivery (Fig. 7) [196]. Subsequently, in January 2021, the FDA issued an action paper on AI-SaMD regulations and frameworks [197, 198] and proposed to issue draft guidance on prescribed change management plans. In addition, in October 2021, with the FDA, Health Canada, and the UK's Medicines and Healthcare products Regulatory Agency jointly identified 10 guiding principles that can inform the development of GMLP and issued a new guidance called “Good Machine Learning Practice for Medical Device Development:

Guiding Principles” [199, 200]. It suggests that regular or ongoing training of the model should be managed to prevent overfitting and unintended bias, as well as to place appropriate controls to manage risk. Since adaptive AI is important for the introduction of AI into MTB, progress in this research area is desirable.

Fourth, evidence analysis for genetic abnormalities based on NGS has been reported to vary considerably among annotation services [135, 201, 202]. For example, there is only moderate agreement between IBM Watson for Genomics (WfG) and OncoKB over their Level 1 treatment action likelihood recommendations [135]. This implies that the accuracy of annotation in tumor profiling tests for solid tumors based on genetic mutation analysis by NGS requires improvement.

In addition to the above, other sensitive issues have been reported, such as the fact that the MTB proposal may be out of sync with the actual trial in which the patient can participate because it does not take into account the patient's medical history (especially drug-induced pneumonitis). Therefore, it is also important that the system can be easily modified to suit the current situation in the clinical field. Furthermore, since this review focuses on the use of AI in MTB, it should also be noted that it presents only limited results among the MTB-related reports so far. Several studies have reported on the various elemental technologies required for MTB, though not specifically focused on AI [203–206], and reviewed AI efforts in clinical diagnosis of cancer (including our previous study) [30, 33, 207, 208].

## Conclusion

This review has shown that AI may be used for various elemental technologies required for MTBs. In particular, the volume of data handled by MTB members is expected to increase in the future, and the introduction of AI into MTB is an urgent requirement to establish a precision medicine system. However, there are several potential challenges of AI, and it is important to progress steadily while solving these challenges individually and simultaneously creating innovative technologies. Imperatively, a win–win relationship between human and AI must be established to create a symbiotic relationship, with a clear understanding of which AI can be beneficial and where it may limit progress.

## Abbreviations

MTB: Molecular tumor board; AI: Artificial intelligence; NGS: Next-generation sequencing; ICT: Information and communication technology; C-CAT: Center for Cancer Genomics and Advanced Therapeutics; CAD: Computer-aided detection/diagnosis; CAde: Computer-aided detection; CADx: Computer-aided diagnosis; ILSVRC: ImageNet Large-Scale Visual Recognition Challenge; FDA: Food and Drug Administration; SaMD: Software as a medical device; NCC: National Cancer Center; RCC: Renal cell carcinoma; NMF: Non-negative matrix factorization; OS: Overall survival; ACU: Acute care unit; EHR: Electronic health record; CMS: Centers for Medicare & Medicaid Services; AUROC: Area under the receiver operating characteristics; NLP: Natural language processing; SAFE: Surrogate-assisted feature extraction; ECOG: Eastern cooperative oncology group; NICE: NLP interpreter for cancer extraction; NCBI: National Center for Biotechnology Information; BioNLP: Biomedical natural language processing; OCP: Oncomine cancer research panel; VMTB: Virtual molecular tumor board; SPS: SaMD pre-specifications; ACP: Algorithm change protocol; GMLP: Good Machine Learning Practice; TPLC: Total product lifecycle; gnomAD: Genome Aggregation Database; COSMIC: the Catalogue Of Somatic Mutations In Cancer; CIVIC: Clinical Interpretation of Variants in Cancer; WfG: IBM Watson for Genomics; NLM: National Library of Medicine; FDAMA: Food and Drug Administration Modernization Act of 1997; HHS: Department of Health and Human Services.

## Acknowledgements

The authors express their gratitude to the past and present members of Division of Medical AI Research and Development, National Cancer Center Research Institute.

## Author contributions

RH, TK, NK, TY, and NY conceived the study. RH, TK, NK, TY, and SY wrote the manuscript. RH, TK, NK, and TY produced the figures. KSudo, MH, KSunami, TK, KT, ST, HM, KK, KA, MK, SK, YY and NY edited and contributed to the final submitted manuscript, and provided critical insights. All authors have read and approved the final manuscript.

## Funding

This work was supported by MHLW ICT infrastructure establishment and implementation of artificial intelligence for clinical and medical research program, Grant Number JP21AC5001 and RIKEN Center for the Advanced Intelligence Project.

## Availability of data and materials

Not applicable.

## Declarations

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Division of Medical AI Research and Development, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. <sup>2</sup>Cancer Translational Research Team, RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan. <sup>3</sup>Department of Experimental Therapeutics, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. <sup>4</sup>Department of Surgery, Graduate School of Medicine, Kyoto University, Yoshida-konoe-cho, Sakyo-ku, Kyoto 606-8303, Japan. <sup>5</sup>Research and Development Group, Hitachi, Ltd., 1-280 Higashi-koigakubo, Kokubunji, Tokyo 185-8601, Japan. <sup>6</sup>Department of Medical Oncology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. <sup>7</sup>Department of Genetic Medicine and Services, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. <sup>8</sup>Department of Laboratory Medicine, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. <sup>9</sup>Department of Diagnostic Pathology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. <sup>10</sup>Division of Molecular Pathology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan.

Received: 31 August 2022 Accepted: 5 October 2022

Published online: 31 October 2022

## References

- Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science*. 2003;300(5617):286–90.
- International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431(7011):931–45.
- McCarthy A. Third generation DNA sequencing: pacific biosciences' single molecule real time technology. *Chem Biol*. 2010;17(7):675–6.
- Even Choev N. Personalized Medicine in Practice: Postgenomics from Multiplicity to Immutability. *Body & Society*. 2019;26(1):26–54.
- Geistlinger J, Ahnert P. Large-scale detection of genetic variation: the key to personalized medicine. In: Knäblein J, editor. *Modern biopharmaceuticals. Design, development, and optimization*. Vol. 1. Weinheim, Germany: Wiley/VCH Verlag; 2005:71–98.
- Offit K. Personalized medicine: new genomics, old lessons. *Hum Genet*. 2011;130(1):3–14.
- Morganti S, Tarantino P, Ferraro E, D'Amico P, Duso BA, Curigliano G. Next generation sequencing (NGS): a revolutionary technology in pharmacogenomics and personalized medicine in cancer. *Adv Exp Med Biol*. 2019;1168:9–30.
- Mosele F, Remon J, Mateo J, Westphalen CB, Barlesi F, Lolkema MP, et al. Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group. *Ann Oncol*. 2020;31(11):1491–505.
- Hulick PJ. Next-generation DNA sequencing (NGS): Principles and clinical applications. UpToDate; 2018. Available online: <https://www.uptodate.com/contents/next-generation-dna-sequencing-ngs-principles-and-clinical-applications>. Accessed 10 Sept 2022.
- van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet*. 2014;30(9):418–26.
- Sheridan C. Illumina claims \$1000 genome win. *Nat Biotechnol*. 2014;32(2):115.
- Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372(9):793–5.
- Ashley EA. Towards precision medicine. *Nat Rev Genet*. 2016;17(9):507–22.
- Nassar SF, Raddassi K, Ubhi B, Doktorski J, Abulaban A. Precision medicine: steps along the road to combat human cancer. *Cells*. 2020;9(9):2056.
- Cheng ML, Berger MF, Hyman DM, Solit DB. Clinical tumour sequencing for precision oncology: time for a universal strategy. *Nat Rev Cancer*. 2018;18(9):527–8.

16. Jean NS, Pinto C, Tenente I, Murray G. Collaboration is key to accelerating diagnostics access to optimize benefits of precision medicines. *Per Med*. 2018;15(3):157–61.
17. Vranic S, Gatalica Z. The role of pathology in the era of personalized (precision) medicine: a brief review. *Acta Med Acad*. 2021;50(1):47–57.
18. Mizuno T, Yoshida T, Sunami K, Koyama T, Okita N, Kubo T, et al. Study protocol for NCCH1908 (UPFRONT-trial): a prospective clinical trial to evaluate the feasibility and utility of comprehensive genomic profiling prior to the initial systemic treatment in advanced solid tumour patients. *Jpn J Clin Oncol*. 2021;51(12):1757–60.
19. Ebi H, Bando H. Precision oncology and the universal health coverage system in Japan. *JCO Precis Oncol*. 2019;3:1–12.
20. Ito M, Fujiwara Y, Kubo T, Matsushita H, Kumamoto T, Suzuki T, et al. Clonal hematopoiesis from next generation sequencing of plasma from a patient with lung adenocarcinoma: a case report. *Front Oncol*. 2020;10:113.
21. Tamborero D, Dienstmann R, Rachid MH, Boekel J, Baird R, Brana I, et al. Support systems to guide clinical decision-making in precision oncology: the Cancer Core Europe Molecular Tumor Board Portal. *Nat Med*. 2020;26(7):992–4.
22. Tamborero D, Dienstmann R, Rachid MH, Boekel J, Lopez-Fernandez A, Jonsson M, et al. The Molecular Tumor Board Portal supports clinical decisions and automated reporting for precision oncology. *Nat cancer*. 2022;3(2):251–61.
23. Lauk K, Peters M-C, Velthaus J-L, Nürnberg S, Ueckert F. Use of process modelling for optimization of molecular tumor boards. *Appl Sci*. 2022;12(7):3485.
24. Gebbia V, Guarini A, Piazza D, Bertani A, Spada M, Verderame F, et al. Virtual multidisciplinary tumor boards: a narrative review focused on lung cancer. *Pulm Ther*. 2021;7(2):295–308.
25. Blasi L, Bordonaro R, Serretta V, Piazza D, Firenze A, Gebbia V. Virtual clinical and precision medicine tumor boards-cloud-based platform-mediated implementation of multidisciplinary reviews among oncology centers in the COVID-19 era: protocol for an observational study. *JMIR Res Protoc*. 2021;10(9):e26220.
26. Hopkins SE, Vidri RJ, Hill MV, Vijayvergia N, Farma JM. A virtual tumor board platform: a way to enhance decision-making for complex malignancies. *J Surg Res*. 2022;278:233–9.
27. Sarker IH, Furhad MH, Nowrozy R. AI-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Comput Sci*. 2021;2(3):1–18.
28. Li P, Ning Y, Fang H. Artificial intelligence translation under the influence of multimedia teaching to study English learning mode. *Int J Electr Eng Educ*. 2021;13:002072092098352.
29. Wang H, Hao L, Sharma A, Kukkar A. Automatic control of computer application data processing system based on artificial intelligence. *J Intell Syst*. 2022;31(1):177–92.
30. Hamamoto R, Suvarna K, Yamada M, Kobayashi K, Shinkai N, Miyake M, et al. Application of artificial intelligence technology in oncology: towards the establishment of precision medicine. *Cancers (Basel)*. 2020;12(12):3532.
31. Benjamens S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*. 2020;3(1):1–8.
32. Komatsu M, Sakai A, Dozen A, Shozu K, Yasutomi S, Machino H, et al. Towards clinical application of artificial intelligence in ultrasound imaging. *Biomedicines*. 2021;9(7):720.
33. Yamada M, Saito Y, Yamada S, Kondo H, Hamamoto R. Detection of flat colorectal neoplasia by artificial intelligence: a systematic review. *Best Pract Res Clin Gastroenterol*. 2021;52:101745.
34. Muehlemaier UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health*. 2021;3(3):e195–203.
35. Asada K, Komatsu M, Shimoyama R, Takasawa K, Shinkai N, Sakai A, et al. Application of artificial intelligence in COVID-19 diagnosis and therapeutics. *J Personalized Med*. 2021;11(9):886.
36. Kobayashi K, Hataya R, Kurose Y, Miyake M, Takahashi M, Nakagawa A, et al. Decomposing normal and abnormal features of medical images for content-based image retrieval of glioma imaging. *Med Image Anal*. 2021;74:102227.
37. Takahashi S, Takahashi M, Kinoshita M, Miyake M, Kawaguchi R, Shinjima N, et al. Fine-tuning approach for segmentation of gliomas in brain magnetic resonance images with a machine learning method to normalize image differences among facilities. *Cancers (Basel)*. 2021;13(6):1415.
38. Yamada M, Saito Y, Imaoka H, Saiko M, Yamada S, Kondo H, et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci Rep*. 2019;9(1):14465.
39. Komatsu M, Sakai A, Komatsu R, Matsuoka R, Yasutomi S, Shozu K, et al. Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. *Appl Sci*. 2021;11(1):371.
40. Dozen A, Komatsu M, Sakai A, Komatsu R, Shozu K, Machino H, et al. Image segmentation of the ventricular septum in fetal cardiac ultrasound videos based on deep learning using time-series information. *Biomolecules*. 2020;10(11):1526.
41. Shozu K, Komatsu M, Sakai A, Komatsu R, Dozen A, Machino H, et al. Model-agnostic method for thoracic wall segmentation in fetal ultrasound videos. *Biomolecules*. 2020;10(12):1691.
42. Jinnai S, Yamazaki N, Hirano Y, Sugawara Y, Ohe Y, Hamamoto R. The development of a skin cancer classification system for pigmented skin lesions using deep learning. *Biomolecules*. 2020;10(8):1123.
43. Salim M, Wählin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. *JAMA Oncol*. 2020;6(10):1581–8.
44. Lizzi F, Atzori S, Aringhieri G, Bosco P, Marini C, Retico A, et al. Residual convolutional neural networks for breast density classification. In: *International Conference on Computer Analysis of Images and Patterns*. Berlin/Heidelberg, Germany: Springer; 2019:258–63.
45. Azuaje F, Kim S-Y, Perez Hernandez D, Dittmar G. Connecting histopathology imaging and proteomics in kidney cancer through machine learning. *J Clin Med*. 2019;8(10):1535.
46. Hamamoto R. Application of artificial intelligence for medical research. *Biomolecules*. 2021;11(1):90.
47. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J biomedical health Inf*. 2017;22(5):1589–604.
48. Asada K, Takasawa K, Machino H, Takahashi S, Shinkai N, Bolatkan A, et al. Single-cell analysis using machine learning techniques and its application to medical research. *Biomedicines*. 2021;9(1):1513.
49. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J Am Med Inform Assoc*. 2018;25(10):1419–28.
50. Hamamoto R, Komatsu M, Takasawa K, Asada K, Kaneko S. Epigenetics analysis and integrated analysis of multiomics data, including epigenetic data. Using *Artif Intell Era Precision Med Biomolecules*. 2020;10(1):62.
51. Xu N. Understanding the reinforcement learning. *J Phys Conf Ser*. 2019;1207(1):012014.
52. Dalal KR. Analysing the role of supervised and unsupervised machine learning in iot. In: *2020 international conference on electronics and sustainable communication systems (ICESC) 2020*. p. 75–9.
53. Peng L, Chen Z, Chen T, Lei L, Long Z, Liu M, et al. Prediction of the age at onset of spinocerebellar ataxia type 3 with machine learning. *Mov Disord*. 2021;36(1):216–24.
54. Xu X, Zhang J, Yang K, Wang Q, Chen X, Xu B. Prognostic prediction of hypertensive intracerebral hemorrhage using CT radiomics and machine learning. *Brain and Behavior*. 2021;11(5):e02085.
55. Zhang H, Chen D, Shao J, Zou P, Cui N, Tang L, et al. Machine learning-based prediction for 4-year risk of metabolic syndrome in adults: a retrospective cohort study. *Risk Manage Healthc Policy*. 2021;14:4361.
56. Yang LS, Perry E, Shan L, Wilding H, Connell W, Thompson AJ, et al. Clinical application and diagnostic accuracy of artificial intelligence in colonoscopy for inflammatory bowel disease: systematic review. *Endoscopy Int open*. 2022;10(07):E1004–13. <https://doi.org/10.1055/a-1846-0642>.
57. Bang CS, Lim H, Jeong HM, Hwang SH. Use of endoscopic images in the prediction of submucosal invasion of gastric neoplasms: automated deep learning model development and usability study. *J Med Internet Res*. 2021;23(4):e25167.

58. Yu W, Hargreaves CA. A review study of the deep learning techniques used for the classification of chest radiological images for Covid-19 diagnosis. *Int J Inf Manage Data Insights* 2022:100100.
59. Peng S-J, Chen Y-W, Yang J-Y, Wang K-W, Tsai J-Z. Automated cerebral infarct detection on computed tomography images based on deep learning. *Biomedicines*. 2022;10(1):122.
60. Cai Y-W, Dong F-F, Shi Y-H, Lu L-Y, Chen C, Lin P, et al. Deep learning driven colorectal lesion detection in gastrointestinal endoscopic and pathological imaging. *World J Clin Cases*. 2021;9(31):9376.
61. Zhang H, Ren F, Wang Z, Rao X, Li L, Hao J, et al. Predicting tumor mutational burden from liver cancer pathological images using convolutional neural network. In: 2019 IEEE international conference on bioinformatics and biomedicine (BIBM) 2019. p. 920-5.
62. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. *Exp Dermatol*. 2018;27(11):1261-7.
63. Goyal M, Knackstedt T, Yan S, Hassanpour S. Artificial intelligence-based image classification methods for diagnosis of skin cancer: challenges and opportunities. *Comput Biol Med*. 2020;127:104065.
64. Chen J, Milot L, Cheung H, Martel AL. Unsupervised clustering of quantitative imaging phenotypes using autoencoder and gaussian mixture model. In: International conference on medical image computing and computer-assisted intervention 2019. p. 575-82.
65. Salgado CM, Vieira SM. Machine learning for patient stratification and classification part 2: unsupervised learning with clustering. *Leveraging data science for global health* 2020. p. 151-68.
66. Li H, Galperin-Aizenberg M, Pryma D, Simone IICB, Fan Y. Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy. *Radiother Oncol*. 2018;129(2):218-26.
67. Li J, Cui L, Tu L, Hu X, Wang S, Shi Y, et al. Research of the distribution of tongue features of diabetic population based on unsupervised learning technology. *Evid Based Complement Altern Med*. 2022;7684714. <https://doi.org/10.1155/2022/7684714>.
68. Hassan NS, Abdulazeez AM, Zeebaree DQ, Hasan DA. Medical images breast cancer segmentation based on K-means clustering algorithm: a review. *Ultrasound*. 2021;27:28.
69. Ma T, Zhang A. Affinity network fusion and semi-supervised learning for cancer patient clustering. *Methods*. 2018;145:16-24.
70. Wang Q, Xia L-Y, Chai H, Zhou Y. Semi-supervised learning with ensemble self-training for cancer classification. In: 2018 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) 2018. p. 796-803.
71. Wenger K, Tirdad K, Cruz AD, Mari A, Basheer M, Kuk C, et al. A semi-supervised learning approach for bladder cancer grading. *Mach Learn Appl*. 2022;9:100347. <https://doi.org/10.1016/j.mlwa.2022.100347>.
72. Li X, Jiang Y, Rodriguez-Andina JJ, Luo H, Yin S, Kaynak O. When medical images meet generative adversarial network: recent development and research opportunities. *Discover Artif Intell*. 2021;1(1):1-20.
73. Chen Y, Yang X-H, Wei Z, Heidari AA, Zheng N, Li Z, et al. Generative adversarial networks in medical image augmentation: a review. *Comput Biol Med*. 2022;105382. <https://doi.org/10.1016/j.combiomed.2021.105063>.
74. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518(7540):529-33.
75. He Z, Li L, Zheng S, Li Y, Situ H. Variational quantum compiling with double Q-learning. *New J Phys*. 2021;23(3):033002.
76. Daoud S, Mdhaffar A, Jmaiel M, Freisleben B. Q-rank: reinforcement learning for recommending algorithms to predict drug sensitivity to cancer therapy. *IEEE J Biomedical Health Inf*. 2020;24(11):3154-61.
77. Liu M, Shen X, Pan W. Deep reinforcement learning for personalized treatment recommendation. *Stat Med*. 2022;41(20):4034-56.
78. Ribba B, Dudal S, Lavé T, Peck RW. Model-informed artificial intelligence: reinforcement learning for precision dosing. *Clin Pharmacol Ther*. 2020;107(4):853-7.
79. Tseng HH, Luo Y, Cui S, Chien JT, Ten Haken RK, Naqa IE. Deep reinforcement learning for automated radiation adaptation in lung cancer. *Med Phys*. 2017;44(12):6690-705.
80. Padmanabhan R, Meskin N, Haddad WM. Reinforcement learning-based control of drug dosing with applications to anesthesia and cancer therapy. *Control applications for biomedical engineering systems* 2020. p. 251-97.
81. Firmino M, Angelo G, Morais H, Dantas MR, Valentim R. Computer-aided detection (CADE) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. *Biomed Eng Online*. 2016;15(1):1-17.
82. Moura DC, López MAG, Cunha P, Posada NGD, Pollan RR, Ramos I, et al. Benchmarking datasets for breast cancer computer-aided diagnosis (CADx). *Iberoamerican congress on pattern recognition*. 2013. p. 326-33.
83. Yanase J, Triantaphyllou E. A systematic survey of computer-aided diagnosis in medicine: past and present developments. *Expert Syst Appl*. 2019;138:112821.
84. Wang L, Yu L. Introductory chapter: computer-aided diagnosis for biomedical applications. In: *Computer architecture in industrial, biomechanical and biomedical engineering*. 2019. p. 1.
85. Doi K. Overview on research and development of computer-aided diagnostic schemes. *Semin Ultrasound CT MRI*. 2004;25(5):404-10.
86. Jarvis T, Thornburg D, Rebecca AM, Teven CM. Artificial intelligence in plastic surgery: current applications, future directions, and ethical implications. *Plast Reconstr Surg Glob Open*. 2020;8(10):e3200.
87. Yu VL. Antimicrobial selection by a computer. *JAMA*. 1979;242(12):1279.
88. Gillies A, Smith P. Can AI systems meet the ethical requirements of professional decision-making in health care? *AI Ethics*. 2021;2(1):41-7.
89. Fujita H, Uchiyama Y, Nakagawa T, Fukuoka D, Hatanaka Y, Hara T, et al. Computer-aided diagnosis: the emerging of three CAD systems induced by Japanese health care needs. *Comput Methods Programs Biomed*. 2008;92(3):238-48.
90. Retson TA, Eghtedari M. Computer-aided detection/diagnosis in breast imaging: a focus on the evolving FDA regulations for using software as a medical device. *Curr Radiol Rep*. 2020;8(6):1-7.
91. Giger ML, Chan HP, Boone J. Anniversary paper: history and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. *Med Phys*. 2008;35(12):5799-820.
92. Castellino RA. Computer aided detection (CAD): an overview. *Cancer Imaging*. 2005;5:17-9.
93. Summers RM. Evaluation of computer-aided detection devices: consensus is developing. *Acad Radiol*. 2012;19(4):377-9.
94. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006;313:504-7.
95. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. A state-of-the-art survey on deep learning theory and architectures. *Electronics*. 2019;8(3):292.
96. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). 2016. p. 770-8.
97. Bali J, Garg R, Bali RT. Artificial intelligence (AI) in healthcare and biomedical research: Why a strong computational/AI bioethics framework is required? *Indian J Ophthalmol*. 2019;67(1):3-6.
98. The US, Food and Drug Administration (FDA). Artificial intelligence and machine learning (AI/ML)-enabled medical devices. 2022. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-ai-ml-enabled-medical-devices>.
99. Acosta JN, Falcone GJ, Rajpurkar P. The need for medical artificial intelligence that incorporates prior images. *Radiology*. 2022;304(2):283-288. <https://doi.org/10.1148/radiol.212830>.
100. Asada K, Kobayashi K, Joutard S, Tubaki M, Takahashi S, Takasawa K, et al. Uncovering prognosis-related genes and pathways by multi-omics analysis in lung cancer. *Biomolecules*. 2020;10(4):524.
101. Kenner BJ, Abrams ND, Chari ST, Field BF, Goldberg AE, Hoos WA, et al. Early detection of pancreatic cancer: applying artificial intelligence to electronic health records. *Pancreas*. 2021;50(7):916-22.
102. Kobayashi K, Bolatkan A, Shiina S, Hamamoto R. Fully-connected neural networks with reduced parameterization for predicting histological types of lung cancer from somatic mutations. *Biomolecules*. 2020;10(9):1249.
103. Zheng Y, Dickson VV, Blecker S, Ng JM, Rice BC, Melkus GD, et al. Identifying patients with hypoglycemia using natural language processing: systematic literature review. *JMIR Diabetes*. 2022;7(2):e34681.

104. Takahashi S, Takahashi M, Tanaka S, Takayanagi S, Takami H, Yamazawa E, et al. A new era of neuro-oncology research pioneered by multi-omics analysis and machine learning. *Biomolecules*. 2021;11(4):565.
105. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: a review. *Biotechnol Adv*. 2021;49:107739.
106. Chowdhary K. Natural language processing. In: *Fundamentals of Artificial Intelligence*. Springer. 2020, pp. 603–49.
107. Huang K, Hussain A, Wang Q-F, Zhang R. Deep learning: fundamentals, theory and applications. 2019. p. 2.
108. Chen X, Xie H, Wang FL, Liu Z, Xu J, Hao T. A bibliometric analysis of natural language processing in medical research. *BMC Med Inf Decis Mak*. 2018;18(1):1–14.
109. Bitterman DS, Miller TA, Mak RH, Savova GK. Clinical natural language processing for radiation oncology: a review and practical primer. *Int J Radiation Oncology\* Biology\* Phys*. 2021;110(3):641–55.
110. Hughes KS, Zhou J, Bao Y, Singh P, Wang J, Yin K. Natural language processing to facilitate breast cancer research and management. *Breast J*. 2020;26(1):92–9.
111. Hao T, Huang Z, Liang L, Weng H, Tang B. Health natural language processing: methodology development and applications. *JMIR Med Inf*. 2021;9(10):e23898.
112. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: a review. *Trends Anaesth Crit Care*. 2021;38:4–9.
113. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ*. 2015;350:H1885. <https://doi.org/10.1136/bmj.h1885>.
114. Zeng J, Banerjee I, Henry AS, Wood DJ, Shachter RD, Gensheimer MF, et al. Natural language processing to identify cancer treatments with electronic medical records. *JCO Clin Cancer Inf*. 2021;5:379–93.
115. Ananthakrishnan AN, Cai T, Savova G, Cheng S-C, Chen P, Perez RG, et al. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Dis*. 2013;19(7):1411–20.
116. Tang Y, Yang J, San Ang P, Dorajoo SR, Foo B, Soh S, et al. Detecting adverse drug reactions in discharge summaries of electronic medical records using Readpeer. *Int J Med Informatics*. 2019;128:62–70.
117. Jagannatha A, Liu F, Liu W, Yu H. Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 1.0). *Drug Saf*. 2019;42(1):99–111.
118. Togra A, Pawar S. Role of automation, natural language processing, artificial intelligence, and machine learning in hospital settings to identify and prevent adverse drug reactions. *J Pharmacovigil Drug Res*. 2022;3(3):3–5.
119. Ujiiie S, Yada S, Wakamiya S, Aramaki E. Identification of adverse drug event-related Japanese articles: natural language processing analysis. *JMIR Med Inf*. 2020;8(11):e22661.
120. Nye B, Li JJ, Patel R, Yang Y, Marshall JJ, Nenkova A, et al. A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. In: *Proceedings of the conference association for computational linguistics meeting*, vol 2018. 2018. p. 197.
121. Bao Y, Deng Z, Wang Y, Kim H, Armengol VD, Acevedo F, et al. Using machine learning and natural language processing to review and classify the medical literature on cancer susceptibility genes. *JCO Clin Cancer Inf*. 2019;1:1–9.
122. Gabarron E, Larbi D, Dorronzoro E, Hasvold PE, Wynn R, Årsand E. Factors engaging users of diabetes social media channels on Facebook, Twitter, and Instagram: observational study. *J Med Internet Res*. 2020;22(9):e21204.
123. Bour C, Ahne A, Schmitz S, Perchoux C, Dessenne C, Fagherazzi G. The use of social media for health research purposes: scoping review. *J Med Internet Res*. 2021;23(5):e25736.
124. Dominy CL, Arvind V, Tang JE, Bellaire CP, Pasik SD, Kim JS, et al. Scoliosis surgery in social media: a natural language processing approach to analyzing the online patient perspective. *Spine Deform*. 2022;10(2):239–46. <https://doi.org/10.1007/s43390-021-00433-0>.
125. Tahami Monfared AA, Stern Y, Doogan S, Irizarry M, Zhang Q. Stakeholder insights in Alzheimer's disease: natural language processing of social media conversations. *J Alzheimers Dis*. 2022;89(2):695–708.
126. Watanabe T, Yada S, Aramaki E, Yajima H, Kizaki H, Hori S. Extracting multiple worries from breast cancer patient blogs using multilabel classification with the natural language processing model bidirectional encoder representations from transformers: infodemiology study of blogs. *JMIR Cancer*. 2022;8(2):e37840.
127. Harada S, Arend R, Dai Q, Levesque JA, Winokur TS, Guo R, et al. Implementation and utilization of the molecular tumor board to guide precision medicine. *Oncotarget*. 2017;8(34):57845.
128. Louie BH, Kato S, Kim KH, Lim HJ, Lee S, Okamura R, et al. Precision medicine-based therapies in advanced colorectal cancer. The University of California San Diego Molecular Tumor Board experience. *Mol Oncol*. 2022;16(13):2575–84. <https://doi.org/10.1002/1878-0261.13202>.
129. Peh KH, Przybylski DJ, Fallon MJ, Bergsbaken JJ, Hutson PR, Yu M, et al. Clinical utility of a regional precision medicine molecular tumor board and challenges to implementation. *J Oncol Pharm Pract*. 2022;10781552221091282. <https://doi.org/10.1177/10781552221091282>.
130. Charo LM, Eskander RN, Sicklick J, Kim KH, Lim HJ, Okamura R, et al. Real-world data from a molecular tumor board: improved outcomes in breast and gynecologic cancers patients with precision medicine. *JCO Precis Oncol*. 2022;6:e2000508.
131. VanderWalde A, Grothey A, Vaena D, Vidal G, ElNaggar A, Bufalino G, et al. Establishment of a molecular tumor board (MTB) and uptake of recommendations in a community setting. *J personalized Med*. 2020;10(4):252.
132. Larson KL, Huang B, Weiss HL, Hull P, Westgate PM, Miller RW, et al. Clinical outcomes of molecular tumor boards: a systematic review. *JCO Precis Oncol*. 2021;5:1122–32.
133. Mano H. Cancer genomic medicine in Japan. *Proc Jpn Acad Ser B*. 2020;96(7):316–21.
134. OncoKB. Precision oncology knowledge base. <https://www.oncokb.org/>.
135. Katsoulakis E, Duffy JE, Hintze B, Spector NL, Kelley MJ. Comparison of annotation services for next-generation sequencing in a large-scale precision oncology program. *JCO Precis Oncol*. 2020;4:212–21.
136. Muinos F, Martinez-Jimenez F, Pich O, Gonzalez-Perez A, Lopez-Bigas N. In silico saturation mutagenesis of cancer genes. *Nature*. 2021;596(7872):428–32.
137. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):1–14.
138. Motzer RJ, Banchereau R, Hamidi H, Powles T, McDermott D, Atkins MB, et al. Molecular subsets in renal cancer determine outcome to checkpoint and angiogenesis blockade. *Cancer Cell*. 2020;38(6):803–17 e4.
139. Motzer RJ, Powles T, Atkins MB, Escudier B, McDermott DF, Suarez C, et al. IMmotion151: a randomized phase III study of atezolizumab plus bevacizumab vs sunitinib in untreated metastatic renal cell carcinoma (mRCC). 2018.
140. Hamamoto R, Takasawa K, Machino H, Kobayashi K, Takahashi S, Bolatkan A, et al. Application of non-negative matrix factorization in oncology: one approach for establishing precision medicine. *Brief Bioinform*. 2022;23(4):bbac246.
141. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science*. 2013;339(6127):1546–58.
142. Dragomir I, Akbar A, Cassidy JW, Patel N, Clifford HW, Contino G. Identifying cancer drivers using DRIVE: a feature-based machine learning model for a pan-cancer assessment of somatic missense mutations. *Cancers (Basel)*. 2021;13(11):2779.
143. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. 2012;12(4):252–64.
144. Li X, Shao C, Shi Y, Han W. Lessons learned from the blockade of immune checkpoints in cancer immunotherapy. *J Hematol Oncol*. 2018;11(1):31.
145. Marin-Acevedo JA, Kimbrough EO, Lou Y. Next generation of immune checkpoint inhibitors and beyond. *J Hematol Oncol*. 2021;14(1):45.
146. Nabet BY, Esfahani MS, Moding EJ, Hamilton EG, Chabon JJ, Rizvi H, et al. Noninvasive early identification of therapeutic benefit from immune checkpoint inhibition. *Cell*. 2020;183(2):363–76 e13.

147. Chowell D, Yoo SK, Valero C, Pastore A, Krishna C, Lee M, et al. Improved prediction of immune checkpoint blockade efficacy across multiple cancer types. *Nat Biotechnol.* 2022;40(4):499–506.
148. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *J Big Data.* 2019;6(1):1–25.
149. Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity (Edinb).* 2020;124(4):525–34.
150. Morin O, Vallieres M, Braunstein S, Ginart JB, Upadhaya T, Woodruff HC, et al. An artificial intelligence framework integrating longitudinal electronic health records with real-world data enables continuous pan-cancer prognostication. *Nat Cancer.* 2021;2(7):709–22.
151. Peterson DJ, Ostberg NP, Blayney DW, Brooks JD, Hernandez-Boussard T. Machine learning applied to electronic health records: identification of chemotherapy patients at high risk for preventable emergency department visits and hospital admissions. *JCO Clin Cancer Inform.* 2021;5:1106–26.
152. Brooks GA, Li L, Uno H, Hassett MJ, Landon BE, Schrag D. Acute hospital care is the chief driver of regional spending variation in Medicare patients with advanced cancer. *Health Aff.* 2014;33(10):1793–800.
153. Yabroff KR, Lamont EB, Mariotto A, Warren JL, Topor M, Meekins A, et al. Cost of care for elderly cancer patients in the United States. *J Natl Cancer Inst.* 2008;100(9):630–41.
154. Wallace EM, Cooney MC, Walsh J, Conroy M, Twomey F. Why do palliative care patients present to the emergency department? Avoidable or unavoidable? *Am J Hospice Palliat Medicine®.* 2013;30(3):253–6.
155. Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end-of-life cancer care from administrative data. *J Clin Oncol.* 2003;21(6):1133–8.
156. Neugut AI, Bates SE. Emergency department visits for emesis following chemotherapy: guideline nonadherence, OP-35, and a path back to the future. *Oncologist.* 2021;26(4):274–6.
157. Csik VP, Li M, Binder AF, Handley NR. Development of an oncology acute care risk prediction model. *JCO Clin Cancer Inf.* 2021;5:266–71.
158. Yuan Q, Cai T, Hong C, Du M, Johnson BE, Lanuti M, et al. Performance of a machine learning algorithm using electronic health record data to identify and estimate survival in a longitudinal cohort of patients with lung cancer. *JAMA Netw Open.* 2021;4(7):e2114723.
159. Zhang Y, Cai T, Yu S, Cho K, Hong C, Sun J, et al. High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP). *Nat Protoc.* 2019;14(12):3426–44.
160. Pishvaian MJ, Blais EM, Bender RJ, Rao S, Boca SM, Chung V, et al. A virtual molecular tumor board to improve efficiency and scalability of delivering precision oncology to physicians and their patients. *JAMIA Open.* 2019;2(4):505–15.
161. Ortiz MV, Kobos R, Walsh M, Slotkin EK, Roberts S, Berger MF, et al. Integrating genomics into clinical pediatric oncology using the molecular tumor board at the Memorial Sloan Kettering Cancer Center. *Pediatr Blood Cancer.* 2016;63(8):1368–74.
162. Boddu S, Walko CM, Bienasz S, Bui MM, Henderson-Jackson E, Naghavi AO, et al. Clinical utility of genomic profiling in the treatment of advanced sarcomas: a single-center experience. *JCO Precis Oncol.* 2018;2:1–8.
163. Abdelkader W, Navarro T, Parrish R, Cotoi C, Germini F, Linkins LA, et al. A deep learning approach to refine the identification of high-quality clinical research articles from the biomedical literature: protocol for algorithm development and validation. *JMIR Res Protoc.* 2021;10(11):e29398.
164. Ebadi A, Xi P, Tremblay S, Spencer B, Pall R, Wong A. Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing. *Scientometrics.* 2021;126(1):725–39.
165. Gurulingappa H, Mateen-Rajpu A, Toldo L. Extraction of potential adverse drug events from medical case reports. *J Biomed Semant.* 2012;3(1):1–10.
166. Zeng J, Cruz-Pico CX, Saridogan T, Shufean MA, Kahle M, Yang D, et al. Natural language processing-assisted literature retrieval and analysis for combination therapy in cancer. *JCO Clin Cancer Inform.* 2022;6:e2100109.
167. Grand A, Muir R, Ferenczi J, Lin J. From MAXSCORE to block-max wand: the story of how Lucene significantly improved query evaluation performance. *ECIR 2020: Advances in information retrieval*, vol 12036. 2020. p. 20–7.
168. Chen HO, Lin PC, Liu CR, Wang CS, Chiang JH. Contextualizing genes by using text-mined co-occurrence features for cancer gene panel discovery. *Front Genet.* 2021;12:771435.
169. Luthra R, Patel KP, Routbort MJ, Broadus RR, Yau J, Simien C, et al. A targeted high-throughput next-generation sequencing panel for clinical screening of mutations, gene amplifications, and fusions in solid tumors. *J Mol Diagn.* 2017;19(2):255–64.
170. Paige SL, Saha P, Priest JR. Beyond gene panels: whole exome sequencing for diagnosis of congenital heart disease. *Circ Genom Precis Med.* 2018;11(3):e002097.
171. Patel NM, Michelini VV, Snell JM, Balu S, Hoyle AP, Parker JS, et al. Enhancing next-generation sequencing-guided cancer care through cognitive computing. *Oncologist.* 2018;23(2):179–85.
172. Itahashi K, Kondo S, Kubo T, Fujiwara Y, Kato M, Ichikawa H, et al. Evaluating clinical genome sequence analysis by Watson for genomics. *Front Med.* 2018;5:305.
173. Jie Z, Zhiying Z, Li L. A meta-analysis of Watson for Oncology in clinical application. *Sci Rep.* 2021;11(1):1–13.
174. Strickland E. IBM Watson, heal thyself: how IBM overpromised and underdelivered on AI health care. *IEEE Spectr.* 2019;56(4):24–31.
175. Madhavan S, Beckman RA, McCoy MD, Pishvaian MJ, Brody JR, Macklin P. Envisioning the future of precision oncology trials. *Nat Cancer.* 2021;2(1):9–11.
176. Zhang X, Yang H, Zhang R. Challenges and future of precision medicine strategies for breast cancer based on a database on drug reactions. *Biosci Rep.* 2019;39(9):BSR20190230. <https://doi.org/10.1042/BSR20190230>.
177. Prasad V. Perspective: the precision-oncology illusion. *Nature.* 2016;537(7619):63.
178. Meric-Bernstam F, Brusco L, Shaw K, Horombe C, Kopetz S, Davies MA, et al. Feasibility of large-scale genomic testing to facilitate enrollment onto genomically matched clinical trials. *J Clin Oncol.* 2015;33(25):2753–62.
179. Asada K, Kaneko S, Takasawa K, Machino H, Takahashi S, Shinkai N, et al. Integrated analysis of whole genome and epigenome data using machine learning technology: toward the establishment of precision oncology. *Front Oncol.* 2021;11:666937.
180. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med.* 2022;28(1):31–8.
181. Kim YH. Artificial intelligence in medical ultrasonography: driving on an unpaved road. *Ultrasonography.* 2021;40(3):313.
182. Diao G, Vidyashankar AN. Assessing genome-wide statistical significance for large p small n problems. *Genetics.* 2013;194(3):781–3.
183. Liang S, Huang W-H, Liang F. Sufficient dimension reduction with deep neural networks for phenotype prediction. In: *Proceedings of the 3rd international conference on statistics: theory and applications (ICSTA'21).* 2021. p. 134.
184. Ling AS, Hay EH, Aggrey SE, Rekaya R. Dissection of the impact of prioritized QTL-linked and-unlinked SNP markers on the accuracy of genomic selection1. *BMC Genomic Data.* 2021;22(1):1–14.
185. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: a review. *J Med Syst.* 2018;42(11):226.
186. Yu H, Yang LT, Zhang Q, Armstrong D, Deen MJ. Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. *Neurocomputing.* 2021;444:92–110.
187. Takahashi S, Asada K, Takasawa K, Shimoyama R, Sakai A, Bolatkan A, et al. Predicting deep learning based multi-omics parallel integration survival subtypes in lung cancer using reverse phase protein array data. *Biomolecules.* 2020;10(10):1460.
188. Ma T, Zhang A. Integrate multi-omics data with biological interaction networks using Multi-view Factorization AutoEncoder (MAE). *BMC Genom.* 2019;20(11):1–11.
189. Huang W-H, Wei Y-C. A split-and-merge deep learning approach for phenotype prediction. *Front Bioscience-Landmark.* 2022;27(3):78.
190. Gandouz M, Holzmann H, Heider D. Machine learning with asymmetric abstention for biomedical decision-making. *BMC Med Inf Decis Mak.* 2021;21(1):1–11.

191. Vokinger KN, Feuerriegel S, Kesselheim AS. Continual learning in medical devices: FDA's action plan and beyond. *Lancet Digit Health*. 2021;3(6):e337–e8.
192. Lee CS, Lee AY. Clinical applications of continual learning machine learning. *Lancet Digit Health*. 2020;2(6):e279–e81.
193. Rivera SC, Liu X, Chan A-W, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digital Health*. 2020;2(10):e549–60. [https://doi.org/10.1016/S2589-7500\(20\)30219-3](https://doi.org/10.1016/S2589-7500(20)30219-3).
194. The US, Food. and Drug Administration (FDA). Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). 2019.
195. Prabhakar B, Singh RK, Yadav KS. Artificial intelligence (AI) impacting diagnosis of glaucoma and understanding the regulatory aspects of AI-based software as medical device. *Comput Med Imaging Graph*. 2021;87:101818.
196. The US, Food. and Drug Administration (FDA). Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). 2019. <https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf>.
197. Stephens K. Fda releases artificial intelligence/machine learning action Plan. *AXIS Imaging News*; 2021.
198. Odaibo SG. Risk management of AI/ML software as a medical device (SaMD): on ISO 14971 and related standards and guidances. *arXiv preprint arXiv:210907905* 2021.
199. The US, Food. and Drug Administration (FDA). Good machine learning practice for medical device development: guiding principles. 2021. 2022. <https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles>.
200. Abramoff MD, Cunningham B, Patel B, Eydelman MB, Leng T, Sakamoto T, et al. Foundational considerations for artificial intelligence using ophthalmic images. *Ophthalmology*. 2022;129(2):e14–32.
201. Sakai K, Takeda M, Shimizu S, Takahama T, Yoshida T, Watanabe S, et al. A comparative study of curated contents by knowledge-based curation system in cancer clinical sequencing. *Sci Rep*. 2019;9(1):1–8.
202. Yang SJ, Pek A. From information overload to actionable insights: digital solutions for interpreting cancer variants from genomic testing. *J Mol Pathol*. 2021;2(4):312–8.
203. Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic Acids Res*. 2018;46(20):10546–62.
204. Lau-Min KS, Asher SB, Chen J, Domchek SM, Feldman M, Joffe S, et al. Real-world integration of genomic data into the electronic health record: the PennChart Genomics Initiative. *Genet Sci*. 2021;23(4):603–5.
205. Kahraman A, Arnold FM, Hanimann J, Nowak M, Pauli C, Britschgi C, et al. MTPilot: an interactive software for visualization of next-generation sequencing results in molecular tumor boards. *JCO Clin Cancer Inf*. 2022;6:e2200032.
206. Luchini C, Lawlor RT, Milella M, Scarpa A. Molecular tumor boards in clinical practice. *Trends Cancer*. 2020;6(9):738–44.
207. Iqbal MJ, Javed Z, Sadia H, Qureshi IA, Irshad A, Ahmed R, et al. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. *Cancer Cell Int*. 2021;21(1):1–11.
208. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Reviews Urol*. 2019;16(7):391–403.
209. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(Database issue):D980–5.
210. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*. 2004;91(2):355–8.
211. Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, et al. Variant interpretation using population databases: lessons from gnomAD. *Hum Mutat*. 2021;43(12):1012–30.
212. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. *JCO Precis Oncol*. 2017;2017:1–16.
213. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet*. 2017;49(2):170–4.
214. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database—update and key issues. *N Engl J Med*. 2011;364(9):852–60.
215. Cline MS, Liao RG, Parsons MT, Paten B, Alquaddoomi F, Antoniou A, et al. BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. *PLoS Genet*. 2018;14(12):e1007752.
216. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v. 2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32(5):557–63.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

