

Revolutionizing cancer care with machine learning: a comprehensive review

Leena Pavitha P.¹,
Sannasi Ganapathy²,
Allan Mark P.³, Nithisha Devi¹ and
Jerad Suresh A.^{1,*}

¹Department of Computer and Information Science, Sri Ramachandra Faculty of Pharmacy, Sri Ramachandra Institute of Higher Education and Research, Chennai, India

²Department of Computer Science and Engineering Education, National Institute of Technical Teachers' Training & Research (NITTTR), Bhopal, India

³Department of Pharmacy Practice, University of Massachusetts Dartmouth, Dartmouth, MA, USA

*E-mail: jeradsuresh1@gmail.com

Received for publication
February 07, 2025.

Abstract

In healthcare, the need for machine learning (ML) analysis in tailored cancer therapy is more pronounced than ever. The escalating volume and complexity of patient data and the growing understanding of diverse cancer subtypes demand sophisticated tools for precise decision-making. ML's ability to sift through immense datasets rapidly and identify nuanced patterns addresses this need, enabling oncologists to deliver personalized treatments considering a patient's unique genetic makeup and evolving health status. As healthcare systems strive for efficiency and cost-effectiveness, the integration of ML improves patient outcomes. It supports resource optimization, making it an indispensable asset in the contemporary fight against cancer. This review critically examines the integration of ML in healthcare, emphasizing its significance across diverse domains. It explores personalized medicine, where ML tailor's treatments to individual genetic profiles, and precision medicine, optimizing drug efficiency. The review focuses on the role of advanced algorithms in clinical decision support systems (CDSS) for enhancing healthcare decision-making. It addresses the various cancer-related issues in different applications, discussing ML's role in toxicity detection, predicting treatment responses, and ultimately contributing to more effective and tailored cancer therapies. This comprehensive exploration focuses on the pivotal impact of ML in shaping modern healthcare and improving patient treatment outcomes.

Keywords

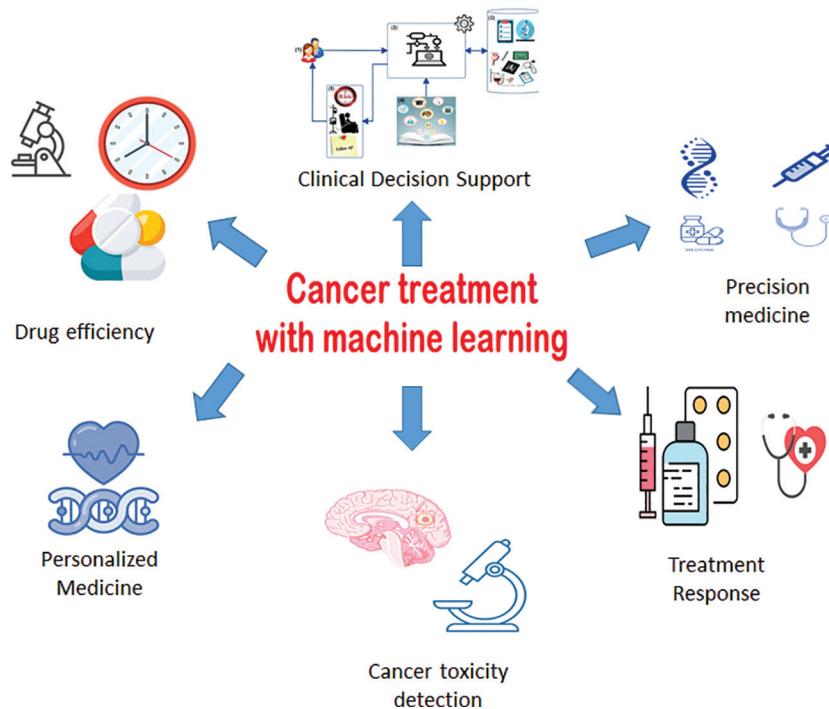
Machine learning, Personalized Medicine, Precision Medicine, Predicting drug efficiency, Clinical decision support systems, Cancer toxicity detection, Treatment Response

1. Introduction

The process of designing cancer treatment trials becomes progressively harder due to various challenges. The expected patient benefits from emerging therapies exist only for selected groups, while their tumor influencers act irregularly and they show unpredictable technology interactions with testing tools, including basic approaches to advanced methods. All analysis methods for treatment mechanism discovery require patient samples so that scientists

can improve clinical trial designs and reduce both trial sizes and treatment expenses [1,2]. Traditional therapeutic trials approve patient eligibility based on clinical characteristics combined with tumor type as well as disease stage along with treatment history [3,4]. The identification of patients who would benefit optimally from therapy becomes possible by adding molecular information from both tumors and patients. Studies analyzing tumor enzyme expression levels have established a connection between diagnostic findings and treatment responses, thus demonstrating

Graphical Abstract



that early diagnosis coupled with precise prognostic exams produce improved treatment results together with reduced medical complications and economical healthcare delivery [5, 6, 7].

Generally, machine learning (ML) is a subset of artificial intelligence (AI). ML applications in oncology became possible through the data explosion produced by professional cancer assessments and innate clinical records. Through examining extensive cancer databases and patient profile data, ML finds previously unknown connections among features of patients and their treatments and medical results. Supervised learning methods create these models through data training that accepts past data for result prediction purposes when new cases need analysis [7–10]. Cancer-related research applications of ML algorithms produce disease projections while improving treatment procedures and creating better patient healthcare by extracting vital predictive information [11]. Deep learning (DL) technology has dramatically impacted image and speech recognition along with automated translation and healthcare improvement through recent advances in algorithm development data storage capabilities and Internet speed and processing strength. AI-powered ML platforms help oncologists by examining radiological images and pathology slides to diagnose cancer more accurately

and forecast treatment outcomes to improve therapeutic decisions. The technology demonstrates promising potential because cancer affects large numbers of patients internationally [10].

Figure 1 shows the PRISMA flowchart that was used to identify the studies through various databases and registers. Generally, it has three major categories such as identification, screening, and inclusion. Here, it starts with the identification of various documents through various databases and registers; the end of the identification process and the duplicate documents are removed. For example, from a total of 250 documents, 100 duplicate documents were removed, which means that the remaining 150 documents are to be screened for further processes.

The screening process is done by checking the quality design, comparator, letter, non-availability of full text, and non-usage of English in documents. Next, it identifies 87 documents among the 150 documents to the included phase. Finally, 87 review articles are considered for inclusion and consideration.

The major contributions of the paper are listed as below:

- (i) Critically examined the integration of ML in healthcare, emphasizing its significance across diverse domains.

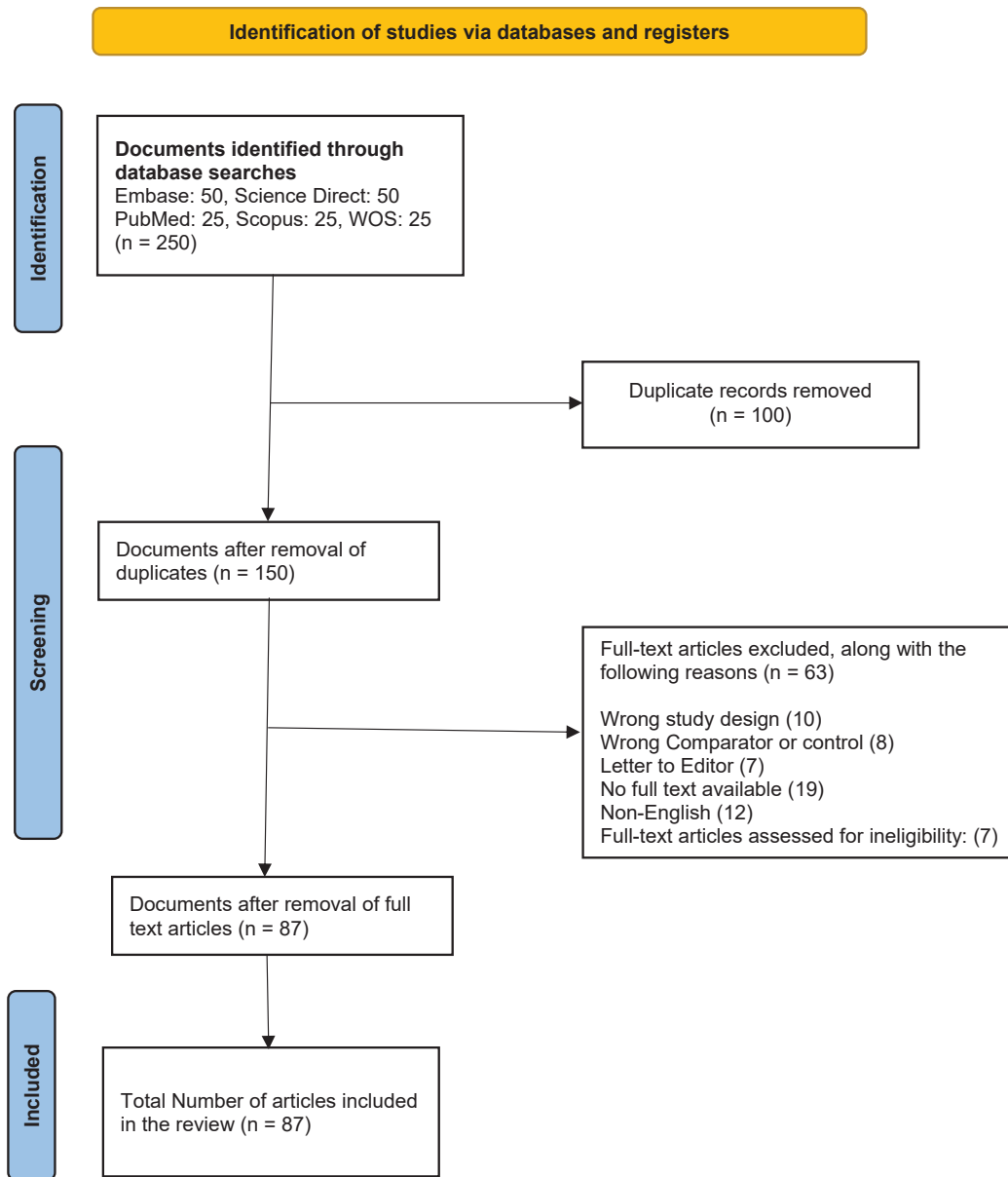


Figure 1: PRISMA flowchart. WOS, Web of Science.

- (ii) Explored the personalized medicine in which ML tailor's treatments to individual genetic profiles, precision medicine, and optimizing the efficiency of drugs.
- (iii) Focuses on the role of advanced algorithms in clinical decision support systems (CDSSs) for enhancing healthcare decision-making.
- (iv) Addresses the various cancer-related issues in different applications, discussing ML's role in toxicity detection, predicting treatment responses, and ultimately contributing to more effective and tailored cancer therapies.
- (v) Conducted comparative analysis based on the

performance evaluation metrics and suggested the suitable models.

The remainder of the paper is organized as follows: Section II discusses personalized medicine. Section III explains the basics of precision medicine. Section IV explains the process of predicting the drug efficiency. Section V explains CDSSs in detail with necessary examples. Section VI describes the role of ML in the cancer toxicity detection process. Section VII explains the role of ML in treatment responses. Section VIII concludes the review by highlighting the contributions and suggestions.

II. Personalized Medicine

The implementation of personalized medicine resembles the process of developing drugs and health devices, with each component requiring distinct expertise and offering opportunities for ML collaboration with AI tools and scientists. Effective information transmission and transitions between components is critical. ML plays a significant role in identifying therapeutic targets, particularly in contexts where vast data from assays need sophisticated statistical analysis to reveal meaningful patterns [12,13]. ML has been instrumental in analyzing molecular pathology imaging techniques and DNA sequence data to identify such patterns for potential therapies [14].

Furthermore, ML-based analyses greatly assist in genetic diagnosis by mining DNA sequence information and diagnosing diseases like cancer through AI-based analyses of blood analytes. Alongside the development of learning systems, numerous initiatives aim to aggregate patient data and materials for ML-based analysis, especially in cancer research. Doctor–patient relations will be profoundly impacted by the integration of ML-driven goods, such as decision assistance tools in Electronic Medical Records (EMRs) and educational systems [15]. Extensive national initiatives, supported by the government, are underway to analyze data using AI from individuals regarding healthcare-related trends, potentially benefiting future clinical and public health strategies and enhancing ML algorithm improvement [16]. Translating personalized medicine into practice shares numerous similarities with the development workflow of drugs and health products. A crucial similarity lies in the necessity for smooth and effective transitions between each workflow stage. Achieving this may be difficult as different technologies and skills are needed at different stages and are often sought separately. As a result, the ideal setting for customized medicine and healthcare entails smoothly combining each patient’s diagnosis, course of treatment, and continuous monitoring into a single process and coordinating changes across pertinent tasks or sub-tasks. The patient’s tumor can be profiled using immunotherapeutic approaches to cancer, which include cell replacement treatments to find particular “neo-antigens” or mutations that may cause the host’s immune system to attack cells having those mutations [17].

If these neo-antigens are found, the patient’s cells (autologous transplantation) or those of a donor are harvested and the patient’s cells are subsequently sensitized to identify the neo-antigens [18]. The use

of AI in ML in personalized medicine encounters numerous constraints. These include challenges in discerning patterns within an individual’s health data that may signify changes in their health status, especially as more data are amassed from various individuals [12].

Consequently, predictions regarding an individual’s health trajectory tend to rely increasingly on their historical data rather than on broader population data. Thus, it becomes imperative to assess the efficacy of healthcare products leveraging AI technologies [19,20]. This imperative stems from the variability in outcomes witnessed with certain healthcare solutions powered by AI or big data, as shown by the uneven performance of IBM’s Watson treatment decision support system (DSS) [21]. AI decision support solutions for ML use complex neural networks and DL methods. When trained on extensive datasets, these algorithms yield highly accurate predictions; thus, comprehending the intricate relationships between input data and resulting predictions can be challenging. The opacity of many AI tools poses a “Black Box” dilemma, potentially undermining confidence and generating apprehension depending on these forecasts when people’s lives are in jeopardy [22].

When assessing certain types of malignancy, fine-needle aspiration (FNA) is the main preoperative evaluation method [23,24]. The category-based Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) is a modern tool that helps physicians make decisions for thyroid cancer. Six diagnostic criteria that indicate the estimated risk of malignancy (ROM) are included in the TBSRTC [25]. According to recent studies, cytologists and ML models agree well when predicting malignancy, revealing notably lower ROM values for TBSRTC III when determined by ML models compared with manual classification [26]. Consequently, as therapeutic standards provide more flexibility in such patients, the probability of lobectomy or repeat FNA may be explored [27].

Furthermore, molecular testing is a precise and non-invasive way to reduce diagnostic and clinical uncertainty. Next-generation sequencing (NGS) produces billions of molecular fragments by enabling the quick study of many genes at once in a single operation [28]. With its large datasets, quick sequencing speed, and precise outputs, NGS has become a paramount component of big data analytics [29]. Such large and complex datasets are excessive for traditional information systems to manage. In this setting, AI manifests as a powerful big data algorithm combining multi-omic data for various learning

tasks, allowing high-level traits to be automatically ML detected or classified [30].

Communication between elements of the immune-inflammatory responses, such as macrophages, neutrophils, and lymphocytes and the cellular makeup of the tumor, such as epithelial cancer cells, fibroblasts, and endothelial cells, is facilitated by the complex web of molecular patterns within the tumor microenvironment, including cytokines, chemokines, and adipocytokines [31]. Combining morphological analysis, molecular markers, AI, and ML will provide more thorough insights into cancer treatment at the individual patient level [32]. In spite of these developments, customized medicine faces several obstacles, such as uneven rating performance, imprecise cytopathological diagnosis, challenges in differentiating between follicular lesions, and imprecise prognostication [33]. AI and ML have enhanced the efficiency and accuracy of diagnosing and treating various types of tumors, underscoring their potential in the medical field. AI technology enables the extraction and analysis of a burgeoning volume of medical data, promising vast opportunities for enhanced medical understanding and decision-making for ML. The initial success of AI with ML in image recognition has led to the emergence of radiogenomics. Radiogenomics, a pioneering field in personalized medicine research, aims to link cancer imaging traits with gene expression to forecast the likelihood of patients experiencing toxicity after radiation treatment. Radiogenomic links between several malignancies, including liver cancer [34], colorectal cancer, and breast cancer, have been found in large part due to AI. However, a major obstacle to AI in radiogenomics is the lack of data.

III. Precision Medicine

Precision medicine in cancer utilizes specific tumor information to aid in diagnosis, treatment planning, assessing treatment efficacy, and forecasting prognosis. Techniques in precision medicine discern patient phenotypes exhibiting atypical treatment responses or distinct healthcare requirements. AI harnesses advanced computational capabilities to derive insights, facilitating reasoning, learning, and enhancing clinical decision-making through augmented intelligence [35]. Progress in precision medicine yields tangible advantages like early disease detection and the increasing prevalence of personalized treatment design via ML in healthcare. Various data collecting and analytics tools enable precision medicine to customize treatment. Precision medicine

involves customizing ML treatment for individual patients, aiming to categorize them into groups with varying disease prognoses or treatment responses. This approach enables targeted interventions for those who would benefit while avoiding unnecessary expenses and side effects for others. DL algorithms are essential for automatically collecting characteristics from medical data when building ML models that can precisely predict tumor recurrence risk and patient reactions to therapies [36].

These forecasts enable doctors or healthcare personnel to provide more accurate and suitable care. Biomarkers of the immunogenic tumor microenvironment, such as Programmed Death-Ligand 1 (PD-L1) expression, tumor mutation burden (TMB), Microsatellite Instability (MSI), and somatic copy number changes, are now used to predict ML responses to immunotherapies. While informative, the biomarker data obtained through biopsy presents challenges due to its invasive nature, longitudinal difficulty, and confinement to a singular tumor region [37,38]. In pursuit of precision medicine objectives for ML, using data from radiomics and pathomics, many researchers have created deep ML models to predict patient biomarkers related to immunotherapy [39]. In addition, ML has made it possible to dynamically modify medication doses for individual patients, either for single or combination treatments, based on continuous gathering of patient-specific data, in addition to its ability to predict patient therapeutic responses [40].

Important variables affecting tumor aggressiveness and influencing clinical decision-making and outcomes include the time of cancer discovery, the precision of cancer diagnosis, and staging. This important area of oncology can perform on par with human specialists while providing advantages in terms of automation and scalability. DL algorithms that correctly diagnose cancer and identify cancer subtypes from histopathologic and other medical pictures have been extensively described [41]. When given enough processing capacity, deep neural networks (DNNs) of ML are powerful algorithms that can analyze huge pictures, such as W/S images of tissue from surgical resections or biopsies stained with hematoxylin and eosin. Another crucial step in the diagnostic procedure is grading and staging the malignancy, which establishes the disease's aggressiveness and progression. The choice between aggressive treatment combining radiation, surgery, and chemotherapy and cautious wait is significantly influenced by staging. The "Gleason score" is used to stage prostate cancer and entails determining the frequency of tumor cells

in two distinct places on a slide [42]. Non-imaging data such as genetic profiles are included in staging and diagnostic procedures in addition to conventional techniques like the Gleason score.

Since the early 2000s, ML has been used to diagnose and stage cancer using molecular data. During that time, microarray-based expression profiles were used to diagnose cancer and identify subtypes using ML approaches, including clustering, support vector machines (SVMs), and artificial neural networks (ANNs). However, it may be difficult to find genetic variations and mutations in NGS data, even with the availability of several computational techniques. This is possible in situations with less coverage or complicated, repeat-rich genome sections. Researchers have examined reframing mutation detection as an ML problem to overcome these challenges [43].

Furthermore, AI has shown promise in directly detecting some mutations from histopathology pictures, especially those clinically significant as response indicators for targeted treatments contributing to the development of ML [44]. “Deep Path” detects six important mutations linked to lung cancer in addition to classifying The Cancer Genome Atlas (TCGA) lung cancer subtypes [45]. Determining the origin of tumor cells can guide targeted therapies tailored to specific sites for ML and have demonstrated higher efficacy than broad systemic chemotherapies [46]. While these therapies may not fully replace molecular pathology assessments, they can offer insights into cellular mechanisms linked to mutations and facilitate screening numerous patients and tumors for subtypes with distinct mutational profiles for ML algorithms [47].

ML with AI is increasingly utilized to comprehend the functional consequences of mutations, such as predicting the effects of non-coding mutations on gene expression and disease susceptibility [48]. ML algorithms can differentiate between various diseases, but they often need more interpretable explanations for their classification decisions, unlike trained pathologists who rely on established image features and extensive experience to evaluate tissue samples [49]. AI holds promise in automating and streamlining routine tasks, potentially reducing the time burden on pathologists. Furthermore, ML has been leveraged to create medication molecules with the intended effects and target specificity *in silico*. However, traditional AI approaches primarily focus on binary classification and need help with complex objectives, such as generating novel molecules computationally [50].

Big data technology encompasses data analysis, mining, and sharing, potentially revolutionizing cancer diagnosis, treatment, prevention, and prognosis [51,52]. However, data translation into actionable patient benefit information needs to grow. This technology can discern visible and hidden features in medical images, extracting and refining these features to ascertain pertinent information regarding diagnosis, treatment, and prognosis [53]. Moreover, it can scrutinize patients’ demographic and clinical data alongside outcome information to anticipate factors impacting cancer patient prognosis [54]. Despite the underutilization of big data reanalysis thus far, its potential cannot be disregarded. By exploring into existing databases, it can yield novel insights. Integration of omics and non-omics data presents a pathway to surmount the challenges in cancer diagnosis, treatment, and monitoring [55].

ML is crucial for examining complex and diverse high-dimensional datasets, particularly in multi-omics and intergroup methods, aiding data integration to unveil cancer molecular mechanisms [56]. This aids in the precision identification of novel dynamic diagnostic and prognostic biomarkers, enhancing precision in cancer care [57]. Genomics, which analyzes data based on nucleotide sequences, has increasingly intertwined with clinical applications [58]. Spatial and single-cell genomics can reconstruct tumorigenesis processes, better comprehend tumors, clarify human pathogenesis, and facilitate targeted drug development [59]. The fusion of ML and genomics enables the identification of cancer subtypes, novel markers, and drug targets, enhancing the understanding of cancer-driving genes and facilitating personalized patient treatment [60]. Transcriptomics bridges genomics and proteomics, employing techniques like quantitative reverse transcription-polymerase chain reaction, RNA sequencing, and microarrays (NGS). ML-driven transcriptomics aids in acute myeloid leukemia diagnosis and calculates triple-negative breast cancer subtypes, surmounting heterogeneity challenges in disease treatment [61]. Figure 2 shows the role of ML in precision medicine.

The various factors influenced through ML in precision medicine include omics data, high-dimensional data, biomarkers, tumor microenvironment, radiomics and pathomics, mutation detection, big data, and molecular pathology; in addition to this are non-imaging data, medication dose modification, clinical decision-making, Gleason score, somatic copy number changes, microsatellite, and tumor mutational burden.

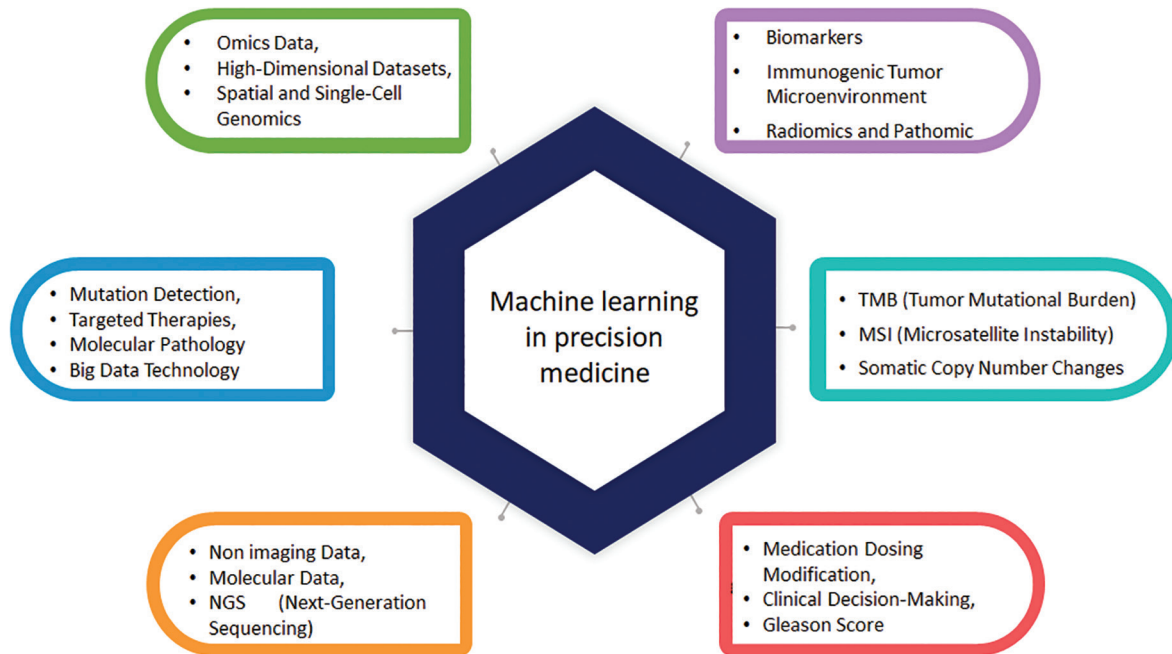


Figure 2: Factors influencing ML in precision medicine. ML, machine learning; NGS, next-generation sequencing; TMB, tumor mutation burden.

IV. Predicting Drug Efficiency

ML techniques have been utilized to forecast drug effectiveness by analyzing molecular characteristics. This research has become significant due to the accessibility of extensive datasets on cancer drug effectiveness, derived from experiments conducted on cell lines. Cell lines are not ideal models because of genetic variances or contamination, but they provide AI models with an enormous of data to work with. As is typical with any dataset, pretreatment procedures, such as confirming cell lines' authenticity or validating in vivo data, are often required to minimize any influence. Quinlan et al. assessed 1,001 cancer cell line responses to 265 different anticancer drugs in research. Based on these discoveries, they created several elastic net models to analyze genomic features, such as gene expression levels and mutations and the efficacy of medications [62].

Compared with other ML techniques, the "Random forests" algorithm has been shown to have higher overall accuracy when used as a medication response prediction tool [63]. Drug response prediction is increasingly done using DL in addition to conventional ML. Almomani et al. trained three DNNs using data from TCGA and the Cancer Cell Line Encyclopedia: one DNN was trained to encode mutation information, another to encode expression information, and a third DNN integrated the results of the

first two DNNs to predict drug response. However, DL techniques have certain limitations such as difficulty in understanding the underlying biological systems that influence the predictions. To guarantee that the neural network's hierarchical structure mirrored the established biological processes, an interpretable DL model using a "visible neural network" (VNN) was created to alienate this problem [64]. Furthermore, crowd-sourcing several models targeted at forecasting medication synergy in certain cancer cell types was a feature of the most recent challenge.

V. CDSSs

CDSS is an information system aiding various sectors like business, medicine, education, or organizational decision-making processes. These systems integrate human expertise with technology to enhance decision-making efficiency and influence ML. CDSSs are extensively applied in numerous medical domains to enhance healthcare service quality, optimize decision timelines, and enhance patients' health-related quality of life [65]. They contribute to cost reduction and decrease medical error rates. CDSSs hold the potential to enhance patient safety via ML by minimizing adverse events, suggesting personalized prescriptions, and facilitating efficient test and medication orders. Moreover, they can enhance healthcare delivery efficiency, curbing expenses through reduced

Table 1: Current and best methods used in cancer detection using ML

Method	Description	Advantages	Limitations	Reference
ML and DL algorithms	Used for analyzing large datasets, including imaging, genetic, and histopathological data. Helps in predicting tumor recurrence risk and patient responses to therapies.	High accuracy in pattern recognition, automation, and scalability. Can analyze massive datasets efficiently.	Requires large, high-quality datasets. Limited interpretability ("Black Box" problem).	[22, 36, 41]
NGS	Enables rapid sequencing of multiple genes, producing vast molecular data for cancer detection.	High throughput, precise genetic mutation detection, and potential for early diagnosis.	High costs, data complexity, and need for specialized expertise.	[28, 29, 43]
Radiogenomics	Combines imaging data with genomic data to identify cancer subtypes and predict treatment outcomes.	Non-invasive, links imaging traits with gene expression, and predicts radiation therapy response.	Limited by the availability of high-quality datasets.	[34]
Molecular testing	Uses genetic markers such as PD-L1, TMB, MSI, and somatic copy number variations to predict therapy response.	High precision, aids in personalized medicine, and helps in selecting targeted therapies.	Invasive biopsy requirement, limited by tumor region sampling.	[37, 38]
Histopathology with AI	AI-powered analysis of stained tissue slides to detect cancer and classify subtypes.	High accuracy, automation, and reduced workload for pathologists.	Requires large labeled datasets for training.	[41, 44, 45]
FNA) and cytopathology (TBSRTC)	FNA is a minimally invasive preoperative evaluation, and TBSRTC categorizes thyroid cancer risk.	Quick, minimally invasive, and widely used in clinical settings.	Subjective interpretation can lead to inconsistent results.	[23–26]
Transcriptomics (RNA sequencing, microarrays)	ML-driven analysis of gene expression to classify cancer subtypes and improve diagnosis.	High precision in detecting gene expression differences.	Data complexity and requires advanced computational tools.	[61]
Big data and AI integration	AI extracts insights from demographic, clinical, and imaging data to predict prognosis.	Improves decision-making, enhances early detection, and enables better patient stratification.	Underutilized potential and challenges in data integration.	[51–54]
ML-based biomarker discovery	ML helps identify predictive biomarkers, such as PD-L1 expression and TMB, to guide immunotherapy decisions.	Identified the predictive biomarkers	Not considered the suitability level.	[39]
Molecular and omics-based analysis	AI integrates genomic, transcriptomic, and proteomic data to improve cancer classification and prognosis.	Performed the effective classification and enriched the cancer diagnosis accuracy.	Not highlighted the false alarm rate and error rate analysis.	[56, 58]

AI, artificial intelligence; DL, deep learning; FNA, fine-needle aspiration; ML, Machine learning; NGS, next-generation sequencing; TBSRTC, The category-based Bethesda System for Reporting Thyroid Cytopathology; TMB, tumor mutation burden.

test duplication, decreased adverse events, and preference for cost-effective generic alternatives. Despite being accessible for clinical oncology for more than 10 years, awareness regarding CDSS varieties and their application in breast cancer treatment remains limited. Table 1 criticisms have emerged regarding the suitability of CDSS in personalized cancer treatment decisions, with perceptions that these systems are better suited to cater to the needs of the “average” patient rather than offering personalized treatment for individual patients. Furthermore, aside from the intricate nature of managing cancer clinically, there is a growing burden of the disease, largely driven by an aging demographic [66].

This escalation in cancer cases and associated costs significantly affects clinical requirements and the overall delivery of care. Health information technology (HIT) is often highlighted as a key facilitator in addressing these challenges. HIT is crucial in leveraging vast amounts of data, commonly known as “big data,” enabling the healthcare system to learn from each patient’s experience, enhance predictive capabilities for treatment options, and, ultimately, enhance the quality of care provided [67]. With the exponential growth in data volumes accessible through electronic health records (EHR) and the technology available to analyze and interpret these data into actionable predictive models, the concept of a rapidly learning health system for cancer care is more achievable than ever. CDSS represents one HIT tool that utilizes patient-specific data through predetermined algorithms to offer clinicians data-driven recommendations, thereby supporting decision-making at the point of care via ML. Electronic and automated CDSS were later devised to address these obstacles. Successful integration and interoperability with EHR significantly enhance the efficacy of CDSS implementation. Nevertheless, the advantages of CDSS in cancer management settings, particularly in offering decision support for disease-specific treatment or supportive care, remain unclear. CDSS holds the potential to facilitate appropriate, evidence-based care, thereby potentially enhancing patient outcomes in cancer treatment Table 1. The presence and application of effective decision-making support play a crucial role in guaranteeing top-notch cancer care results for favorable patient experiences and outcomes [68].

Shared decision-making (SDM) via ML involves the active participation of healthcare professionals and patients in a well-informed decision-making process. It is essential for both parties to base their decisions on the most reliable evidence available

garnered using ML algorithms while also taking into account different possibilities based on the values of the patients. Interactive technologies, which define DSS, enable and encourage informed SDM between patients and practitioners. Through patient-specific data and assessments, DSS helps healthcare practitioners make tailored choices that they usually carry out. These choices affect cancer care’s diagnosis, therapy, monitoring, and follow-up phases. Through the integration of patient data, pertinent health information, and specialized clinical knowledge into medical decision-making processes, CDSS seeks to improve healthcare quality. A CDSS typically comprises software designed to support physicians throughout their decision-making processes actively. It compares each patient’s unique features with an electronic clinical knowledge base, producing patient-specific assessments or recommendations for the physician to consider. At present, CDSS is mostly used via ML at the point of care, allowing physicians to combine their knowledge with the data or suggestions provided by the ML system.

With the advent of CDSS, information and observations that are unavailable or difficult for people to understand may be used. The data source is still necessary for CDSSs that do not rely on known information. Instead of being exclusively designed to follow expert medical knowledge, the decision-making process uses AI, ML, or statistical pattern recognition. Although gaining acceptance in the medical industry, non-knowledge-based CDSSs still face several obstacles, including the inability to understand AI’s reasoning to make recommendations (often known as “black boxes”) and problems with data availability.

VI. ML in Cancer Toxicity Detection

Determining toxicity is a difficult procedure as in vivo systems are intricate. As a result, throughout the pre-clinical or clinical stages, medication safety is among the main reasons for drug withdrawals [69]. To predict drug toxicity and promiscuity computationally, quantitative structure-activity relationship (QSAR) models were traditionally used [70] Table 2. These models establish a relationship between the functional groups of each molecule and their biological features. While models provide a clear mechanical description of the predictions, generating them from different and random information is challenging. ML algorithms for predicting the toxicity of small compounds have gained popularity in recent years because of the establishment of toxicity databases [71]. These approaches rely on their predictions on a model and

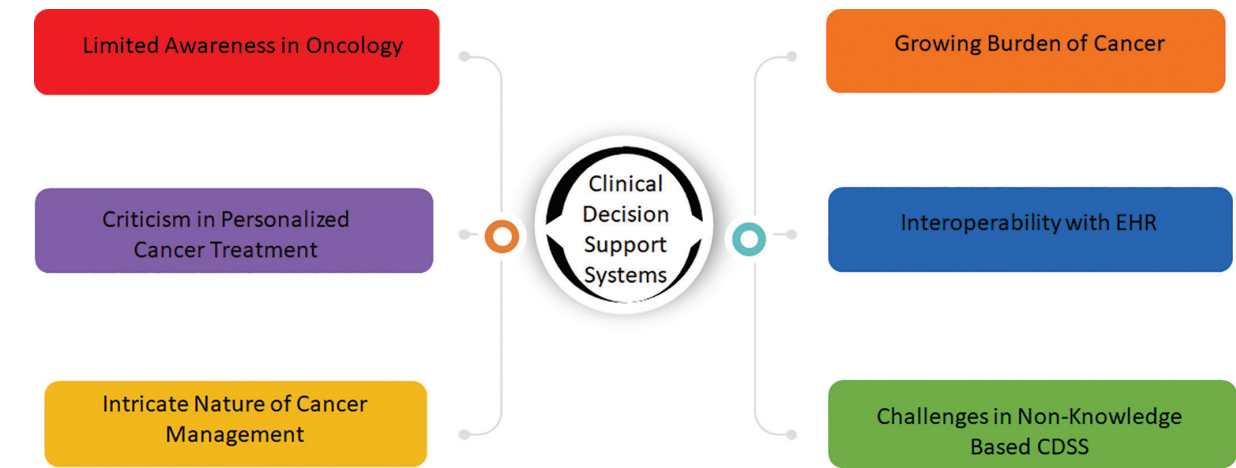


Figure 3: Factors influencing CDSS. CDSS, clinical decision support system; EHR, electronic health records.

Table 2: Current challenges in cancer detection using ML

Aspect	Challenges	Description	Reference
Cancer detection	Lack of high-quality training data	AI models require large, well-annotated datasets, which are often limited for certain cancer subtypes.	[34]
	Model interpretability (“Black Box” problem)	Complex DL models lack transparency, making clinical validation and trust difficult.	[22]
	Heterogeneity in cancer subtypes	Genetic and molecular variations between patients make it difficult to generalize AI predictions.	[57]
	Ethical and privacy concerns	Handling patient genomic data poses ethical and legal challenges regarding data security and bias.	[20]

AI, artificial intelligence; DL, deep learning; ML, machine learning.

use a statistical methodology. A crucial step in computational toxicity prediction involves choosing the appropriate molecular representation, which depends on the specific problem [72]. Various methods can depict molecular structures, such as labeled graphs Mobile Genetic Elements (MGE), numerical characteristics derived from physicochemical properties, or molecular descriptors [73] Table 3. There are a variety of molecular representations and ML algorithms for toxicity prediction. These include the 12 Tox21 Data Challenge endpoints and models for acute oral toxicity (AOT), hepatotoxicity, cardiotoxicity, and endocrine disruption [74]. Classification models are generally used when toxicity is proposed as an active/inactive inquiry based on the endpoint data. The performance of ML approaches varies depending on the dataset because of variations in their complexity, class distributions, or chemical space coverage [75]. This

diversity makes it difficult to evaluate algorithm performance across many hazardous endpoints. Different assessment measures, mostly based on the ML technique and the database used, further complicate the comparison. Finding mechanistic insights or explanations for reported hazardous effects is one of the main challenges facing computational toxicology today. Although ML techniques, especially DL models, are often considered as “black boxes” that can handle complicated problems effectively, their predictions usually lack dependability and reproducibility. Recent models are attempting to incorporate various tools to enhance interpretability.

Radiation-induced toxicity is a significant concern in cancer treatments, particularly regarding the potential for radiomics and ML to predict and assess these toxicities [76]. This approach can aid in tailoring radiation treatment for individual patients. While

radiotherapy is crucial for personalized treatment, high rates of severe acute and late toxicity are often linked. The past 20 years have seen a trend toward quantitative imaging, which might be crucial for capturing tumor heterogeneity and gene expression profiling [77]. Technological developments in medical image processing have made it possible to extract quantitative characteristics on a never-before-seen scale. In radionics, digitized pictures are transformed into high-dimensional data, mined to aid in clinical decision-making via ML [78]. Over the past 5 years, several studies have shown that radiomics and ML are used for accurate prediction and provide prognostic information for different forms of cancer [79]. However, regarding their capacity to forecast results and evaluate treatment-related toxicity, both disciplines are still in their infancy Table 4.

Patients with head and neck cancer have a considerable reduction in quality of life due to radiation-induced xerostomia. To anticipate this side effect, the construction of Normal Tissue Complication Probability (NTCP) models was designed, which mainly depended on baseline patient-reported

xerostomia levels and dose–volume characteristics [80]. Predicting disease response and survival involves using radiomic parameters, such as form, intensity, and textural characteristics, which are extracted from medical pictures. Acute and late symptoms of xerostomia are intimately linked to structural alterations in the parotid glands, namely, volume shrinking following radiation treatment. To use tailored replanning techniques, such as adaptive radiotherapy, which try to protect healthy parotid tissue from high-dose areas, it is essential to predict the volume loss. A common indicator of AOT is the LD_{50} . This metric represents the fatal dosage of a chemical that results in a 50% death rate in test animals following oral intake. Regression and classification methods are two types of ML models that have surfaced recently for AOT prediction. In MGE-based DL models, basic chemical information from molecular graphs is used as input, allowing DL algorithms to learn certain representations independently through graph convolutions. This removes the need for manual computation of numerical descriptors or fingerprints. The Toxicity Database Estimation Software

Table 3: Current trends in cancer treatment using ML

Aspect	Current trends	Description	Reference
Cancer treatment	AI-driven personalized therapy	AI models predict individual patient responses to treatments, improving precision medicine.	[35, 40]
	ML in drug discovery	AI accelerates drug development by predicting the efficacy and toxicity of new compounds.	[50]
	CDSS	AI-powered tools assist oncologists in optimizing treatment strategies based on patient-specific data.	[65]
	Dynamic dose adjustment with ML	ML models personalize drug dosages based on continuous monitoring of patient responses.	[40]

AI, artificial intelligence; CDSS, clinical decision support systems; ML, machine learning.

Table 4: Current challenges in cancer treatment using ML

Aspect	Challenges	Description	Reference
Cancer treatment	Variability in AI performance	AI-based DSSs have shown inconsistent results in real-world applications, such as IBM Watson.	[21]
	Computational and resource constraints	Training advanced ML models requires high-performance computing and large datasets.	[50]
	Resistance to AI Adoption in oncology	Many oncologists remain skeptical about fully integrating AI-driven decision-making into clinical practice.	[66]

AI, artificial intelligence; DSSs, decision support systems; IBM, International Business Machines Corporation; ML, machine learning.

Tool (TEST) was the last ML test set used for comparison with earlier research [81]. The Estimated Toxicity Scores (ESTDs), which achieved a value of 0.788, showed results comparable to state-of-the-art R2 predictions for quantitative toxicity datasets using physicochemical descriptors and the Multi-Task Deep Neural Network (MT-DNN) architecture.

Drug-induced liver damage (DILI) is a major cause of clinical trial failures and drug withdrawals from the market, as it stands out as a very critical toxicological endpoint [82]. More than 700 medications have been connected to hepatotoxicity in recent years. ML models that can cross-validate various datasets, including recognized pharmaceuticals, potentially dangerous substances, natural products, and synthesized bioactive compounds, that are used by the eToxPred algorithm. [83] A large toxicity dataset assembled from Food and Drug Administration (FDA)-approved medications and the Kyoto Encyclopedia of Genes and Genomes was used to train and evaluate these k-NN, k-Nearest Neighbors; LDA, Linear Discriminant Analysis; models. An important field of research is mutagenicity, or the capacity of chemicals to cause harm to DNA. A medication candidate may be rejected if an in vitro genetic toxicology screening

yields even one positive result. Carcinogenicity directly correlates with the degree of DNA damage, including strand breaks, chromosomal abnormalities, mutations, and DNA adducts. The Ames test is an ML assay for bacterial gene mutations that mimics the metabolism of mammals and is very sensitive to finding substances that can potentially cause genetic harm. However, like other toxicity assessments, identifying mutagenicity through experimentation is resource-intensive, time-consuming, and needs perfect reproducibility, making it efficient for meaningful results. Creating predictive models for toxicity necessitates access to dependable extensive datasets for evaluating different chemical substances. The US Tox21 algorithm initiative has crafted numerous in vitro assays employing quantitative high-throughput screening to produce a vast array of toxicity data for thousands of chemical compounds. Leveraging the extensive Tox21 dataset, multiple machine-learning models have been devised to forecast toxicity within the framework of the Tox21 challenge [84].

Table 5 shows a comparative analysis between the various techniques or models available for cancer toxicity detection by considering the description, advantages, and limitations.

Table 5: Cancer toxicity detection methods

Reference	Description	Advantages	Limitations
[70]	Silico toxicology models present themselves as an alternative to computationally analyze, simulate, and predict the toxicity of novel drugs/compounds.	Obtained the cost-efficiency and Speedy of data.	Not considered the performance-based analysis.
[71]	This analysis presented a new model to simulate complex chemical toxicology data sets and compared with the ML methods such as ANN, LDA, SVM, k-NN and NB.	Comparative analysis between the ML models over the drug toxicity.	Not considered the standard evaluation metrics.
[74]	Discussed the advances, challenges, and future perspectives of the drug toxicity prediction process using AI.	The various challenges are identified in the process of toxicity prediction.	Future perspective analysis can be enhanced further by considering the latest technology.
[81]	This review article discussed in detail about the necessary databases and software that are working based on robust computational assessments and toxicity prediction.	This review provides the analysis of existing databases and software methods.	Not done the robust computational assessments and robust toxicity prediction processes.
[82]	Proposed the question about whether clinical trials could be modified/improved for providing the data with better solutions for outstanding issues.	The various issues are discussed by providing the solutions.	The advanced technologies are not considered in the solutions.

AI, artificial Intelligence; ANN, artificial neural network; k-NN, k-Nearest Neighbors; LDA, Linear Discriminant Analysis; ML, machine learning; NB, Naïve Bayes; SVM, support vector machine.

Table 5 shows the performance of the various cancer toxicity detection methods and highlights the better and worst detection method with respect to the detection accuracy. Generally, we will choose the best model that obtains best detection accuracy.

VII. ML in Treatment Response

Determining the relationships between predictor and responder variables is the aim of predictive modeling for therapeutic response. Based on predictor factors, these models anticipate a patient's response to therapy. Predictor factors include a range of clinical and molecular patient characteristics, including age, gender, ethnicity, demography, severity of the first illness, and history of therapy. Measures such as time to treatment failure (e.g., detection of minimal residual disease, changes in blood markers, tumor recurrence, metastasis, cancer-related mortality) or evaluating treatment efficacy within a specific time-frame are response variables that reflect aspects of disease progression in response to treatment. High-throughput technology developments have significantly expanded the clinical and genetic data for researching cancer therapy responses. However, the interpretation and comparability of different *in vitro* and *in vivo* models with human data remains a continuous problem, especially in their capacity to capture the genomic and transcriptome features of patients' primary malignancies. The main goal of using ML in therapeutic predictive modeling is to identify the characteristics associated with therapy response. This knowledge is then used to predict future therapy responses for new patients using supervised learning approaches. Although unsupervised learning techniques have been widely used in cancer research, their main function is to find correlations between input variables; these correlations do not directly predict outputs. In this case, supervised learning techniques include regression-based ML models such as random forests and SVMs [85].

The procedure entails comprehending the relationships between input and output variables and utilizing outputs as a reference point. This allows the final ML model to forecast additional inputs, like new patients, in the future. Three key steps are involved in supervised learning tailored for predicting treatment response in clinical settings: training, parameter estimation from training data using condition list or frequentist methods, and binary classification tasks utilizing SVMs. The flexibility with which SVMs can accommodate misclassified data and non-linear class borders and their ability to handle multi-class settings make them very popular.

ANNs are widely utilized for pattern recognition and image processing tasks. ML is becoming a critical tool for therapy response prediction as more and more cancer patient data, both clinical and molecular, are accessible for computer analysis. Big Data's vastness makes investigating the complex processes influencing treatment response easier. However, the sheer number of individual determinants might overload the system when used as features in ML, resulting in overfitting. Several strategies—including wrapper methods—have been proposed and used for feature selection. High-throughput approaches are very beneficial for understanding cancer growth and treatment response; nevertheless, they are not impervious to experimental noise or missing results. Although certain ML methods, such as random forests, can withstand noise and incomplete data, others may experience a decline in model performance under such conditions. Although these advancements have greatly enhanced the examination of treatment responses, numerous obstacles persist in therapeutic monitoring that warrant comprehensive attention. The accessibility of molecular data in the public domain presents notable hurdles, particularly when swift predictions or result reproductions are necessary. With therapeutic monitoring increasingly accessible and molecular datasets expanding, we anticipate a greater reliance on advanced ML techniques. However, these techniques require sufficient samples to achieve optimal performance.

Figure 4 is helpful for understanding the critical factors of influencing ML in the process of responding to cancer disease treatments. Here, the high-throughput aware factors dominate in the treatment and are also useful for developing the various supervised learning techniques as models. This review article discusses the role of ML in the process of decision-making in the field of medicine. Comparative analyses are also done by considering the available works and suggest the suitable best method to detect the cancer toxicity and precision medicine.

VIII. Conclusion and Future Work

In conclusion, integrating ML into tailored cancer therapy is paramount. As patient data grows in volume and complexity, sophisticated tools are essential for precise decision-making. ML rapidly navigates vast datasets, identifying nuanced patterns and enabling personalized treatments based on genetic makeup and health status. This integration improves patient outcomes and optimizes resource allocation, enhancing efficiency and cost-effectiveness. Critically

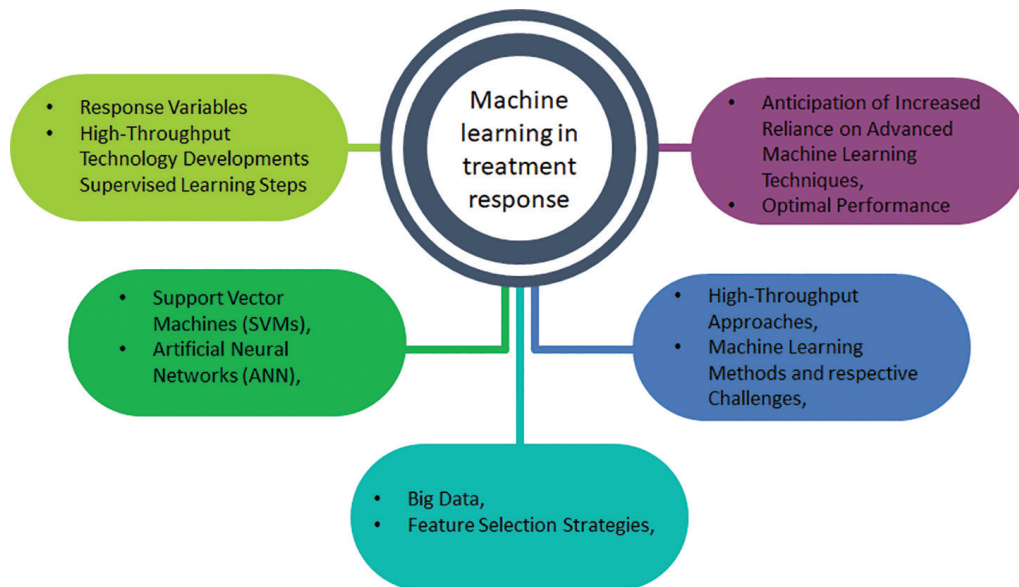


Figure 4: Critical factors influencing ML in treatment response. ANN, artificial neural network; ML, machine learning; SVMs, support vector machines.

examining ML's role across domains like personalized and precision medicine improves its transformative impact. By enhancing CDSS and facilitating tailored therapies, ML reshapes contemporary healthcare, continually advancing cancer care and patient treatment outcomes. This review has done a critical investigation on the role of ML in various emerging applications including healthcare, emphasizing its significance across diverse domains. Moreover, it discusses about the personalized medicine that uses ML algorithms for maintaining the individual genetic profiles, and precision medicine and the efficiency of drugs. In addition, this review article focused on the usage of advanced algorithms over CDSSs to enrich the healthcare decisions. Finally, this paper addressed cancer disease issues over the various medical applications.

a. Limitations

The limitations of this study include not considering the role of neural networks and reinforcement learning, and no comparative analysis done with respect to all the necessary evaluation metrics.

b. Future work

This work can be enhanced further by discussing the role of DL algorithms and optimization techniques over the decision-making process in various healthcare applications.

Authors' Contributions

LP, ND, SG, and AM were the major contributors in writing the manuscript and screening the articles. JS was the research supervisor and final editor. All authors agreed on the study's design and approved the manuscript's final version.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] Holford NH, Kimko HC, Monteleone JP, Peck CC, "Simulation of clinical trials", *Annu Rev Pharmacol Toxicol*, Vol. 40, No. 1, pp. 209–234, 2000. <http://dx.doi.org/10.1146/annurev.pharmtox.40.1.209>
- [2] Kuo I-Y, Hsieh C-H, Kuo W-T, Chang C-P, Wang Y-C, "Recent advances in conventional and unconventional vesicular secretion pathways in the tumor microenvironment", *J Biomed Sci.*, Vol.

29, Article No. 56, 2022. <http://dx.doi.org/10.1186/s12929-022-00837-8>

[3] Neppelenbroek KH, Campanha NH, Spolidorio DMP, Spolidorio LC, Seó RS, Pavarina AC, "Molecular fingerprinting methods for the discrimination between *C. albicans* and *C. dubliniensis*", *Oral Dis.*, Vol. 12, No. 3, pp. 242-253, 2006. <http://dx.doi.org/10.1111/j.1601-0825.2005.01189.x>

[4] Miró O, Sánchez M, Espinosa G, Coll-Vinent B, Bragulat E, Millá J, "Analysis of patient flow in the emergency department and the effect of an extensive reorganisation", *Emerg Med J.*, Vol. 20, No. 2, pp. 143-148, 2003. <http://dx.doi.org/10.1136/emj.20.2.143>

[5] Gamberi T, Chiappetta G, Fiaschi T, Modesti A, Sorbi F, Magherini F, "Upgrade of an old drug: Auranofin in innovative cancer therapies to overcome drug resistance and to increase drug effectiveness", *Med Res Rev.*, Vol. 42, No. 3, pp. 1111-1146, 2022. <http://dx.doi.org/10.1002/med.21872>

[6] Luca AR, Ursuleanu TF, Gheorghe L, Grigorovici R, Iancu S, Hlusneac M, et al., "Impact of quality, type and volume of data used by deep learning models in the analysis of medical images", *Inform Med Unlocked*, Vol. 29, Article No. 100911, 2022. <http://dx.doi.org/10.1016/j.imu.2022.100911>

[7] Nalini T, Revathi S, "An enhanced clustering algorithm implemented on biological data in data mining", *International Journal of Pharma and Bio Sciences*, Vol. 4, No. 2, pp. 1281 – 1286, 2013.

[8] Kaur I, Doja MN, Ahmad T, "Data mining and machine learning in cancer survival research: An overview and future recommendations", *J Biomed Inform*, Vol. 128, Article No. 104026, 2022. <http://dx.doi.org/10.1016/j.jbi.2022.104026>

[9] Napolitano G, Fox C, Middleton R, Connolly D, "Pattern-based information extraction from pathology reports for cancer registration", *Cancer Causes Control*, Vol. 21, No. 11, pp. 1887-1894, 2010. <http://dx.doi.org/10.1007/s10552-010-9616-4>

[10] Peng C, Cheng J, Cheng Q, "A supervised learning model for high-dimensional and large-scale data", *ACM Transactions on Intelligent System Technology*, Vol. 8, No. 2, pp. 1-23, 2017. <http://dx.doi.org/10.1145/2972957>

[11] Parikh RB, Manz CR, Nelson MN, Evans CN, Regli SH, Connor O, et al., "Clinician perspectives on machine learning prognostic algorithms in the routine care of patients with cancer: a qualitative study", *Supportive Care in Cancer*, Vol. 30, pp. 4363-4372, 2022.

[12] Mak K-K, Pichika MR, "Artificial intelligence in drug development: present status and

future prospects", *Drug Discovery Today*, Vol. 24, No. 3, pp. 773-780, 2019. <http://dx.doi.org/10.1016/j.drudis.2018.11.014>

[13] Pun FW, Ozerov IV, Zhavoronkov A, "AI-powered therapeutic target discovery", *Trends Pharmacol Sci.*, Vol. 44, No. 9, pp. 561-572, 2023. <http://dx.doi.org/10.1016/j.tips.2023.06.010>

[14] Stenzinger A, Alber M, Allgäuer M, Jurmeister P, Bockmayr M, Budczies J, et al., "Artificial intelligence and pathology: From principles to practice and future applications in histomorphology and molecular profiling", *Semin Cancer Biology*, Vol. 84, pp. 129-143, 2022. <http://dx.doi.org/10.1016/j.semcancer.2021.02.011>

[15] Lee WW, Alkureishi ML, "The impact of EMRs on communication within the doctor-patient relationship", In: Papadakos, P., Bertman, S. (eds) *Distracted Doctoring*. Springer, Cham., pp. 101-120, 2017. https://doi.org/10.1007/978-3-319-48707-6_9

[16] Khorram-Manesh A, Dulebenets MA, Goniewicz K, "Implementing public health strategies-the need for educational initiatives: A systematic review", *Int J Environ Res Public Health*, Vol. 18, No. 11, Article No. 5888, 2021. <http://dx.doi.org/10.3390/ijerph18115888>

[17] Drake CG, "Combination immunotherapy approaches", *Ann Oncol.*, Vol. 23, Suppl. 8, pp. 41-46, 2012. <http://dx.doi.org/10.1093/annonc/mds262>

[18] Watanabe N, Woo SL-Y, Papageorgiou C, Celechovsky C, Takai S, "Fate of donor bone marrow cells in medial collateral ligament after simulated autologous transplantation", *Microsc Res Tech*, Vol. 58, No. 1, pp. 39-44, 2002. <http://dx.doi.org/10.1002/jemt.10115>

[19] Aldwin CM, Spiro A, Levenson MR, Cupertino AP, "Longitudinal findings from the Normative Aging Study: III. Personality, individual health trajectories, and mortality", *Psychol Aging*, Vol. 16, No. 3, pp. 450-465, 2001. <http://dx.doi.org/10.1037//0882-7974.16.3.450>

[20] Ktistakis IP, Goodman G, Britzolaki A, "Advances in Assistive Technologies: Selected Papers in Honour of. Bourbakis", Vol. 3, pp. 11-31, 2021. <https://doi.org/10.1007/978-3-030-87132-1>

[21] Lee KY, Kim J, "Artificial intelligence technology trends and IBM Watson references in the medical field", *Korean Med Educ Rev*, Vol. 18, No. 2, pp. 51-57, 2016. <http://dx.doi.org/10.17496/kmer.2016.18.2.51>

[22] Ratti E, Graves M., "Explainable machine learning practices: opening another black box for reliable medical AI", *AI and Ethics*, Vol. 2, pp. 801-814, 2022. <https://doi.org/10.1007/s43681-022-00141-z>

[23] Boyd LA, Earnhardt RC, Dunn JT, Frierson HF, Hanks JB, "Preoperative evaluation and predictive

value of fine-needle aspiration and frozen section of thyroid nodules" *J Am Coll Surg.*, Vol. 187, No. 5, pp. 494–502, 1998. [http://dx.doi.org/10.1016/s1072-7515\(98\)00221-x](http://dx.doi.org/10.1016/s1072-7515(98)00221-x)

[24] Mahmood AA, Murgod R, Swarup Badapanda S, "Artificial Intelligence in Oncology: Present Potential, Prospective Prospects, and Ethical Reviews", *International Journal of Trends in OncoScience*, pp. 37–45, 2024.

[25] Zhou H, Baloch ZW, Nayar R, Bizzarro T, Fadda G, Adhikari-Guragain D, et al., "Noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP): Implications for the risk of malignancy (ROM) in the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC)", *Cancer Cytopathol.*, Vol. 126, No. 1, pp. 20–26, 2018. <http://dx.doi.org/10.1002/cncy.21926>

[26] Elliott Range DD, Dov D, Kovalsky SZ, Henao R, Carin L, Cohen J, "Application of a machine learning algorithm to predict malignancy in thyroid cytopathology: Machine learning and thyroid cytopathology", *Cancer Cytopathology*, Vol. 128, No. 4, pp. 287–295, 2020. <http://dx.doi.org/10.1002/cncy.22238>

[27] Balentine CJ, Domingo RP, Patel R, Laucirica R, Suliburk JW, "Thyroid lobectomy for indeterminate FNA: not without consequences", *J Surg Res.*, Vol. 184, No. 1, pp. 189–192, 2013. <http://dx.doi.org/10.1016/j.jss.2013.05.076>

[28] Park Y, Heider D, Hauschild A-C, "Integrative analysis of next-generation sequencing for next-generation cancer research toward artificial intelligence", *Cancers (Basel)*, Vol. 13, No. 13, Article No. 3148, 2021. <http://dx.doi.org/10.3390/cancers13133148>

[29] Tripathi R, Sharma P, Chakraborty P, Varadwaj PK, "Next-generation sequencing revolution through big data analytics", *Front Life Sci.*, Vol. 9, No. 2, pp. 119–49, 2016. <http://dx.doi.org/10.1080/21553769.2016.1178180>

[30] Biswas N, Chakrabarti S, "Artificial intelligence (AI)-based systems biology approaches in multi-omics data analysis of cancer", *Front Oncology*, Vol. 10, Article No. 588221, 2020. <http://dx.doi.org/10.3389/fonc.2020.588221>

[31] Perelmuter VM, Tashireva LA, Manskikh VN, Denisov EV, Savelieva OE, Kaygorodova EV, et al., "Heterogeneity and plasticity of immune inflammatory responses in the tumor microenvironment: Their role in the antitumor effect and tumor aggressiveness", *Biology Bull Rev.*, Vol. 8, No. 5, pp. 431–448, 2018. <http://dx.doi.org/10.1134/s2079086418050055>

[32] Rodellar J, Alf  rez S, Acevedo A, Molina A, Merino A, "Image processing and machine learning in

the morphological analysis of blood cells", *Int J. Lab Hematology*, Vol. 40, Suppl. 1, pp. 46–53, 2018. <http://dx.doi.org/10.1111/ijlh.12818>

[33] Kini SR, Miller JM, Hamburger JL, Smith-Purslow MJ, "Cytopathology of follicular lesions of the thyroid gland", *Diagn Cytopathology*, Vol. 1, No. 2, pp. 123–132, 1985. <http://dx.doi.org/10.1002/dc.2840010208>

[34] Saini A, Breen I, Pershad Y, Naidu S, Knuttinen MG, Alzubaidi S, et al., "Radiogenomics and radiomics in liver cancers", *Diagnostics (Basel)*, Vol. 9, No. 1, pp. 1–23, 2018. <http://dx.doi.org/10.3390/diagnostics9010004>

[35] Jussupow E, Spohrer K, Heinzl A, Gawlitza J, "Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence", *Inf Syst Res.*, Vol. 32, No. 3, pp. 713–735, 2021. <http://dx.doi.org/10.1287/isre.2020.0980>

[36] Liang M, Ting Y, Fu H, "Estimating individualized optimal combination therapies through outcome weighted deep learning algorithms", *Statistics in Medicine*, Vol. 37, No. 27, pp. 3869–3886, 2018. <http://arxiv.org/abs/1804.05378>

[37] Kumari M, Benzeval M, "Collecting biomarker data in longitudinal surveys", *Advances in Longitudinal Survey Methodology*, pp. 26–46, 2021.

[38] Mahmood AA, Jha AM, Manivannan K, "Precision Medicine: Personalizing The Fight Against Cancer", *International Journal of Trends in OncoScience.*, pp. 10–18, 2024. DOI: <https://doi.org/10.22376/ijtos.2023.2.1.10-18>

[39] Dia AK, Ebrahimpour L, Yolchuyeva S, Tonneau M, Lamaze FC, Orain M, et al., "The Cross-Scale Association between Pathomics and Radiomics Features in Immunotherapy-Treated NSCLC Patients: A Preliminary Study", *Cancers*, Vol. 16, No. 2, pp. 1–24, 2024. <https://doi.org/10.3390/cancers16020348>

[40] Huang C, Clayton EA, Matyunina LV, McDonald LD, Benigno BB, Vannberg F, et al., "Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy", *Scientific Reports*, Vol. 8, No. 1, Article No. 16444., 2018. <http://dx.doi.org/10.1038/s41598-018-34753-5>

[41] Hong R, Liu W, DeLair D, Razavian N, Feny   D, "Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models", *Cell Rep Med.*, Vol. 2, No. 9, Article No. 100400, 2021. <http://dx.doi.org/10.1016/j.xcrm.2021.100400>

[42] Wolfgruber TK, Presting GG, "JunctionViewer: customizable annotation software for

repeat-rich genomic regions", *BMC Bioinformatics*, Vol. 11, No. 1, pp. 1-23, 2010. <http://dx.doi.org/10.1186/1471-2105-11-23>

[43] Strug J, Strug B, "Machine learning approach in mutation testing", In: *Testing Software and Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 200–214, 2012.

[44] Zhang H, Ji J, Liu Z, Lu H, Qian C, Wei C, et al., "Artificial intelligence for the diagnosis of clinically significant prostate cancer based on multimodal data: a multicenter study", *BMC Med.*, Vol. 21, No. 1, Article No. 270, 2023. <http://dx.doi.org/10.1186/s12916-023-02964-x>

[45] Corrales-Rodriguez L, Soulières D, Weng X, Tehfe M, Florescu M, Blais N, "Mutations in NSCLC and their link with lung cancer-associated thrombosis: a case-control study", *Thromb Res.*, Vol. 133, No. 1, pp. 48-51, 2014. <http://dx.doi.org/10.1016/j.thromres.2013.10.042>

[46] Schweitzer N, Vogel A, "Systemic therapy of cholangiocarcinoma: From chemotherapy to targeted therapies", *Best Pract Res Clin Gastroenterol*, Vol. 29, No. 2, pp. 345–353, 2015. <http://dx.doi.org/10.1016/j.bpg.2015.01.002>

[47] Yardley-Jones A, Anderson D, Parke DV, "The toxicity of benzene and its metabolism and molecular pathology in human risk assessment", *Br J Ind Med.*, Vol. 48, No. 7, pp. 437–444, 1991. <http://dx.doi.org/10.1136/oem.48.7.437>

[48] Fredriksson NJ, Ny L, Nilsson JA, Larsson E, "Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types", *Nat Genet.*, Vol. 46, No. 12, pp. 1258–1263, 2014. <http://dx.doi.org/10.1038/ng.3141>

[49] Gruver AM, Lu H, Zhao X, Fulford AD, Soper MD, Ballard D, et al., "Pathologist-trained machine learning classifiers developed to quantitate celiac disease features differentiate endoscopic biopsies according to modified Marsh score and dietary intervention response", *Diagn Pathology*, Vol. 18, No. 1, Article No. 122, 2023. doi: 10.1186/s13000-023-01412-x

[50] Atkins D, Makridis CA, Alterovitz G, Ramoni R, Clancy C, "Developing and implementing predictive models in a learning healthcare system: Traditional and artificial intelligence approaches in the Veterans Health Administration", *Annu Rev Biomed Data Sci.*, Vol. 5, No. 1, pp. 393–413, 2022. <http://dx.doi.org/10.1146/annurev-biodatasci-122220-110053>

[51] Zhang S, Han P, Wu C, "Calibration techniques encompassing survey sampling, missing data analysis and causal inference", *Int Stat Rev.*, Vol. 91, No. 2, pp. 165–92, 2023. <http://dx.doi.org/10.1111/insr.12518>

[52] Ali WM, Alhumaidi MS, "Artificial Intelligence for Cancer Diagnosis & Radiology", *International Journal of Trends in OncoScience*, Vol. 13-18, 2023.

[53] Pei Q, Luo Y, Chen Y, Li J, Xie D, Ye T, "Artificial intelligence in clinical applications for lung cancer: diagnosis, treatment and prognosis", *Clinical Chemistry and Laboratory Medicine (CCLM)*, Vol. 60, pp. 1974–1983, 2022.

[54] Ye W, Luo C, Liu F, Liu Z, Chen F, "CD96 correlates with immune infiltration and impacts patient prognosis: A pan-cancer analysis", *Front Oncol.*, Vol. 11, Article No. 634617, 2021. <http://dx.doi.org/10.3389/fonc.2021.634617>

[55] Wulfschle JD, Pawletz CP, Steeg PS, Petricoin EF, Liotta L., "Proteomic approaches to the diagnosis, treatment, and monitoring of cancer. In: *New Trends in Cancer for the 21st Century*", In: *Proceedings of the International Symposium on Cancer: New Trends in Cancer for the 21st Century*. Valencia, Spain; US: Springer, pp. 59–68, 2002.

[56] Picard M, Scott-Boyer M-P, Bodein A, Périn O, Droit A, "Integration strategies of multi-omics data for machine learning analysis", *Comput Struct Biotechnol J.*, Vol. 19, pp. 3735–3746, 2021. <http://dx.doi.org/10.1016/j.csbj.2021.06.030>

[57] Al-Tashi Q, Saad MB, Muneer A, Qureshi R, Mirjalili S, Sheshadri A, et al., "Machine learning models for the identification of prognostic and predictive cancer biomarkers: A systematic review", *Int J Mol Sci.*, Vol. 24, No. 9, Article No. 7781, pp. 1-42, 2023. <http://dx.doi.org/10.3390/ijms24097781>

[58] Dewey FE, Pan S, Wheeler MT, Quake SR, Ashley EA, "DNA sequencing: clinical applications of new DNA sequencing technologies", *Circulation*, Vol. 125, No. 7, pp. 931–944, 2012. <http://dx.doi.org/10.1161/CIRCULATIONAHA.110.972828>

[59] Klein CA, Hölzel D, "Systemic cancer progression and tumor dormancy: mathematical models meet single cell genomics", *Cell Cycle.*, Vol. 5, No. 16, pp. 1788–1798, 2006. <http://dx.doi.org/10.4161/cc.5.16.3097>

[60] Kasula BY, "Harnessing Machine Learning for Personalized Patient Care", *Transactions on Latest Trends in Artificial Intelligence*, Vol. 4, No. 4, pp. 1-9, 2023.

[61] Thiyagarajan S, Chakravarthy T, Arivoli PV, "Diagnosing Breast Cancer with Machine Learning Algorithms", *International Journal of Pharmaceutical Research*, Vol. 23, pp. 1-13, 2020.

[62] Quinlan AR, "BEDTools: the Swiss-army tool for genome feature analysis", *Current protocols in*

bioinformatics, Vol. 47, pp. 11–22, 2014. <https://doi.org/10.1002/0471250953.bi1112s47>

[63] Singh Y, Bhatia PK, Sangwan O, “A review of studies on machine learning techniques”, *International Journal of Computer Science and Security*, Vol. 1, No. 1, pp. 70–84, 2007. <https://www.cscjournals.org/manuscript/Journals/IJCSS/Volume1/Issue1/IJCSS-7.pdf>

[64] Almomani F., “Prediction the performance of multistage moving bed biological process using artificial neural network (ANN)”, *Sci Total Environ.*, Vol. 744, Article No. 140854, 2020. <http://dx.doi.org/10.1016/j.scitotenv.2020.140854>

[65] Mahmoud AS, Alkhenizan A, Shafiq M, Alsoghayer S, “The impact of the implementation of a clinical decision support system on the quality of healthcare services in a primary care setting”, *J Family Med Prim Care*, Vol. 9, No. 12, pp. 6078–6084, 2020. http://dx.doi.org/10.4103/jfmpc.jfmpc_1728_20

[66] Anderson G, Horvath J, “The growing burden of chronic disease in America”, *Public Health Rep.*, Vol. 119, No. 3, pp. 263–270, 2004. <http://dx.doi.org/10.1016/j.phr.2004.04.005>

[67] Lenouvel E, Novak L, Biedermann A, Kressig RW, Klöppel S, “Preventive treatment options for fear of falling within the Swiss healthcare system: A position paper”, *Z Gerontol Geriat.*, Vol. 55, pp. 597–602, 2022. <https://doi.org/10.1007/s00391-021-01957-w>

[68] Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al., “Clinical decision support systems for the practice of evidence-based medicine”, *J Am Med Inform Assoc.*, Vol. 8, No. 6, pp. 527–534, 2001. <http://dx.doi.org/10.1136/jamia.2001.0080527>

[69] Almkvist O., “Neuropsychological features of early Alzheimer’s disease: preclinical and clinical stages”, *Acta Neurol. Scand. Suppl.*, Vol. 165, pp. 63–71, 1996. <http://dx.doi.org/10.1111/j.1600-0404.1996.tb05874.x>

[70] Roncaglioni A, Toropov AA, Toropova AP, Benfenati E., “In silico methods to predict drug toxicity”, *Current Opinion Pharmacology*, Vol. 13, No. 5, pp. 802–806, 2013. <http://dx.doi.org/10.1016/j.coph.2013.06.001>

[71] Judson R, Elloumi F, Setzer RW, Li Z, Shah I., “A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model”, *BMC Bioinformatics*, Vol. 9, No. 1, Article No. 241, 2008. <http://dx.doi.org/10.1186/1471-2105-9-241>

[72] Raies AB, Bajic VB., “In silico toxicology: computational methods for the prediction of chemical toxicity: Computational methods for the prediction of

chemical toxicity”, *Wiley Interdisciplinary Rev Comput. Mol Sci.*, Vol. 6, No. 2, pp. 147–172, 2016. <http://dx.doi.org/10.1002/wcms.1240>

[73] Paster I, Shacham M, Brauner N, “Investigation of the relationships between molecular structure, molecular descriptors, and physical properties”, *Ind Eng Chem Res.*, Vol. 48, No. 21, pp. 9723–9734, 2009. <http://dx.doi.org/10.1021/ie801318y>

[74] Tran TT, Wibowo S, Tayara A, Chong H, “Artificial Intelligence in Drug Toxicity Prediction: Recent Advances, Challenges, and Future Perspectives”, *Journal of Chemical Information and Modeling*, Vol. 63, No. 9, pp. 2628–2643, 2023.

[75] Huang B, von Lilienfeld OA. Ab initio, “Machine Learning in chemical compound space”, *Chem. Rev.*, Vol. 121, No. 16, pp. 10001–10036, 2021. <http://dx.doi.org/10.1021/acs.chemrev.0c01303>

[76] Hildebrandt MAT, Komaki R, Liao Z, Gu J, Chang JY, Ye Y, et al., “Genetic variants in inflammation-related genes are associated with radiation-induced toxicity following treatment for non-small cell lung cancer”, *PLoS One*, Vol. 5, No. 8, Article No. e12402, 2010. <http://dx.doi.org/10.1371/journal.pone.0012402>

[77] Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin S-F, et al, “Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling”, *Sci Transl. Med.*, Vol. 4, Article No. 157, 2012. <http://dx.doi.org/10.1126/scitranslmed.3004330>

[78] Banning M, “A review of clinical decision making: models and current research”, *J Clin Nurs*, Vol. 17, No. 2, pp. 187–95, 2008. <http://dx.doi.org/10.1111/j.1365-2702.2006.01791.x>

[79] Giraud P, Giraud P, Gasnier A, El Ayachy R, Kreps S, Foy J-P, et al., “Radiomics and machine learning for radiotherapy in head and neck cancers”, *Front Oncol.*, Vol. 9, No. 174, 2019. <http://dx.doi.org/10.3389/fonc.2019.00174>

[80] Palma G, Monti S, Conson M, Pacelli R, Cella L, “Normal tissue complication probability (NTCP) models for modern radiation therapy”, *Semin Oncol.*, Vol. 46, No. 3, pp. 210–218, 2019. <http://dx.doi.org/10.1053/j.seminoncol.2019.07.006>

[81] Toropov AA, Toropova AP, Raska I Jr, Leszczynska D, Leszczynski J, “Comprehension of drug toxicity: software and databases”, *Comput. Biol. Med.*, Vol. 45, pp. 20–25, 2014. <http://dx.doi.org/10.1016/j.combiomed.2013.11.013>

[82] Watkins PB, Seligman PJ, Pears JS, Avigan MI, Senior JR, “Using controlled clinical trials to learn more about acute drug-induced liver injury”, *Hepatology*, Vol. 48, No. 5, pp. 1680–1689, 2008. <http://dx.doi.org/10.1002/hep.22633>

[83] Smusz S, Kurczab R, Bojarski AJ, "A multidimensional analysis of machine learning methods performance in the classification of bioactive compounds", *Chemometr Intell Lab Syst*, Vol. 128, pp. 89–100, 2013. <http://dx.doi.org/10.1016/j.chemolab.2013.08.003>

[84] Lu H, Ma X, Huang K, Azimi M., "Carbon trading volume and price forecasting in China using multiple

machine learning models", *J Clean Prod.*, Vol. 249, Article No. 119386, 2020. <http://dx.doi.org/10.1016/j.jclepro.2019.119386>

[85] Peng S, Wang W, Chen Y, Zhong X, Hu Q, "Regression-based hyperparameter learning for support vector machines", *IEEE Trans Neural Netw Learn Syst.*, Vol. 35, No. 12, pp. 18799-18813, 2024. <http://dx.doi.org/10.1109/TNNLS.2023.3321685>