

## Review

# Systems Biology in Cancer Diagnosis Integrating Omics Technologies and Artificial Intelligence to Support Physician Decision Making

Alaa Fawaz, Alessandra Ferraresi  and **Ciro Isidoro** \* 

Laboratory of Molecular Pathology, Department of Health Sciences, Università del Piemonte Orientale, 28100 Novara, Italy; 20041632@studenti.uniupo.it (A.F.); alessandra.ferraresi@med.uniupo.it (A.F.)

\* Correspondence: ciro.isidoro@med.uniupo.it; Tel.: +39-0321-660507; Fax: +39-0321-620421

**Abstract:** Cancer is the second major cause of disease-related death worldwide, and its accurate early diagnosis and therapeutic intervention are fundamental for saving the patient's life. Cancer, as a complex and heterogeneous disorder, results from the disruption and alteration of a wide variety of biological entities, including genes, proteins, mRNAs, miRNAs, and metabolites, that eventually emerge as clinical symptoms. Traditionally, diagnosis is based on clinical examination, blood tests for biomarkers, the histopathology of a biopsy, and imaging (MRI, CT, PET, and US). Additionally, omics biotechnologies help to further characterize the genome, metabolome, microbiome traits of the patient that could have an impact on the prognosis and patient's response to the therapy. The integration of all these data relies on gathering of several experts and may require considerable time, and, unfortunately, it is not without the risk of error in the interpretation and therefore in the decision. Systems biology algorithms exploit Artificial Intelligence (AI) combined with omics technologies to perform a rapid and accurate analysis and integration of patient's big data, and support the physician in making diagnosis and tailoring the most appropriate therapeutic intervention. However, AI is not free from possible diagnostic and prognostic errors in the interpretation of images or biochemical-clinical data. Here, we first describe the methods used by systems biology for combining AI with omics and then discuss the potential, challenges, limitations, and critical issues in using AI in cancer research.



**Citation:** Fawaz, A.; Ferraresi, A.; Isidoro, C. Systems Biology in Cancer Diagnosis Integrating Omics Technologies and Artificial Intelligence to Support Physician Decision Making. *J. Pers. Med.* **2023**, *13*, 1590. <https://doi.org/10.3390/jpm13111590>

Academic Editor: Daniele Giansanti

Received: 17 October 2023

Revised: 7 November 2023

Accepted: 8 November 2023

Published: 10 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** artificial intelligence; medical technology; smart health; digital health; omics technologies; imaging; diagnosis; personalized medicine

## 1. Introduction

Delayed diagnoses, misdiagnoses, and missed diagnoses impact patient health and safety, and have great societal consequences. Mistakes in diagnosis may account for up to 60% of all medical errors and are accountable for up to 80,000 deaths in U.S. medical centers each year [1]. Typically, clinicians have limited time to make decisions based on the interpretation of huge amounts of laboratory, imaging, and clinical data, and this increases the risk of underestimating (or sometimes overestimating) some data. Furthermore, subjective factors, such as personal experience and medical specialty, are potential bias factors that influence the accuracy of diagnosis [2].

Artificial Intelligence (AI), a field of computer science used for prediction and automation, has emerged as a potential solution to promote a precision approach in healthcare and is expected to reduce errors caused by human judgment in various medical domains [3].

Cancer is the leading cause of death in people, accounting for an estimated 10 million deaths by 2020 [4]. It is a complex disease resulting from anomalies in physiological processes involving genes, coding and non-coding RNAs, proteins, metabolites, and other biomolecules [5,6]. To understand such a complex disease from its onset to its progression, multi-omics analysis of these numerous bio-entities is required. Modern biotechnologies

allow for the high throughput analysis of the sequence and expression of many genes (genomics and epigenomics), proteins and their post-translational modifications (proteomics, phospho-proteomics and glycol-proteomics), RNAs (RNA transcriptomics), non-coding RNAs (including miRNAs and long-non-coding RNAs), and metabolites (metabolomics) from the same organism [7]. However, a platform where all these big data are integrated to uncover correlations and synergisms among the biological pathways and processes is required. Systems biology combines the power of AI and of multi-omics technologies for modeling the signaling and metabolic signature of a given cancer. This is instrumental for designing effective diagnostic and prognostic markers and novel and patient-tailored therapeutic interventions.

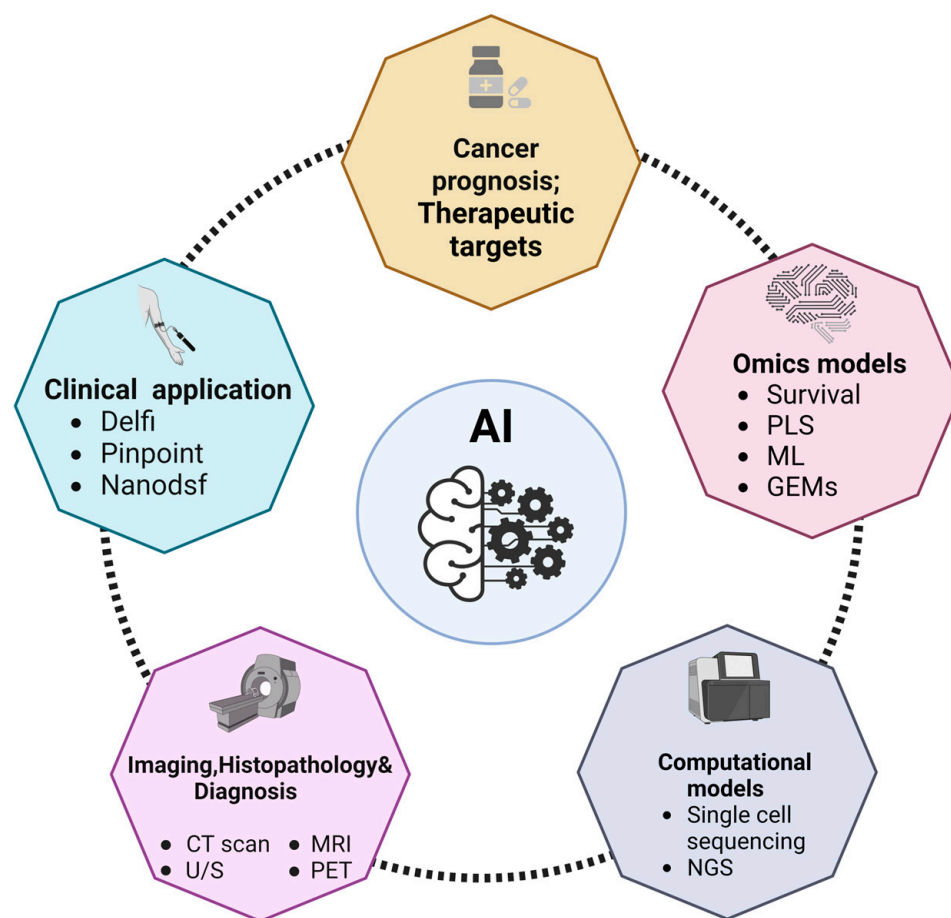
Despite difficulties in providing individualized and data-driven care, advancements in screening, diagnosis, treatment, and survival rate in cancer patients have been remarkable in recent decades [8]. Early detection and prognosis prediction represent two crucial clinical needs for limiting cancer progression. Body and organ computed scan methodologies, the histopathology imaging of biopsies, and a range of blood tests for detecting biomarkers are instrumental in the initial diagnosis process and for determining cancer staging, the grade of malignancy, and prognosis. These approaches do not provide information on the molecular alterations that precede and follow the onset of cancer. Molecular and omics technologies can provide a genetic, epigenetic, and metabolic profile of the tumor that can better define such alterations thus helping to determine the most appropriate treatment as well as predict the response to therapy [9,10].

The development and extensive use of high-throughput technologies has ushered in the era of biological and medical big data. This has led to the accumulation of data sets on a large scale, thereby opening a wide range of potential applications for data-driven methods in cancer treatment, spanning from basic research to clinical practice: molecular tumor characterization, tumor heterogeneity, drug discovery and potential therapeutic strategies. As a result, the data-driven research field of bioinformatics adapts data mining techniques, such as systems biology, machine learning, and deep learning, which are discussed in this review paper. Systems biology uses a data-driven approach to identify important signaling pathways. The pathway-oriented analysis is extremely important in cancer research because it helps researchers comprehend the molecular features and heterogeneity of tumors and tumor subtypes [11]. In this context, the proper clinical care for cancer patients can be improved by the introduction of AI in cancer detection, diagnosis, and treatment [12–15].

AI-based technologies applied to oncology aim at improving clinical practice, including but not limited to the early and accurate diagnosis and prediction of personalized outcomes (i.e., prognosis and therapy response), by acquiring a profound perception of tumor molecular biology through the association of multiple biological parameters [16].

#### *Artificial Intelligence in Medicine at Glance*

AI is meant to mimic human cognitive abilities in elaborating the information but at a much higher speed and with no emotional interference. The main types of AI that apply to cancer-patient healthcare include machine learning (ML) and its evolved subtype deep learning (DL), which can assist in making a rapid and more accurate diagnosis (based on biochemical, clinical data, and medical imaging), in discovering and developing new drugs, in designing personalized therapy, in predicting the therapy response, and in guiding the robotic surgery [17,18] (Figure 1).



**Figure 1.** Overview of the applications of AI to cancer diagnosis and oncology research field. The scheme depicts the main fields of application of AI discussed in this review. Abbreviations: computed tomography, CT; gene expression models, GEMs; machine learning, ML; magnetic resonance imaging, MRI; nano differential scanning fluorimetry, Nanodsf; next-generation sequencing, NGS; positron emission tomography, PET; partial least squares analysis, PLS; ultrasound imaging, U/S.

Current AI systems have been involved to be used in a variety of clinical settings, including (i) image-based computer-aided discovery and diagnosis in various medical specialties, (ii) the translation of genomic information for recognizing genetic variants using high-throughput sequencing technologies, and (iii) the prediction and tracking of patient's prognosis [19,20]. Moreover, they have been implemented as well in (iv) the discovery of new biomarkers by combining omics and phenotype data, (v) the detection of health status using biological signals (e.g., enzyme activity and protein concentration) obtained from wearable devices, and (vi) the production and implementation of autonomous robots in medical procedures [19,20].

The creation of AI models that predict the properties of vast and interconnected networks found in living organisms would allow for a thorough examination of how signaling molecules generate functional cellular reactions. Machine learning (ML) algorithms, a subset of AI, are capable of making decisive interpretations of large, complex data sets, making them an effective tool for analyzing and comprehending multi-omics data for patient-specific observations [20]. We can anticipate the remarkable growth of AI in the medical field in light of the digital acquisition of high-dimensional and annotated medical data, the progress of ML methods, open ML data science, and advancements in computational power and storage services [20]. AI is expected to make it easier to diagnose specific illnesses in patients. Commonly, deep learning (DL) architectures are analogous to artificial neural networks of multiple non-linear tiers. Over the past decade, a large variety of DL designs have been developed depending on the input data type and the purpose

of the research. Moreover, the assessment of the model's efficiency has revealed that DL application on cancer prognosis surpasses other traditional ML techniques. DL frameworks have also been used in cancer diagnosis, classification, and treatment by utilizing genomic profiles and phenotype information. Systems biology has been an effective method to comprehend the complex molecular profile of cancers, interpret the mechanisms of tumor progression, and allow for the amalgamation of omics data as well as the characterization of diverse tumors [21,22].

## 2. Omics Data for Identifying Cancer Metabolic Biomarkers

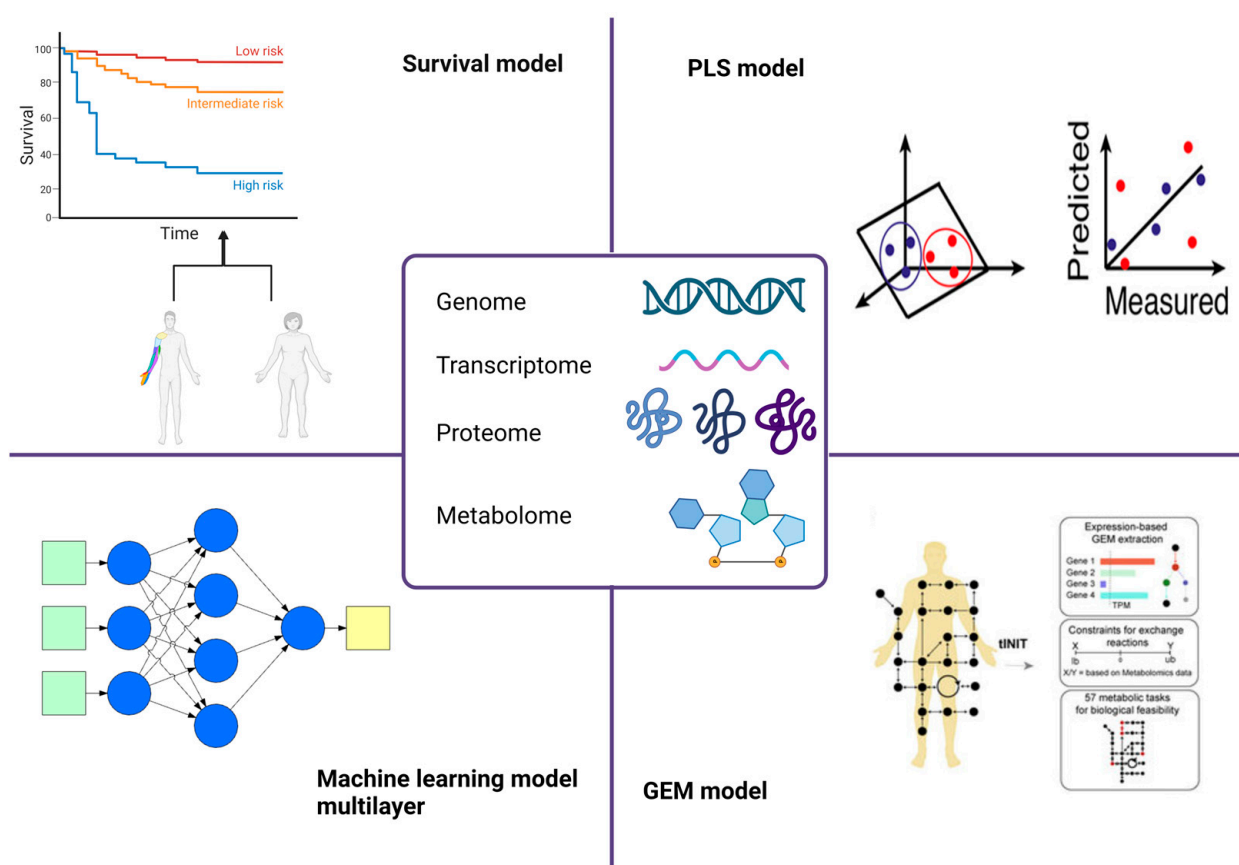
Omics technologies allow for the in depth analysis of the molecular characteristics of cancer at both bulk and single-cell level, providing a wealth of multi-omics data that challenge the capability of scientists and medical doctor to combine for drawing a consistent picture of the multilayer complexity of cancer biology. Genomic, epigenomic, transcriptomic, proteomic, and metabolomic data can be elaborated using appropriate models for making predictions about prognosis and treatment response in a patient-tailored (personalized) manner [13,15,22].

### 2.1. Survival Models

To find cancer metabolic biomarkers, survival models have been used more frequently than partial least squares (PLS) models, ML models, and gene expression modeling (GEM) [23] (Figure 2). The Kaplan–Meier method, the log-rank test, and/or the Cox regression model are representative survival models used in cancer studies. These models are used to describe the likelihood of survival (or survival curve) for a group of patients after treatment, compare the survival curves of two or more treatment groups, and describe the effects of multiple explanatory (independent) variables, profiles of gene expression, and metabolite concentration) on survival curves, respectively. In contrast to Kaplan–Meier models, which must discretize their data, the Cox regression model has the advantage of processing continuous values directly, minimizing data loss [24]. In their study, based on GEM of seven major metabolic pathways, Peng and colleagues identified 30 tumor subtypes in 33 different cancer types (such as breast invasive carcinoma, cholangiocarcinoma, colorectal cancer, glioblastoma multiforme, gastrointestinal tumors, lung cancer, pancreatic cancer, and ovarian serous cystadenocarcinoma, among others) and evaluated the clinical utility of so-called metabolic expression subtypes. For this, correlations between metabolic expression subtypes and their corresponding prognosis were investigated using the Kaplan–Meier method, log-rank test, and Cox regression model. Consequently, subtypes with upregulated lipid metabolism appeared to have a better prognosis than subtypes with upregulated glyceimic, nucleotide, vitamin, and cofactor metabolism. The association of various somatic mutations in cancer driver genes with metabolic expression subtypes has also been discovered. Two transcription factors, SNAI1 and RUNX1, were identified from knockdown studies as potential therapeutic targets for a subtype of cancer with upregulated carbohydrate metabolism that consistently had a poor prognosis across cancer types [23].

### 2.2. PLS Models

Partial least squares regression (PLS) was initially created as a regression model that processes numerous independent variables that are correlated and produce numerous dependent variables, which many statistical and ML techniques cannot directly handle. PLS models and their variations, particularly PLS-discriminant analysis (PLS-DA) are frequently used for the analysis of omics data with a focus on metabolomics [25]. PLS-DA has been primarily used to extract insights from large datasets of omics data, such as identifying metabolites from metabolome data that differentiate between cancer cells in their various statuses. PLS-DA might have an overfitting issue too, like other data mining techniques, so it needs thorough validation, frequently performed through cross-validation [26].



**Figure 2.** Overview of the omics technologies exploited in cancer diagnosis/prognosis. The scheme depicts the main omics models currently used in biomarker identification. Abbreviations: gene expression modeling, GEM; partial least squares analysis, PLS.

PLS-DA and its variants have been used to analyze metabolome data to identify a variety of cancers, including breast cancer, glioma, non-small cell lung cancer, oral precancerous cells, cervical precancerous lesions, and prostate cancer [27,28]. Among its advantages, PLS-DA allows for the analysis of highly collinear and noisy data. Moreover, the calibration model provides a subset of useful statistics, including prediction accuracy, scores and loading plots. However, a potential limitation has emerged when this method was applied to metabolomics; the use of this model by non-experts may produce inaccurate results, owing to a lack of appropriate statistical validation [29] (Table 1).

**Table 1.** Summary of the main advantages and limitations of PLS models.

Advantages	Limitations
Ability to robustly handle more descriptor variables	Higher risk of overlooking ‘real’ correlations
Provide more predictive accuracy	Sensitivity to the relative scaling of the descriptor variables
Low risk of chance correlation	

### 2.3. Genome-Scale Metabolic Models

Gene expression modeling (GEM) is a computational model based on the law of mass conservation of metabolites and allows for the prediction of metabolic fluxes for entire biochemical reactions taking place inside a cell by using numerical optimization [30,31]. Technically, GEM describes the participation of each metabolite for an entire set of biochemical reactions in the form of a stoichiometric matrix and is simulated using varied forms



of objective functions and constraints that reflect genetic and environmental conditions of interest. As a result, GEM allows for the efficient simulation of a target cell's metabolic phenotypes under a wide range of genetic and environmental conditions. GEM can also be integrated with omics data, such as RNA-seq, for building a cell-specific model and thereafter modeling multicellular organisms. In comparison with ML models, GEMs generate more interpretable prediction outcomes that grasp a cell-specific metabolic phenotype. GEM simulations, however, demand consideration. Due to the possibility of biologically incorrect objective functions or constraints, it is advised to proceed with the analysis of the predicted intracellular metabolic flux distributions from GEMs with caution. A representative issue is the use of constraints that do not accurately reflect a culture medium. Finally, GEMs do not directly produce additional data for regulatory and signaling networks, which are also crucial for understanding the physiology of a cell [32,33] (Table 2).

**Table 2.** Summary of the main advantages and limitations of GEM models.

Advantages	Limitations
Explore metabolism in multiple cell types	Uncertainties in the estimated parameters regarding quantitative flux predictions
Validating or discovering biomarkers for screening, diagnostics, prognostics, and/or patient stratification	Ambiguous normalization of experimentally quantified fluxes
Identify cancer-specific metabolic features that constitute generic potential drug targets for cancer treatment	

#### 2.4. Machine Learning Models

The classification task of disease prediction has been thoroughly studied in medical oncology and cancer research, based on well-established machine learning algorithms for dealing with binary or multi-class learning problems. Patient categorization would allow for the development of ML-based predictive models capable of assessing risk stratification with generalizable performance. Based on images and genetic data, DL models were trained to classify and detect disease subtypes. These data-driven approaches demonstrated the superiority of ML-based frameworks for leveraging heterogeneous datasets for improved diagnosis and treatment [34].

#### 2.5. Deep Neural Networks (DNNs)

Deep neural network (DNN) models are rapidly evolving and becoming more sophisticated. They have been widely used in biomedical research across the board. Initially, large-scale imaging and video data aided its development. While most biomedical data sets are not considered big data, the rapid data accumulation enabled by NGS made it suitable for the application of DNN models that require a large amount of training data [35]. In 2019, for example, Samiei et al. used TCGA-based large-scale cancer data as benchmark datasets for bioinformatics machine learning research, such as Image-Net in computer vision [36]. Following that, large-scale public cancer data sets like the TCGA encouraged the widespread use of DNNs in cancer research [37] (Table 3).

**Table 3.** Summary of the main advantages and limitations of DNN models.

Advantages	Limitations
Ability to handle complex data and relationships	Massive data requirement
Effective at producing high-quality results	High processing and computational power
Extremely scalable because of its capacity to analyze large volumes of data	Black box problem making them hard to debug and understand how they make decisions

### 2.6. Graph Neural Networks (GNNs)

Graph neural networks (GNNs) have achieved great results and are being progressively employed in a node classification task. It offers a strategy to acquire novel representations of nodes by combining the features of its local neighborhood and connectivity. Recently, some GNN-based approaches have been proposed to forecast the molecular subtyping of cancer. Rhee et al. created a graph convolutional network (GCN)-based model to investigate the gene–gene alliance and information transmission for cancer subtyping [38]. Lee et al. developed a GCN model with a focus on the mechanisms to learn pathway-level representations of cancer samples for their subtype classification [39]. Even though GNNs are strong, it is reported that they are susceptible when the structure of the graph and nodes' features are polluted with noise [40]. Thus, a robust GNN model is required for the precise and stable prediction of cancer subtypes [41] (Table 4).

**Table 4.** Summary of the main advantages and limitations of GNN models.

Advantages	Limitations
Rapid processing of massive data	Limited to a fixed number of points
Reliable performance in mining deep-level topological information	Time and space complexity are higher
Extracting text relationship and reasoning the structure of graphics and images	Less handling of edges of graphs based on their types and relations

### 3. Computational Models for the Prediction of Cancer Metabolic Biomarkers

Single-cell sequencing allows for the study of the molecular changes occurring in individual cells within the tumor mass. Nonetheless, attributing a specific cellular annotation (in terms of cell type or metabolic state) is challenging, in particular to distinguish cancer cells in single-cell or spatial sequencing experiments. The information provided by high-throughput single-cell sequencing provides not only the description of distinct cellular annotations but also the functional annotation of single cells, for example the estimation of the differentiation potential, vulnerability to metabolic changes, and a prediction of cellular crosstalk [42]. However, the use of this technology also raises computational difficulties [43]. One of the major challenges in single-cell data analysis is to attribute a cell annotation to each cell analyzed [44]. The magnitude of the generated datasets renders the manual annotation processes unfeasible, whereas the peculiarities of data generation have stimulated the spread of novel and creative classification methods [45]. This limitation is particularly found in datasets coming from cancer tissues, in which the variability in the transcriptomic states does not conform to traditionally defined cell types [46,47].

In addition to the genome data, the transcriptome, proteome, and metabolome data offer snapshots of a cell's phenotype space. As shown by PCAWG58 and TCGA59, which also provide transcriptome data in addition to genome data, the transcriptome, particularly RNA sequencing (RNA-seq), is the most frequently generated omics data among these. To perform more complex transcriptomic analyses, bulk RNA-seq has evolved into single-cell RNA-seq (scRNA-seq) and spatial RNA-seq. To enable a greater understanding of cell phenotypes, massive amounts of proteome and metabolome data are being generated for various human cells [48,49]. The Human Metabolome Database (HMDB) and Human Protein Atlas (HPA) are representative databases for the human proteome and metabolome, respectively. Integrative omics analysis has gained importance since these omics data are complementary to one another, and multiple omics data are frequently generated for a target cell [50,51].

Several studies have combined NGS data with ML to propose a novel data-driven methodology in systems biology [52]. Several network-based ML models have been implemented to analyze cancer data and aid in the understanding of novel mechanisms in cancer development [53,54]. Furthermore, the use of DNN models for large-scale data analysis enhanced the accuracy of computational models for the prediction of the muta-

tional landscape, molecular subtyping and drug repurposing [55–58]. A growing number of DNN-based applications have recently integrated multi-omics and systems biology data into the learned models. Such approaches aim to apply the DNN model to well-established biomedical knowledge, thereby improving our understanding of diseases and therapeutic effects in novel ways [59,60].

A common aim of NGS data analysis in cancer research is the identification of potential biomarkers that are predictive of specific cancer types or subtypes. A variety of bioinformatics tools and ML models, for example, aim to identify a molecular signature that is significantly altered in cancer cells on a genomic, transcriptomic, or epigenomic level. Statistical and ML methods are typically used to identify the best set of biomarkers, such as single nucleotide polymorphisms (SNPs), mutations, or differentially expressed genes that are important in cancer progression. Previously, those markers had to be discovered or validated using time-consuming *in vitro* analysis. As a result, systems biology provides *in silico* solutions to validate such findings by utilizing biological pathways or gene ontology data [61].

#### 4. AI in Cancer Prognosis

Detecting and predicting the course of the disease are key components to controlling tumor enlargement and providing adequate treatment to cancer patients. With the understanding that cancer can affect individuals differently, AI has been utilized to isolate subgroups within the patient population based on prognosis and survival data. Aside from segmentation, AI has pinpointed biomarkers that can indicate the recurrence of the disease. AI has been implemented to prognosticate high-risk neuroblastoma patients. Utilizing combined gene expression and copy number variations, an unsupervised learning algorithm called auto encoder determined significant features, which were then used for division into two clusters [62]. In a separate study, Francescatto et al. employed the integrative network fusion framework together with an ML classifier to distinguish features that could differentiate between distinct outcomes of patients [63].

DL-based neural networks have also been applied to breast cancer survival prognosis. To prevent overfitting effects due to the vast size of omics data, the SALMON survival analysis algorithm operates on eigengene matrices of co-expression network modules. To enhance robustness, it brings together traditional cancer biomarkers and multi-omics information and pinpoints key feature genes and cytobands [64]. The use of a DL-based algorithm allows for the combination of the information from the same gene across different types of omics data, thus resulting in a successful and insightful analysis [65].

#### 5. AI in the Identification of Therapeutic Targets

A subset of alternative network approaches to identifying cancer targets are provided by network-based biology analysis algorithms. More importantly, because different algorithms can look at network data from different angles, they can compensate for each other to provide accurate biological explanations [66].

Interactome data can be organized and represented in the form of network structures to explain the molecular mechanisms underlying carcinogenesis, where the nodes are biological entities (genes, proteins, mRNAs, and metabolites) while the edges represent the associations–interactions between them (gene co-expression, signaling transduction, gene regulation, and physical interaction between proteins) [67,68]. AI algorithms could efficiently process biological network data by implementing classification, clustering, and prediction tasks in biological networks using machines or programs that enhance human intelligence [69]. As a result, AI algorithms will be able to elucidate the complexity of cancer behavior that rely on the interactions between genes and their products in biological network structures [70], allowing us to better understand carcinogenesis and identify novel anti-cancer targets [71].

One of the fundamental needs of precision oncology is anticipating therapy response for a patient population. The advantages of ML strategies have been tried for treatment



response displaying and expectation following both center-based and component choice-based strategies [72]. The profound neural system-based examination has been used to predict therapy response. MOLI, a multi-omics late mix strategy in light of a profound neural system, consolidates somatic transformation, and duplicates number variation and quality articulation information to anticipate medication reaction conduct. MOLI is additionally utilized for board medication information, and information on medications with a similar target [73].

The Support Vector Machine (SVM) and the Leave-One-Out Cross-Validation (LOOCV) models have been employed to detect significant changes in RNA and miRNA transcriptomics data between from pancreatic ductal adenocarcinoma specimens and normal tissues. These features (selected RNAs and miRNAs) in combination with miRNA target expression data were further exploited to identify efficient diagnostic markers that were validated in other distinct datasets and biologically interpreted by pathway analysis of the corresponding target genes [74]. Moreover, ML-based analysis has been utilized to discover specific anticancer drug targets for breast tumors [75]. The characteristic genes extracted from multi-omics data of breast cancer with the aid of capsule network-based modeling were compared with well-known oncogenes, and novel genes were identified [76].

Recently, a comprehensive examination of nine cancers has demonstrated that proteomics data combined with gene expression, miRNAs expression and genomics is more effective in predicting the responsiveness of drugs and molecules specifically designed to target them. This research was conducted across 58 cell lines over nine cancers with Bayesian Efficient Multiple Kernel Learning (BEMKL) models [72]. This confirms the robustness of multi-omics data analysis across cancer types.

## 6. AI Clinical Application

The DELFI technology, which uses a blood test to indirectly evaluate the packing of DNA in the nucleus of a cell by assessing the bulk and amount of cell-free DNA present in the flow from various regions of the genome, is one example of AI in clinical practice. Cancer cells release DNA into the bloodstream when they die. DELFI uses ML to investigate millions of cell-free DNA pieces for unusual design in order to distinguish the occurrence of cancer. The strategy provides a perspective on cell-free DNA known as the “fragmentome” and only requires low-coverage genome sequencing, allowing the technology to be economically affordable in a screening setting [77].

The DELFI methodology finds that patients who were later diagnosed positive for cancer had a wide fluctuation in their fragmentome profiles, while those who had a negative cancer diagnosis had predictable fragmentome profiles. Overall, the technique was able to distinguish more than 90 percent of patients with lung cancer (including those with early stages) and displaying different subtypes [78].

Another study focused on glioblastoma, whose diagnosis is based on resection or biopsy which can be especially arduous and perilous in the case that the tumor mass is located in a deep position. Moreover, tracking cancer progression also necessitates repeated biopsies that are often impracticable. Consequently, there is an urgent requirement to identify biomarkers to diagnose and follow-up glioblastoma evolution by limiting the invasive approaches. Recently, an innovative cancer detection method has been developed based on plasma denaturation profiles obtained by a novel use of differential scanning fluorimetry. By comparing the denaturation profiles of blood samples collected from glioma patients and from healthy subjects, the researchers demonstrated that ML-based algorithms can automatically distinguish the cancer patients from the healthy individuals (with a precision around 92%). Additionally, this high-throughput workflow can be applied to any type of cancer and may represent a potent pan-cancer diagnostic and monitoring tool that requires only a plain blood test [79].

Among the limitations of the current approaches, tissue biopsy presents a fixed overview of the tumor that fails to record the intratumor distinguishment and dynamic changes occurring during carcinogenesis, also determined by clonal pressure caused by the

applied medication [80]. On top of that, it is an invasive procedure, which usually cannot be performed multiple times on request, making this system unfeasible to be conducted as a regular practice for cancer patients' long-term supervision and treatment adjustment. The emergence of liquid biopsy has been a revolutionary development for the current clinical practice, offering great potential to improve the management of ongoing cancer patients for the diagnosis, prognosis, and tailoring of treatment. This approach presents the advantage of being a minimally invasive procedure that utilizes tumor-derived materials obtained from several body fluids, such as peripheral blood, urine, pleural liquid, saliva, or ascites [81]. This solution is not limited by space or time, and it supplies clinically meaningful information related to both primary and metastatic malignant lesions. Among the components of tumor-derived materials that can be analyzed by liquid biopsy, circulating tumor cells, cell-free circulating nucleic acids, and extracellular vesicles are the most extensively studied and characterized cancer markers and are used for various objectives, for instance, the early detection of cancer, staging, prognosis, drug resistance, and minimal residual disease [82].

Another AI approach is the PinPoint test, a cost-effective AI-driven blood test for cancer that is meant to upgrade rapid cancer referral paths. The test is found on an algorithm that uses ML to investigate regular constituents, as well as the patient's age and sex. It can calibrate and combine these individual variables into one solid and highly precise result, such as the likelihood that a patient has cancer [83]. The PinPoint test has been crafted as a decision support tool to give medical professionals the data they need to better sort patients when they initially present with symptoms. Those with high risk can be given precedence for speedy examination in secondary care, while those with the lowest risk can be securely excluded from the "2 week wait" pathway for further discussion with their physicians [84]. This strategy of pinpointing those at the greatest risk for prioritization will promote early detection, contribute to a more dependable pathway, and assist in decreasing post-pandemic delays [85].

## 7. AI imaging in Cancer Diagnosis

In the field of cancer imaging, AI displays a great utility in three main clinical tasks: tumor detection, characterization, and monitoring [86]. The localization of objects of interest in radiographs is referred to as detection, and it is a subset of computer-aided detection (CADe). AI-based detection tools can be used to reduce observational errors and serve as a first line of defense against omission errors [87].

Characterization in general includes tumor segmentation, diagnosis, and staging. It can also include a disease-specific prognosis as well as outcome prediction based on specific treatment modalities. Segmentation determines the extent of abnormalities and can range from simple 2D measurements of the maximum in plane tumor diameter to more involved volumetric segmentations that assess the entire tumor as well as any surrounding tissues. This information could be exploited for future diagnostic purposes as well as for calculating the appropriate dose administration during radiation planning. AI has the capability to significantly improve the efficiency, reproducibility, and reliability of tumor measurements through automated segmentation. In computer-aided diagnosis (CADx) systems, systematic processing of quantitative tumor features is used, allowing for more reproducible descriptors. In the case of inconsistencies in interpretation by different human readers, CADx systems have been used to diagnose lung nodules in thin-section CT and prostate lesions in multiparametric MRI [88].

Staging is another aspect of tumor characterization in which tumors are classified into predefined groups based on the size and spread of the tumor mass, thus providing information regarding the expected clinical course and for the decision of the most appropriate treatment strategies [89]. The application of AI-based methods to cancer imaging allows for the estimation of tumor size, shape, morphology, texture, and kinetics. Additionally, the use of dynamic assessment of contrast uptake on MRI enables physicians to characterize the tumor mass in terms of heterogeneity, phenotypes of spatial features and dynamic

characteristics [90]. Another variable taken in consideration from AI-based tools is entropy, a mathematical descriptor of randomness that provides information on how heterogeneous the pattern is within the tumor, thereby describing the heterogeneous pattern of vascular system uptake (contrast uptake) within tumors imaged on contrast-enhanced breast MRI. As demonstrated by the NCI's The Cancer Genome Atlas (TCGA) breast cancer dataset, such analyses could reflect the heterogeneous nature of angiogenesis and treatment susceptibility [91].

DL systems have been used to simultaneously detect and classify prostate lesions. For training convolutional neural networks (CNNs) for prostate cancer diagnosis by MRI, both *de novo* training [92] and the transfer learning of pre-trained models [93] have been successful. The implementation of CNNs models with anatomically aware features has been shown to improve their performance [94,95]. In addition to MRI, AI techniques for prostate cancer classification have shown promising results by integrating ultrasound data, specifically radiofrequency. Again, both traditional ML and DL approaches were used to train classifiers to estimate the grading of prostate cancer by exploiting temporal ultrasound data [96].

## 8. Critical Issues, Challenges, and Limitations

The accuracy and consistency of AI systems are frequently restricted by their training data and the hardware used. We must keep in mind that AI can make mistakes in some situations because its decision-making ability is predictive and probabilistic. As a result, there are no clear regulations or guidelines in place to determine who is legally liable when AI malfunctions occur or causes issues while providing a service. Another factor to take in consideration is that most of the places where the potential of AI in healthcare has been evaluated are basically high-income and resource-driven areas. When used in low-income countries with a shortage of well-trained physicians and oncological specialists, AI-based prediction tools are expected to have a greater impact and increment the success of cancer treatment.

The improvement in the AI interpretation is a crucial step toward mitigating this risk and providing a decision-making rationale. One limitation is represented by the lack of a human verification step in the process unless a physician supervises the AI system. As a result, no one expects AI to entirely replace medical professionals. AI-based precision medicine will be critical for cancer treatment in the future. Living databases will exploit extremely complex models capable of making a personalized therapy selection, estimation of the drug dose, follow-up schedule, and so on. However, the transition from artificial narrow intelligence to artificial general intelligence will result in the automation of all the steps involved in cancer prediction, diagnosis, and treatment.

Despite its numerous benefits, AI presents several challenges and constraints that hinder it from fully functioning in cancer research. Particularly, three layers of complexity must be considered: (i) cancer is a highly heterogeneous organoid-like structure that, at the time of diagnosis, is made up of many different cancer subclones embedded in a stroma (the tumor microenvironment) that itself contributes to cancer progression; (ii) as cancer progresses, tumor evolution leads to increased intratumor heterogeneity so that by the time therapy is started, the targeted cancer may not respond; (iii) cancers with the same molecular and histological signatures behave differently in each single patient because of individual epigenetic and immunological modulations [97–99]. Thus, the final clinical outcome will depend on the complex interplay between the cancer (with its multiple subclones) and the tumor microenvironment (which includes the stroma composition and the inflammatory and immune response), and, finally, the general pathophysiological condition of the patient (e.g., the body mass, the adipose tissue mass, the nutrition status, the psychological status, the immune status, etc.). This poses an important limit to the capability of AI in predicting the therapy efficacy and the prognosis, which once again stresses the fundamental role of the clinician that cannot be substituted by an algorithm.

The new era of innovation brings with it many challenges that should be overcome to drastically improve oncology procedures at several levels. The lack of inclusive and different datasets for training represents a significant obstacle to the widespread adoption of AI algorithms and decision-support systems in cancer care. Most of the powerful AI models require a large sample size to efficiently train the tool. Although there are dimensionality reduction and feature selection methods for addressing these aspects, proper implementation is critical for achieving better and reliable results. The number and type of data annotated influences the constructions of algorithms, and an imbalance in data from patients differing for gender, age, race, nutritional state, lifestyle, and environment will affect AI and ML training. Thus, the lack of sensible data may increase the risk of missed diagnosis. Therefore, experts are fundamental in data curation and data annotation to provide reliable datasets to be used for training AI classifier and predictors models.

In medical data sets, particularly in the case of cancer data, classes are typically distributed unequally. The continuous use of AI- and ML-based tools for diagnosis and treatment decisions can be risky due to distributional shifts, which means that target data may not match the ongoing patient data employed to train the model, resulting in incorrect outputs. Predictions made by AI at the time of diagnosis likely changes during the course of the therapy and the evolution of the disease along with changes in patient's habit (style of life, diet, medications, etc.).

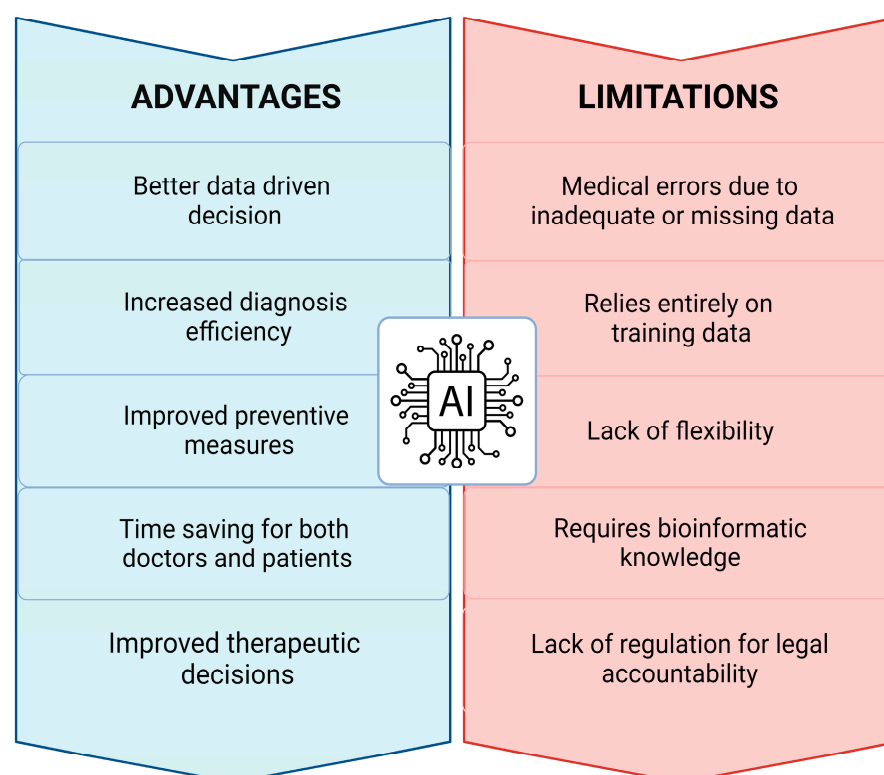
Changes in technology, healthcare, and population, such as the gene pool, are likely to have an impact on the relationship between the data items. The actual application of AI models in clinics is not being actively considered. The predictions achieved with these models frequently require validation in the clinical practice to assist medical experts in confirming diagnosis decisions.

Significant issues regarding data availability and interpretability caused by AI's "black box" process, in parallel with the emergence of an inherent bias toward limited cohorts that reduces the reproducibility of AI models and perpetuates disparities in the healthcare, collectively prevent the widespread application of AI in clinics. Additionally, the distribution of AI-based technologies in many developing countries may be hampered by a lack of knowledge in computing algorithms and technologies of the physicians.

Taken together, the clinically relevant achievements discussed in the present review need to become more solid to be translated into the right treatment for the right patient. Hence, the rapidly ongoing evolution of AI-based medical data analysis will significantly improve the treatments in cancer.

## 9. Conclusions and Perspectives

In this paper, we present an overview of the models applied in diagnosing and identifying therapeutic targets, and we discussed the challenges and future perspectives of AI in cancer research (Figure 3). As the power and potential of AI are increasingly demonstrated, in the coming future several other biomedical fields may exploit the use of AI in their routine clinical practice. AI methodologies' accuracy and predictive power must be significantly improved, as well as demonstrated efficacy comparable to, or better than, human experts in controlled studies [100]. Up to now, AI shows early promising results in the management of several disease conditions, but more efforts in prospective trials and in the education of physicians, technologists, and physicists are needed before it can be widely used. Although there will always be a "black box" for human experts to view AI-generated results, data visualization tools are becoming more widely available to provide some visual understanding of how algorithms make decisions [101]. It is to be stressed that AI is meant to complement the medical doctor facilitating his work, but it will not replace the medical doctor.



**Figure 3.** Advantages and limitations of AI. The scheme summarizes the main benefits along with the current concerns related to the use of AI in the clinical practice.

**Author Contributions:** Conceptualization, A.F. (Alaa Fawaz) and C.I.; writing—original draft preparation, A.F. (Alaa Fawaz) and A.F. (Alessandra Ferraresi); writing—review and editing, C.I.; visualization, A.F. (Alaa Fawaz) and A.F. (Alessandra Ferraresi). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** A.F. (Alessandra Ferraresi) is the recipient of a post-doctoral fellowship from Fondazione Umberto Veronesi (FUV 2023).

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

Artificial Intelligence, AI; Bayesian Efficient Multiple Kernel Learning, BEMKL; computed tomography, CT; computer-aided detection, CADE; computer-aided diagnosis, CADx; convolutional neural networks, CNNs; deep learning, DL; deep neural network, DNN; gene expression modeling, GEM; graph convolutional network, GCN; graph neural networks, GNNs; Human Metabolome Database, HMDB; Human Protein Atlas, (HPA); Leave-One-Out Cross-Validation, LOOCV; machine learning, ML; magnetic resonance imaging, MRI; nano differential scanning fluorimetry, Nanodsf; next-generation sequencing, NGS; positron emission tomography, PET; partial least squares-discriminant analysis, PLS-DA; single nucleotide polymorphisms, SNPs; single-cell RNA sequencing, scRNA-seq; Support Vector Machine, SVM; The Cancer Genome Atlas, TCGA; ultrasound imaging, US.



## References

1. Committee on Diagnostic Error in Health Care; Board on Health Care Services; Institute of Medicine; The National Academies of Sciences, Engineering, and Medicine. *Improving Diagnosis in Health Care*; Balogh, E.P., Miller, B.T., Ball, J.R., Eds.; National Academies Press: Washington, DC, USA, 2015. Available online: <http://www.ncbi.nlm.nih.gov/books/NBK338596/> (accessed on 3 May 2023).
2. Rodziewicz, T.L.; Houseman, B.; Hipskind, J.E. Medical Error Reduction and Prevention. In *StatPearls*; StatPearls Publishing: St. Petersburg, FL, USA, 2023. Available online: <http://www.ncbi.nlm.nih.gov/books/NBK499956/> (accessed on 3 May 2023).
3. Taylor, N. Duke Report Identifies Barriers to Adoption of AI Healthcare Systems. MedTech Dive. Available online: <https://www.medtechdive.com/news/duke-report-identifies-barriers-to-adoption-of-ai-healthcare-systems/546739/> (accessed on 3 May 2023).
4. Bray, F.; Laversanne, M.; Weiderpass, E.; Soerjomataram, I. The Ever-Increasing Importance of Cancer as a Leading Cause of Premature Death Worldwide. *Cancer* **2021**, *127*, 3029–3030. [\[CrossRef\]](#)
5. Ponomarenko, E.A.; Poverennaya, E.V.; Ilgisonis, E.V.; Pyatnitskiy, M.A.; Kopylov, A.T.; Zgoda, V.G.; Lisitsa, A.V.; Archakov, A.I. The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* **2016**, *2016*, 7436849. [\[CrossRef\]](#)
6. Nadhan, R.; Kashyap, S.; Ha, J.H.; Jayaraman, M.; Song, Y.S.; Isidoro, C.; Dhanasekaran, D.N. Targeting Oncometabolites in Peritoneal Cancers: Preclinical Insights and Therapeutic Strategies. *Metabolites* **2023**, *13*, 618. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Hasin, Y.; Seldin, M.; Lusis, A. Multi-Omics Approaches to Disease. *Genome Biol.* **2017**, *18*, 83. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Perkins, D.O.; Jeffries, C.; Sullivan, P. Expanding the ‘Central Dogma’: The Regulatory Role of Nonprotein Coding Genes and Implications for the Genetic Liability to Schizophrenia. *Mol. Psychiatry* **2005**, *10*, 69–78. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Tsakiroglou, M.; Evans, A.; Pirmohamed, M. Leveraging Transcriptomics for Precision Diagnosis: Lessons Learned from Cancer and Sepsis. *Front. Genet.* **2023**, *14*, 1100352. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Haga, Y.; Minegishi, Y.; Ueda, K. Frontiers in Mass Spectrometry–Based Clinical Proteomics for Cancer Diagnosis and Treatment. *Cancer Sci.* **2023**, *114*, 1783–1791. [\[CrossRef\]](#)
11. Janes, K.A.; Yaffe, M.B. Data-Driven Modelling of Signal-Transduction Networks. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 820–828. [\[CrossRef\]](#)
12. Luo, J.; Pan, M.; Mo, K.; Mao, Y.; Zou, D. Emerging Role of Artificial Intelligence in Diagnosis, Classification and Clinical Management of Glioma. *Semin. Cancer Biol.* **2023**, *91*, 110–123. [\[CrossRef\]](#)
13. Wang, S.; Wang, S.; Wang, Z. A Survey on Multi-Omics-Based Cancer Diagnosis Using Machine Learning with the Potential Application in Gastrointestinal Cancer. *Front. Med.* **2023**, *9*, 1109365. [\[CrossRef\]](#)
14. Liao, J.; Li, X.; Gan, Y.; Han, S.; Rong, P.; Wang, W.; Li, W.; Zhou, L. Artificial Intelligence Assists Precision Medicine in Cancer Treatment. *Front. Oncol.* **2023**, *12*, 998222. [\[CrossRef\]](#) [\[PubMed\]](#)
15. He, X.; Liu, X.; Zuo, F.; Shi, H.; Jing, J. Artificial Intelligence-Based Multi-Omics Analysis Fuels Cancer Precision Medicine. *Semin. Cancer Biol.* **2023**, *88*, 187–200. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Dembrower, K.; Wählin, E.; Liu, Y.; Salim, M.; Smith, K.; Lindholm, P.; Eklund, M.; Strand, F. Effect of Artificial Intelligence-Based Triaging of Breast Cancer Screening Mammograms on Cancer Detection and Radiologist Workload: A Retrospective Simulation Study. *Lancet Digit. Health* **2020**, *2*, e468–e474. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Future Healthc. J.* **2019**, *6*, 94–98. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Bohr, A.; Memarzadeh, K. The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in Healthcare*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 25–60. [\[CrossRef\]](#)
19. Venkatesan, D.; Elangovan, A.; Winster, H.; Pasha, M.Y.; Abraham, K.S.; Satheshkumar, J.; Sivaprakash, P.; Niraikulam, A.; Gopalakrishnan, A.V.; Narayanasamy, A.; et al. Diagnostic and therapeutic approach of artificial intelligence in neuro-oncological diseases. *Biosens. Bioelectron. X* **2022**, *11*, 100188. [\[CrossRef\]](#)
20. Swanson, K.; Wu, E.; Zhang, A.; Alizadeh, A.A.; Zou, J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell* **2023**, *186*, 1772–1791. [\[CrossRef\]](#)
21. Mohammed, A.; Biegert, G.; Adamec, J.; Helikar, T. Identification of Potential Tissue-Specific Cancer Biomarkers and Development of Cancer versus Normal Genomic Classifiers. *Oncotarget* **2017**, *8*, 85692–85715. [\[CrossRef\]](#)
22. Zhang, Y.; Xiong, S.; Wang, Z.; Liu, Y.; Luo, H.; Li, B.; Zou, Q. Local Augmented Graph Neural Network for Multi-Omics Cancer Prognosis Prediction and Analysis. *Methods* **2023**, *213*, 1–9. [\[CrossRef\]](#)
23. Peng, X.; Chen, Z.; Farshidfar, F.; Xu, X.; Lorenzi, P.L.; Wang, Y.; Cheng, F.; Tan, L.; Mojumdar, K.; Du, D.; et al. Molecular Characterization and Clinical Relevance of Metabolic Expression Subtypes in Human Cancers. *Cell Rep.* **2018**, *23*, 255–269.e4. [\[CrossRef\]](#)
24. Yokota, K.; Uchida, H.; Sakairi, M.; Abe, M.; Tanaka, Y.; Tainaka, T.; Shiota, C.; Sumida, W.; Oshima, K.; Makita, S.; et al. Identification of Novel Neuroblastoma Biomarkers in Urine Samples. *Sci. Rep.* **2021**, *11*, 4055. [\[CrossRef\]](#)
25. Barker, M.; Rayens, W. Partial Least Squares for Discrimination. *J. Chemom.* **2003**, *17*, 166–173. [\[CrossRef\]](#)
26. Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.-A. MixOmics: An R Package for ‘omics Feature Selection and Multiple Data Integration. *PLoS Comput. Biol.* **2017**, *13*, e1005752. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Westerhuis, J.A.; Hoefsloot, H.C.J.; Smit, S.; Vis, D.J.; Smilde, A.K.; van Velzen, E.J.J.; van Duijnhoven, J.P.M.; van Dorsten, F.A. Assessment of PLS-DA Cross Validation. *Metabolomics* **2008**, *4*, 81–89. [\[CrossRef\]](#)

28. Brereton, R.G.; Lloyd, G.R. Partial Least Squares Discriminant Analysis: Taking the Magic Away: PLS-DA: Taking the Magic Away. *J. Chemom.* **2014**, *28*, 213–225. [\[CrossRef\]](#)
29. Gromski, P.S.; Muhamadali, H.; Ellis, D.I.; Xu, Y.; Correa, E.; Turner, M.L.; Goodacre, R. A Tutorial Review: Metabolomics and Partial Least Squares-Discriminant Analysis—A Marriage of Convenience or a Shotgun Wedding. *Anal. Chim. Acta* **2015**, *879*, 10–23. [\[CrossRef\]](#)
30. Gu, C.; Kim, G.B.; Kim, W.J.; Kim, H.U.; Lee, S.Y. Current Status and Applications of Genome-Scale Metabolic Models. *Genome Biol.* **2019**, *20*, 121. [\[CrossRef\]](#)
31. Fang, X.; Lloyd, C.J.; Palsson, B.O. Reconstructing Organisms in Silico: Genome-Scale Models and Their Emerging Applications. *Nat. Rev. Microbiol.* **2020**, *18*, 731–743. [\[CrossRef\]](#)
32. Thiele, I.; Palsson, B.O. A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction. *Nat. Protoc.* **2010**, *5*, 93–121. [\[CrossRef\]](#)
33. O'Brien, J.E.; Monk, J.M.; Palsson, B.O. Using Genome-Scale Models to Predict Biological Capabilities. *Cell* **2015**, *161*, 971–987. [\[CrossRef\]](#)
34. Chand, S. A comparative study of breast cancer tumor classification by classical machine learning methods and deep learning method. *Mach. Vis. Appl.* **2020**, *31*, e270.
35. Angermueller, C.; Pärnamaa, T.; Parts, L.; Stegle, O. Deep Learning for Computational Biology. *Mol. Syst. Biol.* **2016**, *12*, 878. [\[CrossRef\]](#)
36. Samiei, M.; Würfl, T.; Deleu, T.; Weiss, M.; Dutil, F.; Fevens, T.; Boucher, G.; Lemieux, S.; Cohen, J.P. The TCGA Meta-Dataset Clinical Benchmark. *arXiv* **2019**, arXiv:1910.08636.
37. Jin, S.; Zeng, X.; Xia, F.; Huang, W.; Liu, X. Application of Deep Learning Methods in Biological Networks. *Brief. Bioinform.* **2021**, *22*, 1902–1917. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Rhee, S.; Seo, S.; Kim, S. Hybrid Approach of Relation Network and Localized Graph Convolutional Filtering for Breast Cancer Subtype Classification. *arXiv* **2018**, arXiv:1711.05859.
39. Lee, S.; Lim, S.; Lee, T.; Sung, I.; Kim, S. Cancer Subtype Classification and Modeling by Pathway Attention and Propagation. *Bioinformatics* **2020**, *36*, 3818–3824. [\[CrossRef\]](#)
40. Dai, H.; Li, H.; Tian, T.; Huang, X.; Wang, L.; Zhu, J.; Song, L. Adversarial Attack on Graph Structured Data. In Proceedings of the 35th International Conference on Machine Learning (ICML 2018), Stockholm, Sweden, 10–15 July 2018. Available online: <https://proceedings.mlr.press/v80/dai18b.html> (accessed on 3 May 2023).
41. Zhang, X.; Zitnik, M. GNNGuard: Defending Graph Neural Networks against Adversarial Attacks. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2020; Volume 33, pp. 9263–9275. Available online: <https://papers.nips.cc/paper/2020/hash/690d83983a63aa1818423fd6edd3bfdb-Abstract.html> (accessed on 3 May 2023).
42. Abdelaal, T.; Michielsen, L.; Cats, D.; Hoogduin, D.; Mei, H.; Reinders, M.J.T.; Mahfouz, A. A Comparison of Automatic Cell Identification Methods for Single-Cell RNA Sequencing Data. *Genome Biol.* **2019**, *20*, 194. [\[CrossRef\]](#)
43. Tan, Y.; Cahan, P. SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species. *Cell Syst.* **2019**, *9*, 207–213.e2. [\[CrossRef\]](#)
44. Hu, J.; Li, X.; Hu, G.; Lyu, Y.; Susztak, K.; Li, M. Iterative Transfer Learning with Neural Network for Clustering and Cell Type Classification in Single-Cell RNA-Seq Analysis. *Nat. Mach. Intell.* **2020**, *2*, 607–618. [\[CrossRef\]](#)
45. Andreatta, M.; Corria-Osorio, J.; Müller, S.; Cubas, R.; Coukos, G.; Carmona, S.J. Interpretation of T Cell States from Single-Cell Transcriptomics Data Using Reference Atlases. *Nat. Commun.* **2021**, *12*, 2965. [\[CrossRef\]](#)
46. Michielsen, L.; Reinders, M.J.T.; Mahfouz, A. Hierarchical Progressive Learning of Cell Identities in Single-Cell Data. *Nat. Commun.* **2021**, *12*, 2799. [\[CrossRef\]](#)
47. Ranjan, B.; Schmidt, F.; Sun, W.; Park, J.; Honardoost, M.A.; Tan, J.; Rayan, N.A.; Prabhakar, S. ScConsensus: Combining Supervised and Unsupervised Clustering for Cell Type Identification in Single-Cell RNA Sequencing Data. *BMC Bioinform.* **2021**, *22*, 186. [\[CrossRef\]](#)
48. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.E.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal. *Sci. Signal.* **2013**, *6*, pl1. [\[CrossRef\]](#)
49. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, D.R.; Kibbe, W.A.; Staudt, L.M. Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **2016**, *375*, 1109–1112. [\[CrossRef\]](#)
50. Goldman, M.J.; Craft, B.; Hastie, M.; Repčeka, K.; McDade, F.; Kamath, A.; Banerjee, A.; Luo, Y.; Rogers, D.; Brooks, A.N.; et al. Visualizing and Interpreting Cancer Genomics Data via the Xena Platform. *Nat. Biotechnol.* **2020**, *38*, 675–678. [\[CrossRef\]](#)
51. Manzoni, C.; Kia, D.A.; Vandrovicova, J.; Hardy, J.; Wood, N.W.; Lewis, P.A.; Ferrari, R. Genome, Transcriptome and Proteome: The Rise of Omics Data and Their Integration in Biomedical Sciences. *Brief. Bioinform.* **2018**, *19*, 286–302. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Creixell, P.; Reimand, J.; Haider, S.; Wu, G.; Shibata, T.; Vazquez, M.; Mustonen, V.; Gonzalez-Perez, A.; Pearson, J.; Sander, C.; et al. Pathway and Network Analysis of Cancer Genomes. *Nat. Methods* **2015**, *12*, 615–621. [\[PubMed\]](#)
53. Ngiam, K.Y.; Khor, I.W. Big Data and Machine Learning Algorithms for Health-Care Delivery. *Lancet Oncol.* **2019**, *20*, e262–e273. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Reyna, M.A.; Haan, D.; Paczkowska, M.; Verbeke, L.P.C.; Vazquez, M.; Kahraman, A.; Pulido-Tamayo, S.; Barenboim, J.; Wadi, L.; Dhingra, P.; et al. Pathway and Network Analysis of More than 2500 Whole Cancer Genomes. *Nat. Commun.* **2020**, *11*, 729. [\[CrossRef\]](#)

55. Luo, P.; Ding, Y.; Lei, X.; Wu, F.-X. DeepDriver: Predicting Cancer Driver Genes Based on Somatic Mutations Using Deep Convolutional Neural Networks. *Front. Genet.* **2019**, *10*, 13. [\[CrossRef\]](#)
56. Jiao, W.; Atwal, G.; Polak, P.; Karlic, R.; Cuppen, E.; PCAWG Tumor Subtypes and Clinical Translation Working Group; Danyi, A.; de Ridder, J.; van Herpen, C.; Lolkema, M.P.; et al. A Deep Learning System Accurately Classifies Primary and Metastatic Cancers Using Passenger Mutation Patterns. *Nat. Commun.* **2020**, *11*, 728. [\[CrossRef\]](#)
57. Chaudhary, K.; Poirion, O.B.; Lu, L.; Garmire, L.X. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2018**, *24*, 1248–1259. [\[CrossRef\]](#)
58. Gao, F.; Wang, W.; Tan, M.; Zhu, L.; Zhang, Y.; Fessler, E.; Vermeulen, L.; Wang, X. DeepCC: A Novel Deep Learning-Based Framework for Cancer Molecular Subtype Classification. *Oncogenesis* **2019**, *8*, 44. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Zeng, X.; Zhu, S.; Liu, X.; Zhou, Y.; Nussinov, R.; Cheng, F. DeepDR: A Network-Based Deep Learning Approach to in Silico Drug Repositioning. *Bioinformatics* **2019**, *35*, 5191–5198. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Issa, N.T.; Stathias, V.; Schürer, S.; Dakshanamurthy, S. Machine and Deep Learning Approaches for Cancer Drug Repurposing. *Semin. Cancer Biol.* **2021**, *68*, 132–142. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Park, Y.; Heider, D.; Hauschild, A.-C. Integrative Analysis of Next-Generation Sequencing for Next-Generation Cancer Research toward Artificial Intelligence. *Cancers* **2021**, *13*, 3148. [\[CrossRef\]](#)
62. Zhang, L.; Lv, C.; Jin, Y.; Cheng, G.; Fu, Y.; Yuan, D.; Tao, Y.; Guo, Y.; Ni, X.; Shi, T. Deep Learning-Based Multi-Omics Data Integration Reveals Two Prognostic Subtypes in High-Risk Neuroblastoma. *Front. Genet.* **2018**, *9*, 477. [\[CrossRef\]](#)
63. Francescato, M.; Chierici, M.; Dezfooli, S.R.; Zandonà, A.; Jurman, G.; Furlanello, C. Multi-Omics Integration for Neuroblastoma Clinical Endpoint Prediction. *Biol. Direct* **2018**, *13*, 5. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Huang, Z.; Zhan, X.; Xiang, S.; Johnson, T.S.; Helm, B.; Yu, C.Y.; Zhang, J.; Salama, P.; Rizkalla, M.; Han, Z.; et al. SALMON: Survival Analysis Learning with Multi-Omics Neural Networks on Breast Cancer. *Front. Genet.* **2019**, *10*, 166. [\[CrossRef\]](#)
65. Xie, G.; Dong, C.; Kong, Y.; Zhong, J.F.; Li, M.; Wang, K. Group Lasso Regularized Deep Learning for Cancer Prognosis from Multi-Omics and Clinical Features. *Genes* **2019**, *10*, 240. [\[CrossRef\]](#)
66. Chen, L.; Wu, J. Bio-Network Medicine. *J. Mol. Cell Biol.* **2015**, *7*, 185–186. [\[CrossRef\]](#)
67. Song, H.; Chen, L.; Cui, Y.; Li, Q.; Wang, Q.; Fan, J.; Yang, J.; Zhang, L. Denoising of MR and CT Images Using Cascaded Multi-Supervision Convolutional Neural Networks with Progressive Training. *Neurocomputing* **2022**, *469*, 354–365. [\[CrossRef\]](#)
68. Zhang, L.; Zhang, L.; Guo, Y.; Xiao, M.; Feng, L.; Yang, C.; Wang, G.; Ouyang, L. MCDB: A Comprehensive Curated Mitotic Catastrophe Database for Retrieval, Protein Sequence Alignment, and Target Prediction. *Acta Pharm. Sin. B* **2021**, *11*, 3092–3104. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Zhou, Y.; Wang, F.; Tang, J.; Nussinov, R.; Cheng, F. Artificial Intelligence in COVID-19 Drug Repurposing. *Lancet Digit. Health* **2020**, *2*, e667–e676. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Suhail, Y.; Cain, M.P.; Vanaja, K.; Kurywach, P.A.; Levchenko, A.; Kalluri, R.; Kshitiz. Systems Biology of Cancer Metastasis. *Cell Syst.* **2019**, *9*, 109–127. [\[CrossRef\]](#)
71. Barabási, A.-L.; Oltvai, Z.N. Network Biology: Understanding the Cell's Functional Organization. *Nat. Rev. Genet.* **2004**, *5*, 101–113. [\[CrossRef\]](#)
72. Ali, M.; Khan, S.A.; Wennerberg, K.; Aittokallio, T. Global Proteomics Profiling Improves Drug Sensitivity Prediction: Results from a Multi-Omics, Pan-Cancer Modeling Approach. *Bioinformatics* **2018**, *34*, 1353–1362. [\[CrossRef\]](#)
73. Sharifi-Noghabi, H.; Zolotareva, O.; Collins, C.C.; Ester, M. MOLI: Multi-Omics Late Integration with Deep Neural Networks for Drug Response Prediction. *Bioinformatics* **2019**, *35*, i501–i509. [\[CrossRef\]](#)
74. Kwon, M.-S.; Kim, Y.; Lee, S.; Namkung, J.; Yun, T.; Yi, S.G.; Han, S.; Kang, M.; Kim, S.W.; Jang, J.-Y.; et al. Integrative Analysis of Multi-Omics Data for Identifying Multi-Markers for Diagnosing Pancreatic Cancer. *BMC Genom.* **2015**, *16*, S4. [\[CrossRef\]](#)
75. Gautam, P.; Jaiswal, A.; Aittokallio, T.; Al-Ali, H.; Wennerberg, K. Phenotypic Screening Combined with Machine Learning for Efficient Identification of Breast Cancer-Selective Therapeutic Targets. *Cell Chem. Biol.* **2019**, *26*, 970–979.e4. [\[CrossRef\]](#)
76. Peng, C.; Zheng, Y.; Huang, D.-S. Capsule Network Based Modeling of Multi-Omics Data for Discovery of Breast Cancer-Related Genes. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *17*, 1605–1612. [\[CrossRef\]](#)
77. Mazzone, P.J.; Sears, C.R.; Arenberg, D.A.; Gaga, M.; Gould, M.K.; Massion, P.P.; Nair, V.S.; Powell, C.A.; Silvestri, G.A.; Vachani, A.; et al. Evaluating Molecular Biomarkers for the Early Detection of Lung Cancer: When Is a Biomarker Ready for Clinical Use? An Official American Thoracic Society Policy Statement. *Am. J. Respir. Crit. Care Med.* **2017**, *196*, e15–e29. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Seijo, L.M.; Peled, N.; Ajona, D.; Boeri, M.; Field, J.K.; Sozzi, G.; Pio, R.; Zulueta, J.J.; Spira, A.; Massion, P.P.; et al. Biomarkers in Lung Cancer Screening: Achievements, Promises, and Challenges. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **2019**, *14*, 343–357. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Tsvetkov, P.O.; Eyraud, R.; Ayache, S.; Bougaev, A.A.; Malesinski, S.; Benazha, H.; Gorokhova, S.; Buffat, C.; Dehais, C.; Sanson, M.; et al. An AI-Powered Blood Test to Detect Cancer Using NanoDSF. *Cancers* **2021**, *13*, 1294. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Parikh, A.R.; Leshchiner, I.; Elagina, L.; Goyal, L.; Levovitz, C.; Siravegna, G.; Livitz, D.; Rhrissorakrai, K.; Martin, E.E.; Van Seventer, E.E.; et al. Liquid versus Tissue Biopsy for Detecting Acquired Resistance and Tumor Heterogeneity in Gastrointestinal Cancers. *Nat. Med.* **2019**, *25*, 1415–1421. [\[CrossRef\]](#)
81. Lu, T.; Li, J. Clinical Applications of Urinary Cell-Free DNA in Cancer: Current Insights and Promising Future. *Am. J. Cancer Res.* **2017**, *7*, 2318–2332. [\[PubMed\]](#)



82. Heitzer, E.; Haque, I.S.; Roberts, C.E.S.; Speicher, M.R. Current and Future Perspectives of Liquid Biopsies in Genomics-Driven Oncology. *Nat. Rev. Genet.* **2019**, *20*, 71–88. [\[CrossRef\]](#)
83. Savage, R.; Messenger, M.; Neal, R.D.; Ferguson, R.; Johnston, C.; Lloyd, K.L.; Neal, M.D.; Sansom, N.; Selby, P.; Sharma, N.; et al. Development and Validation of Multivariable Machine Learning Algorithms to Predict Risk of Cancer in Symptomatic Patients Referred Urgently from Primary Care: A Diagnostic Accuracy Study. *BMJ Open* **2022**, *12*, e053590. [\[CrossRef\]](#)
84. Cohen, J.D.; Li, L.; Wang, Y.; Thoburn, C.; Afsari, B.; Danilova, L.; Douville, C.; Javed, A.A.; Wong, F.; Mattox, A.; et al. Detection and Localization of Surgically Resectable Cancers with a Multi-Analyte Blood Test. *Science* **2018**, *359*, 926–930. [\[CrossRef\]](#)
85. Cree, I.A.; Uttley, L.; Woods, H.B.; Kikuchi, H.; Reiman, A.; Harnan, S.; Whiteman, B.L.; Philips, S.T.; Messenger, M.; Cox, A.; et al. The Evidence Base for Circulating Tumour DNA Blood-Based Biomarkers for the Early Detection of Cancer: A Systematic Mapping Review. *BMC Cancer* **2017**, *17*, 697. [\[CrossRef\]](#)
86. Aerts, H.J.W.L.; Velazquez, E.R.; Leijenaar, R.T.H.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat. Commun.* **2014**, *5*, 4006. [\[CrossRef\]](#)
87. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J.W.L. Artificial Intelligence in Radiology. *Nat. Rev. Cancer* **2018**, *18*, 500–510. [\[CrossRef\]](#)
88. Chan, H.-P.; Hadjiiski, L.; Zhou, C.; Sahiner, B. Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography—a review. *Acad. Radiol.* **2008**, *15*, 535–555. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Rasch, C.; Barillot, I.; Remeijer, P.; Touw, A.; van Herk, M.; Lebesque, J.V. Definition of the Prostate in CT and MRI: A Multi-Observer Study. *Int. J. Radiat. Oncol. Biol. Phys.* **1999**, *43*, 57–66. [\[CrossRef\]](#)
90. Chen, W.; Giger, M.L.; Li, H.; Bick, U.; Newstead, G.M. Volumetric Texture Analysis of Breast Lesions on Contrast-Enhanced Magnetic Resonance Images. *Magn. Reson. Med.* **2007**, *58*, 562–571. [\[CrossRef\]](#) [\[PubMed\]](#)
91. Zhu, B.; Song, N.; Shen, R.; Arora, A.; Machiela, M.J.; Song, L.; Landi, M.T.; Ghosh, D.; Chatterjee, N.; Baladandayuthapani, V.; et al. Integrating Clinical and Multiple Omics Data for Prognostic Assessment across Human Cancers. *Sci. Rep.* **2017**, *7*, 16954. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Liu, S.; Zheng, H.; Feng, Y.; Li, W. Prostate Cancer Diagnosis Using Deep Learning with 3D Multiparametric MRI. *arXiv* **2017**, arXiv:1703.04078.
93. Chen, Q.; Xu, X.; Hu, S.; Li, X.; Zou, Q.; Li, Y. A Transfer Learning Approach for Classification of Clinical Significant Prostate Cancers from MpMRI Scans. *Proc. SPIE* **2017**, *10134*, 101344F.
94. Seah, J.C.Y.; Tang, J.S.N.; Kitchen, A. Detection of Prostate Cancer on Multiparametric MRI. In *Medical Imaging 2017: Computer-Aided Diagnosis*; Armato, S.G., Petrick, N.A., Eds.; SPIE: Orlando, FL, USA, 2017; Volume 10134, p. 1013429.
95. Mehrtash, A.; Sedghi, A.; Ghafoorian, M.; Taghipour, M.; Tempany, C.M.; Wells, W.M.; Kapur, T.; Mousavi, P.; Abolmaesumi, P.; Fedorov, A. Classification of Clinical Significance of MRI Prostate Findings Using 3D Convolutional Neural Networks. *Proc. SPIE Int. Soc. Opt. Eng.* **2017**, *10134*, 101342A.
96. Azizi, S.; Bayat, S.; Yan, P.; Tahmasebi, A.; Nir, G.; Kwak, J.T.; Xu, S.; Wilson, S.; Iczkowski, K.A.; Lucia, M.S.; et al. Detection and Grading of Prostate Cancer Using Temporal Enhanced Ultrasound: Combining Deep Neural Networks and Tissue Mimicking Simulations. *Int. J. Comput. Assist. Radiol. Surg.* **2017**, *12*, 1293–1305. [\[CrossRef\]](#)
97. Al Bakir, M.; Huebner, A.; Martínez-Ruiz, C.; Grigoriadis, K.; Watkins, T.B.K.; Pich, O.; Moore, D.A.; Veeriah, S.; Ward, S.; Laycock, J.; et al. The Evolution of Non-Small Cell Lung Cancer Metastases in TRACERx. *Nature* **2023**, *616*, 534–542. [\[CrossRef\]](#)
98. Martínez-Ruiz, C.; Black, J.R.M.; Puttick, C.; Hill, M.S.; Demeulemeester, J.; Cadieux, E.L.; Thol, K.; Jones, T.P.; Veeriah, S.; Naceur-Lombardelli, C.; et al. Genomic–Transcriptomic Evolution in Lung Cancer and Metastasis. *Nature* **2023**, *616*, 543–552. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Chen, C.; Wang, Z.; Ding, Y.; Wang, L.; Wang, S.; Wang, H.; Qin, Y. DNA Methylation: From Cancer Biology to Clinical Perspectives. *Front. Biosci. Landmark* **2022**, *27*, 326. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Olaronke, I.; Oluwaseun, O. Big Data in Healthcare: Prospects, Challenges and Resolutions. In Proceedings of the 2016 Future Technologies Conference (FTC), San Francisco, CA, USA, 6–7 December 2016; pp. 1152–1157.
101. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* **2017**, *542*, 115–118. [\[CrossRef\]](#) [\[PubMed\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.