

Review

Not peer-reviewed version

Systems Biology in Cancer Diagnosis Integrating Omics Technologies and Artificial Intelligence to support physician decision-making

Alaa Fawaz , [Alessandra Ferraresi](#) , [Ciro Isidoro](#) *

Posted Date: 19 October 2023

doi: [10.20944/preprints202310.1262.v1](https://doi.org/10.20944/preprints202310.1262.v1)

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



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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.

Review

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

Alaa Fawaz ¹, Alessandra Ferraresi ¹ and Ciro Isidoro ^{1,*}

¹ Laboratory of Molecular Pathology, Department of Health Sciences, Università del Piemonte Orientale, Novara, Italy; A.Fa. 20041632@studenti.uniupo.it; A.Fe. alessandra.ferraresi@med.uniupo.it; C.I. ciro.isidoro@med.uniupo.it

* 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 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, histopathology of biopsy, and imaging (MRI, CT, PET, 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.

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

1. Introduction

According to a recent review, mistakes in diagnosis account for 60% of all medical errors and an estimated 40,000 to 80,000 deaths in U.S. medical centers each year. Typically, clinicians have limited time to make decisions based on the interpretation of huge amounts of data, and this increases the risk of underestimating (or sometimes overestimating) some data. Other factors of potential biases affecting the accuracy of the diagnosis are personal experience and medical specialty.

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 [1].

Cancer is the leading cause of death in people, accounting for an estimated 10 million deaths by 2020 [2]. It is a complex disease resulting from anomalies in physiological processes involving genes, coding and non-coding RNAs, proteins, metabolites, and other biomolecules [3,4]. 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 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 [5]. 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.

Current diagnostic practice includes histopathology imaging and a range of blood tests which are instrumental in the initial diagnosis and for determining cancer staging; however, these approaches display some limitations in dissecting the molecular basis beyond the onset of the disorder. To overcome these aspects, molecular and omics technologies can provide a genetic, epigenetic, and metabolic profile of the tumor [6,7].

Despite difficulties in providing individualized and data-driven care, advancements in screening, diagnosis, treatment, and survival rate have been remarkable in recent decades [8]. Early detection and prognosis prediction represent two crucial clinical needs for limiting cancer progression. In this context, the proper clinical care for cancer patients can be improved by the introduction of AI in cancer detection, diagnosis, and treatment [9–12].

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 [13].

AI-based technologies applied to oncology aim at improving clinical practice, including but not limited to 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 [14].

Artificial intelligence in medicine at glance

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) translation of genomic information for recognizing genetic variants using high-throughput sequencing technologies, and (iii) prediction and tracking of patient's prognosis [15]. 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 obtained from wearable devices, and (vi) the production and implementation of autonomous robots in medical procedures [16].

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 [17]. 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 [18]. 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 been also 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 the amalgamation of omics data as well as the characterization of diverse tumors [19].

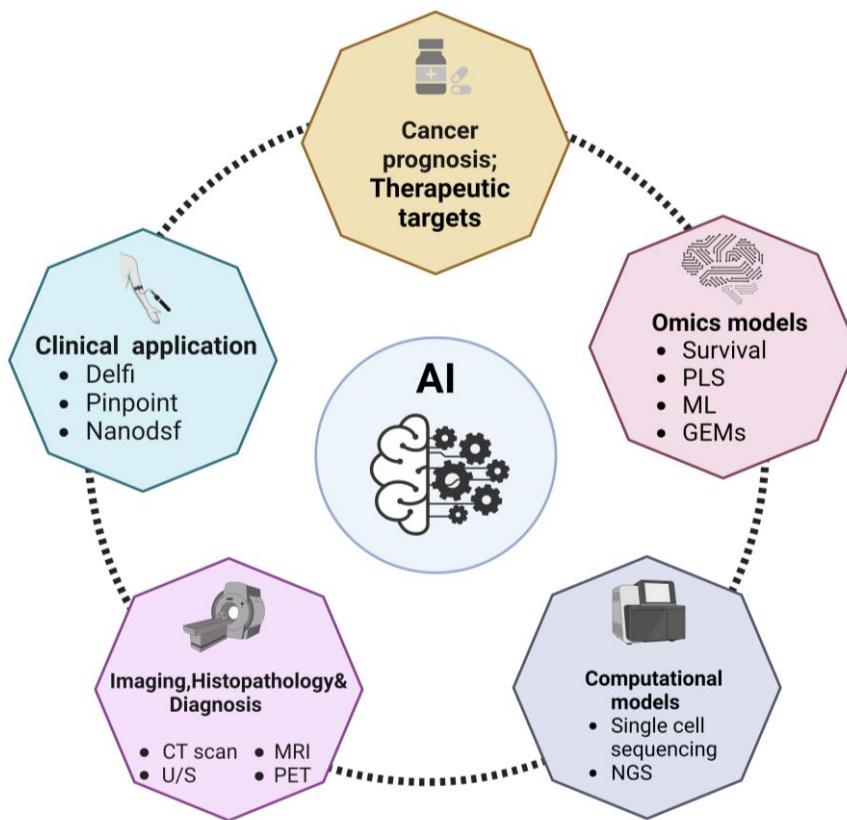


Figure 1. Overview of the applications of AI to cancer diagnosis and oncology research field. The scheme depicts the main fields of application of AI discussed in this review. Abbreviations: computed tomography, CT; gene expression models, GEMs; machine learning, ML; magnetic resonance imaging, MRI;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

2. Omics data for identifying cancer metabolic biomarkers

Omics technologies allow to analyze in depth 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 prediction about prognosis and treatment response in a patient-tailored (personalized) manner [10,12,20].

Survival models

To find cancer metabolic biomarkers, survival models have been used more frequently than partial least squares (PLS) models, machine learning models, and gene expression modelling (GEM) [21]. 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 [22].

In their study, based on gene expression profiles of seven major metabolic pathways, Peng and colleagues identified 30 tumor subtypes in 33 different cancer types 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 glycemic, 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 [21].

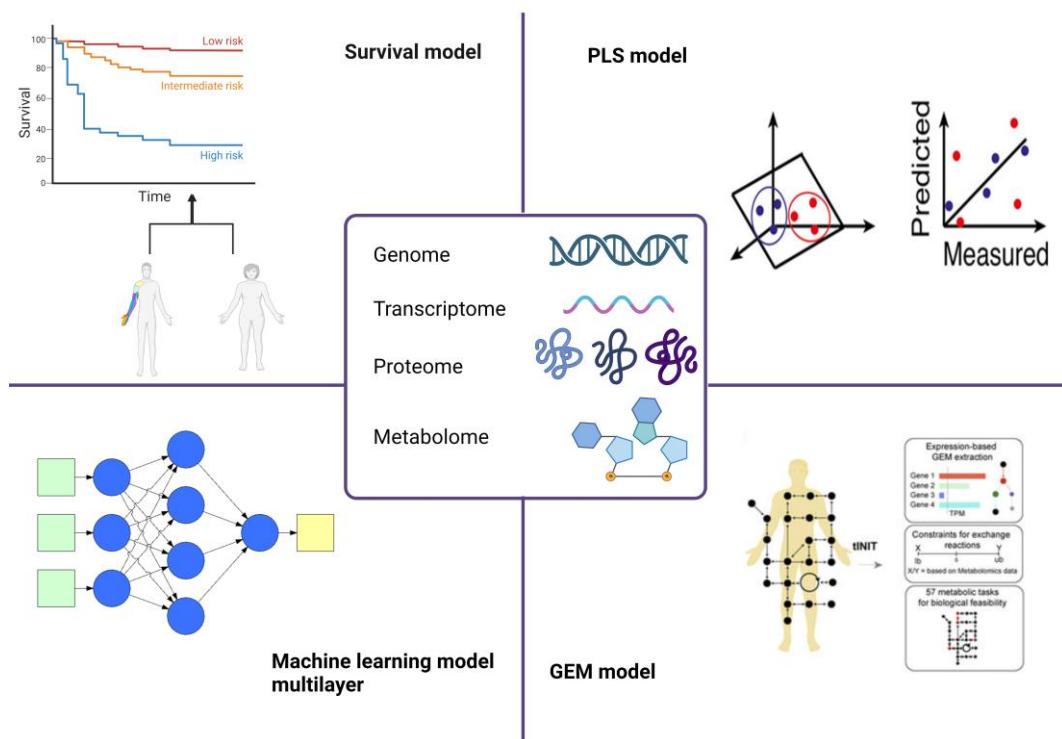


Figure 2. Overview of the omics technologies exploited in cancer diagnosis/prognosis. The scheme depicts the main omics models currently used in biomarker identification. Abbreviations: gene expression modelling, GEM; partial least squares analysis, PLS.

PLS models

Partial least squares regression (PLS) was initially created as a regression model that processes numerous independent variables that are correlated and also produce numerous dependent variables, which many statistical and ML techniques cannot directly handle. PLS models and their variations, particularly PLS-discriminant analysis (PLS-DA) were frequently used for the analysis of omics data with a focus on metabolomics [23]. PLS-DA has been primarily used to extract insights from large datasets of omics data, such as identifying metabolites from metabolome data that are the best at differentiating between cancer cells with various statuses or people with various health conditions. PLS-DA might have an overfitting issue, too, like other data mining techniques, so it needs thorough validation, frequently done through cross-validation [24].

PLS-DA or 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 [25,26]. Among its advantages, PLS-DA allows to analyze 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 [27].

The Materials and Methods should be described with sufficient details to allow others to replicate and build on the published results. Please note that the publication of your manuscript implicates that you must make all materials, data, computer code, and protocols associated with the publication available to readers. Please disclose at the submission stage any restrictions on the availability of materials or information. New methods and protocols should be described in detail while well-established methods can be briefly described and appropriately cited.

Research manuscripts reporting large datasets that are deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, please state that they will be provided during review. They must be provided prior to publication.

Interventionary studies involving animals or humans, and other studies that require ethical approval, must list the authority that provided approval and the corresponding ethical approval code.

Table 1. The table summarizes 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	

Genome-scale metabolic models

Gene expression modeling (GEM) is a computational model based on the law of mass conservation of metabolites and allows predicting metabolic fluxes for entire biochemical reactions taking place inside a cell by using numerical optimization [28,29]. 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 efficiently simulating 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, which allows for building a cell-specific model, which can be especially useful for 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 cultural 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 [30,31].

Table 2. The table summarizes the main advantages and limitations of GEM models.

Advantages	Limitations
Explore metabolism in multiple cell types	Uncertainties in the estimated parameters in 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	

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 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 [32].

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 [33]. 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 [34]. Following that, large-scale public cancer data sets like the TCGA encouraged the widespread use of DNNs in the cancer research [35].

Table 3. The table summarizes 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

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 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 [36]. 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 [37]. 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 [38]. Thus, a robust GNN model is required for precise and stable prediction of cancer subtypes [39].

Table 4. The table summarizes 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 to study 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 a 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 [40]. However, the use of this technology also raises computational difficulties [41]. One of the major challenges in single-cell data analysis is to attribute a cell annotation to each cell analyzed [42]. 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 [43]. This limitation is particularly found in datasets coming from cancer tissues, in which the variability in the transcriptomic states is not conform to traditionally defined cell types [44,45].

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 [46,47]. 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 [48,49].

Several studies have combined NGS data with ML to propose a novel data-driven methodology in systems biology [50]. Several network-based ML models have been implemented to analyze cancer data and aid in the understanding of novel mechanisms in cancer development [51,52]. Furthermore, the use of DNN models for large-scale data analysis enhanced the accuracy of computational models for the prediction of the mutational landscape, molecular subtyping and drug repurposing [53–56]. 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 [57,58].

A common aim of NGS data analysis in cancer research is the identification of potential biomarkers predictive of specific cancer types or subtypes. A variety of bioinformatics tools and ML models, for example, aim to identify a molecular signature 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 [59].

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 [60]. In a separate study, Francescatto et al. employed the integrative network fusion framework together with a ML classifier to distinguish features that could differentiate between distinct outcomes of patients [61].

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 [62]. The use of a DL-based algorithm allows to combine the information from the same gene across different types of omics data, thus resulting in a successful and insightful analysis [63].

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 each other to provide accurate biological explanations [64].

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) [65,66]. 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 [67]. 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 [68], allowing us to better understand carcinogenesis and identify novel anti-cancer targets [69].

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 [70]. 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, duplicate 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 [71].

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 [72]. Moreover, ML-based analysis has been utilized to discover specific anticancer drug targets for breast tumors [73]. 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 [74].

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 the Bayesian Efficient Multiple Kernel Learning (BEMKL) models [70]. 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 technology to be economically affordable in a screening setting [75].

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 [76].

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 and patient comorbidities. 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, it has been developed an innovative cancer detection method 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 [77].

Among the limitations of the current approaches, tissue biopsy presents a fixed overview of the tumor that fails to record the intratumor distinction and dynamic changes occurring during carcinogenesis, also determined by clonal pressure caused by the applied medication [78]. On top of that, it is an invasive procedure, which cannot be usually done 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 diagnosis, prognosis, and tailoring of treatment. This approach presents the advantage to be a minimally invasive procedure that utilizes tumor-derived materials obtained from several body fluids, such as peripheral blood, urine, pleural liquid, saliva, or ascites [79]. 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, early detection of cancer, staging, prognosis, drug resistance, and minimal residual disease [80].

Another AI approach is 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 [81]. 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 [82]. 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 [83].

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 [84]. 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 [85].

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 [86].

Staging is another aspect of tumor characterization in which tumors are classified into predefined groups based on size and spread of the tumor mass, thus providing information regarding the expected clinical course and for the decision of the most appropriated treatment strategies [87]. 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 [88]. 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 [89].

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 [90] and transfer learning of pre-trained models [91] have been successful. The implementation of CNNs models with anatomically aware features has been shown to improve their performance [92,93]. 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 [94].

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 [Garavaglia, Vallino et al., 2023 Book chapter, in press; [95–97]. 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. In medical data sets, particularly in the case of cancer data, classes are typically distributed unequally. 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. 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 to be validated 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 prevented 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 for being 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.

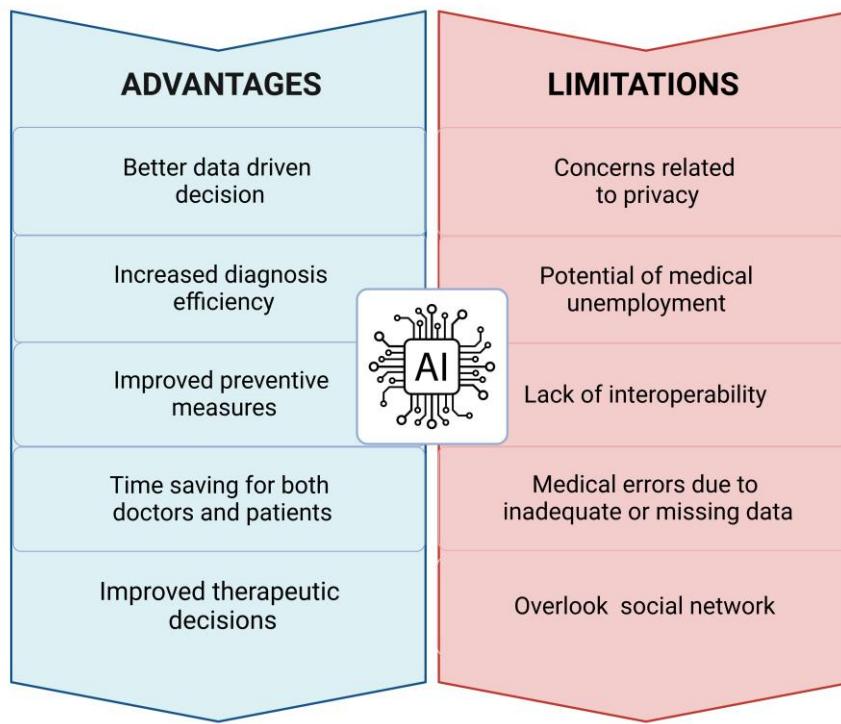


Figure 3. Advantages and limitations of AI. The scheme summarizes the main benefits along with the current concerns related to the use of AI in the clinical practice.

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. 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 [98]. 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 [99].

Supplementary Materials: Not applicable.

Author Contributions: Conceptualization, A.Fa. and C.I.; writing—original draft preparation, A.Fa. and A.Fe.; writing—review and editing, C.I.; visualization, A.Fa. All authors have read and agreed to the published version of the manuscript.

Funding: Please add: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: A.Fe. is 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 modelling, 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. Taylor N. "Duke Report Identifies Barriers to Adoption of AI Healthcare Systems." *MedTech Dive*. Accessed May 3, 2023. <https://www.medtechdive.com/news/duke-report-identifies-barriers-to-adoption-of-ai-healthcare-systems/546739/>.
2. Bray, Freddie, Mathieu Laversanne, Elisabete Weiderpass, and Isabelle Soerjomataram. "The Ever-Increasing Importance of Cancer as a Leading Cause of Premature Death Worldwide." *Cancer* 127, no. 16 (August 15, 2021): 3029–3030.
3. Ponomarenko, Elena A., Ekaterina V. Poverennaya, Ekaterina V. Ilgisonis, Mikhail A. Pyatnitskiy, Arthur T. Kopylov, Victor G. Zgoda, Andrey V. Lisitsa, and Alexander I. Archakov. "The Size of the Human Proteome: The Width and Depth." *International Journal of Analytical Chemistry* 2016 (2016): 7436849.
4. Nadhan, Revathy, Srishti Kashyap, Ji Hee Ha, Muralidharan Jayaraman, Yong Sang Song, Ciro Isidoro, and Danny N. Dhanasekaran. "Targeting Oncometabolites in Peritoneal Cancers: Preclinical Insights and Therapeutic Strategies." *Metabolites* 13, no. 5 (May 2023): 618. Accessed May 3, 2023. <https://www.mdpi.com/2218-1989/13/5/618>.
5. Hasin, Yehudit, Marcus Seldin, and Aldons Lusis. "Multi-Omics Approaches to Disease." *Genome Biology* 18, no. 1 (May 5, 2017): 83.
6. Tsakiroglou, Maria, Anthony Evans, and Munir Pirmohamed. "Leveraging Transcriptomics for Precision Diagnosis: Lessons Learned from Cancer and Sepsis." *Frontiers in Genetics* 14 (March 10, 2023): 1100352. Accessed May 3, 2023. <https://www.frontiersin.org/articles/10.3389/fgene.2023.1100352/full>.
7. Haga, Yoshimi, Yuriko Minegishi, and Koji Ueda. "Frontiers in Mass Spectrometry-Based Clinical Proteomics for Cancer Diagnosis and Treatment." *Cancer Science* 114, no. 5 (2023): 1783–1791. Accessed May 3, 2023. <https://onlinelibrary.wiley.com/doi/10.1111/cas.15731>.
8. Perkins, D. O., C. Jeffries, and P. Sullivan. "Expanding the 'Central Dogma': The Regulatory Role of Nonprotein Coding Genes and Implications for the Genetic Liability to Schizophrenia." *Molecular Psychiatry* 10, no. 1 (January 2005): 69–78.
9. Luo, Jiefeng, Mika Pan, Ke Mo, Yingwei Mao, and Donghua Zou. "Emerging Role of Artificial Intelligence in Diagnosis, Classification and Clinical Management of Glioma." *Seminars in Cancer Biology* 91 (2023): 110–123. Accessed May 3, 2023. <https://linkinghub.elsevier.com/retrieve/pii/S1044579X23000457>.
10. Wang, Suixue, Shuling Wang, and Zhengxia Wang. "A Survey on Multi-Omics-Based Cancer Diagnosis Using Machine Learning with the Potential Application in Gastrointestinal Cancer." *Frontiers in Medicine* 9 (January 10, 2023): 1109365. Accessed May 3, 2023. <https://www.frontiersin.org/articles/10.3389/fmed.2022.1109365/full>.
11. Liao, Jinzhuang, Xiaoying Li, Yu Gan, Shuangze Han, Pengfei Rong, Wei Wang, Wei Li, and Li Zhou. "Artificial Intelligence Assists Precision Medicine in Cancer Treatment." *Frontiers in Oncology* 12 (January 4, 2023): 998222. Accessed May 3, 2023. <https://www.frontiersin.org/articles/10.3389/fonc.2022.998222/full>.
12. He, Xiujing, Xiaowei Liu, Fengli Zuo, Hubing Shi, and Jing Jing. "Artificial Intelligence-Based Multi-Omics Analysis Fuels Cancer Precision Medicine." *Seminars in Cancer Biology* 88 (2023): 187–200. Accessed May 3, 2023. <https://linkinghub.elsevier.com/retrieve/pii/S1044579X22002632>.
13. Janes, Kevin A., and Michael B. Yaffe. "Data-Driven Modelling of Signal-Transduction Networks." *Nature Reviews Molecular Cell Biology* 7, no. 11 (November 2006): 820–828.
14. Dembrower, Karin, Erik Wåhlin, Yue Liu, Mattie Salim, Kevin Smith, Peter Lindholm, Martin Eklund, and Fredrik Strand. "Effect of Artificial Intelligence-Based Triaging of Breast Cancer Screening Mammograms on Cancer Detection and Radiologist Workload: A Retrospective Simulation Study." *The Lancet. Digital Health* 2, no. 9 (September 2020): e468–e474.
15. Coccia, Mario. "Artificial Intelligence Technology in Oncology: A New Technological Paradigm." *arXiv*, May 14, 2019. Accessed May 3, 2023. <http://arxiv.org/abs/1905.06871>.
16. Mary A, Majeed AKM (2015) Artificial intelligence and medical science: a survey. *Int J Adv Res Eng Manag* 1(1):66–72.

17. Arsene, C. Artificial Intelligence in healthcare The Future is Amazing. *Healthcare Weekly*. 2019.
18. Uzialko A. Artificial Intelligence Will Change Healthcare as We Know it.(2020); Business News Daily.
19. Mohammed, Akram, Greyson Biegert, Jiri Adamec, and Tomáš Helikar. "Identification of Potential Tissue-Specific Cancer Biomarkers and Development of Cancer versus Normal Genomic Classifiers." *Oncotarget* 8, no. 49 (October 17, 2017): 85692–85715.
20. Zhang, Yongqing, Shuwen Xiong, Zixuan Wang, Yuhang Liu, Hong Luo, Beichen Li, and Quan Zou. "Local Augmented Graph Neural Network for Multi-Omics Cancer Prognosis Prediction and Analysis." *Methods* 213 (2023): 1–9. Accessed May 3, 2023. <https://linkinghub.elsevier.com/retrieve/pii/S1046202323000348>.
21. Peng, Xinxin, Zhongyuan Chen, Farshad Farshidfar, Xiaoyan Xu, Philip L. Lorenzi, Yumeng Wang, Feixiong Cheng, et al. "Molecular Characterization and Clinical Relevance of Metabolic Expression Subtypes in Human Cancers." *Cell Reports* 23, no. 1 (April 3, 2018): 255–269.e4.
22. Yokota, Kazuki, Hiroo Uchida, Minoru Sakairi, Mayumi Abe, Yujiro Tanaka, Takahisa Tainaka, Chiyo Shirota, et al. "Identification of Novel Neuroblastoma Biomarkers in Urine Samples." *Scientific Reports* 11, no. 1 (February 18, 2021): 4055.
23. Barker, Matthew, and William Rayens. "Partial Least Squares for Discrimination." *Journal of Chemometrics* 17, no. 3 (March 24, 2003): 166–173. Accessed May 3, 2023. <https://onlinelibrary.wiley.com/doi/10.1002/cem.785>.
24. Rohart, Florian, Benoît Gautier, Amrit Singh, and Kim-Anh Lê Cao. "MixOmics: An R Package for 'omics Feature Selection and Multiple Data Integration." *PLOS Computational Biology* 13, no. 11 (November 3, 2017): e1005752. Accessed May 3, 2023. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005752>.
25. Westerhuis, Johan A., Huub C. J. Hoefsloot, Suzanne Smit, Daniel J. Vis, Age K. Smilde, Ewoud J. J. van Velzen, John P. M. van Duijnhoven, and Ferdi A. van Dorsten. "Assessment of PLSDA Cross Validation." *Metabolomics* 4, no. 1 (March 1, 2008): 81–89. Accessed May 3, 2023. <https://doi.org/10.1007/s11306-007-0099-6>.
26. Brereton, Richard G., and Gavin R. Lloyd. "Partial Least Squares Discriminant Analysis: Taking the Magic Away: PLS-DA: Taking the Magic Away." *Journal of Chemometrics* 28, no. 4 (2014): 213–225. Accessed May 3, 2023. <https://onlinelibrary.wiley.com/doi/10.1002/cem.2609>.
27. Gromski, Piotr S., Howbeer Muhamadali, David I. Ellis, Yun Xu, Elon Correa, Michael L. Turner, and Royston Goodacre. "A Tutorial Review: Metabolomics and Partial Least Squares-Discriminant Analysis--a Marriage of Convenience or a Shotgun Wedding." *Analytica Chimica Acta* 879 (June 16, 2015): 10–23.
28. Gu, Changdai, Gi Bae Kim, Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. "Current Status and Applications of Genome-Scale Metabolic Models." *Genome Biology* 20, no. 1 (June 13, 2019): 121.
29. Fang, Xin, Colton J. Lloyd, and Bernhard O. Palsson. "Reconstructing Organisms in Silico: Genome-Scale Models and Their Emerging Applications." *Nature Reviews Microbiology* 18, no. 12 (December 2020): 731–743.
30. Thiele, Ines, and Bernhard Ø Palsson. "A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction." *Nature Protocols* 5, no. 1 (January 2010): 93–121.
31. O'Brien, Edward J., Jonathan M. Monk, and Bernhard O. Palsson. "Using Genome-Scale Models to Predict Biological Capabilities." *Cell* 161, no. 5 (May 21, 2015): 971–987.
32. Chand S. "A comparative study of breast cancer tumor classification by classical machine learning methods and deep learning method". *Mach Vis Appl.* 2020;31.
33. Angermueller, Christof, Tanel Pärnamaa, Leopold Parts, and Oliver Stegle. "Deep Learning for Computational Biology." *Molecular Systems Biology* 12, no. 7 (July 29, 2016): 878.
34. Samiei, Mandana, Tobias Würfl, Tristan Deleu, Martin Weiss, Francis Dutil, Thomas Fevens, Geneviève Boucher, Sébastien Lemieux, and Joseph Paul Cohen. "The TCGA Meta-Dataset Clinical Benchmark." *arXiv*, October 18, 2019. Accessed May 3, 2023. <http://arxiv.org/abs/1910.08636>.
35. Jin, Shuting, Xiangxiang Zeng, Feng Xia, Wei Huang, and Xiangrong Liu. "Application of Deep Learning Methods in Biological Networks." *Briefings in Bioinformatics* 22, no. 2 (March 22, 2021): 1902–1917.
36. Rhee, Sungmin, Seokjun Seo, and Sun Kim. "Hybrid Approach of Relation Network and Localized Graph Convolutional Filtering for Breast Cancer Subtype Classification." *arXiv*, June 15, 2018. Accessed May 3, 2023. <http://arxiv.org/abs/1711.05859>.
37. Lee, Sangseon, Sangsoo Lim, Taeheon Lee, Inyoung Sung, and Sun Kim. "Cancer Subtype Classification and Modeling by Pathway Attention and Propagation." *Bioinformatics (Oxford, England)* 36, no. 12 (June 1, 2020): 3818–3824.
38. Dai, Hanjun, Hui Li, Tian Tian, Xin Huang, Lin Wang, Jun Zhu, and Le Song. "Adversarial Attack on Graph Structured Data." In *Proceedings of the 35th International Conference on Machine Learning*, 1115–1124. PMLR, 2018. Accessed May 3, 2023. <https://proceedings.mlr.press/v80/dai18b.html>.
39. Zhang, Xiang, and Marinka Zitnik. "GNNGuard: Defending Graph Neural Networks against Adversarial Attacks." In *Advances in Neural Information Processing Systems*, 33:9263–9275. Curran Associates, Inc.,

2020. Accessed May 3, 2023. <https://papers.nips.cc/paper/2020/hash/690d83983a63aa1818423fd6edd3bfdb-Abstract.html>.

- 40. Abdelaal, Tamim, Lieke Michielsen, Davy Cats, Dylan Hoogduin, Hailiang Mei, Marcel J. T. Reinders, and Ahmed Mahfouz. "A Comparison of Automatic Cell Identification Methods for Single-Cell RNA Sequencing Data." *Genome Biology* 20, no. 1 (September 9, 2019): 194.
- 41. Tan, Yuqi, and Patrick Cahan. "SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species." *Cell Systems* 9, no. 2 (August 28, 2019): 207-213.e2.
- 42. Hu, Jian, Xiangjie Li, Gang Hu, Yafei Lyu, Katalin Susztak, and Mingyao Li. "Iterative Transfer Learning with Neural Network for Clustering and Cell Type Classification in Single-Cell RNA-Seq Analysis." *Nature Machine Intelligence* 2, no. 10 (October 2020): 607-618.
- 43. Andreatta, Massimo, Jesus Corria-Osorio, Sören Müller, Rafael Cubas, George Coukos, and Santiago J. Carmona. "Interpretation of T Cell States from Single-Cell Transcriptomics Data Using Reference Atlases." *Nature Communications* 12, no. 1 (May 20, 2021): 2965.
- 44. Michielsen, Lieke, Marcel J. T. Reinders, and Ahmed Mahfouz. "Hierarchical Progressive Learning of Cell Identities in Single-Cell Data." *Nature Communications* 12, no. 1 (May 14, 2021): 2799.
- 45. Ranjan, Bobby, Florian Schmidt, Wenjie Sun, Jinyu Park, Mohammad Amin Honardoost, Joanna Tan, Nirmala Arul Rayan, and Shyam Prabhakar. "ScConsensus: Combining Supervised and Unsupervised Clustering for Cell Type Identification in Single-Cell RNA Sequencing Data." *BMC bioinformatics* 22, no. 1 (April 12, 2021): 186.
- 46. Gao, Jianjiong, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S. Onur Sumer, Yichao Sun, et al. "Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal." *Science Signaling* 6, no. 269 (April 2, 2013): pl1.
- 47. Grossman, Robert L., Allison P. Heath, Vincent Ferretti, Harold E. Varmus, Douglas R. Lowy, Warren A. Kibbe, and Louis M. Staudt. "Toward a Shared Vision for Cancer Genomic Data." *The New England Journal of Medicine* 375, no. 12 (September 22, 2016): 1109-1112.
- 48. Goldman, Mary J., Brian Craft, Mim Hastie, Kristupas Repečka, Fran McDade, Akhil Kamath, Ayan Banerjee, et al. "Visualizing and Interpreting Cancer Genomics Data via the Xena Platform." *Nature Biotechnology* 38, no. 6 (June 2020): 675-678.
- 49. Manzoni, Claudia, Demis A. Kia, Jana Vandrovčová, John Hardy, Nicholas W. Wood, Patrick A. Lewis, and Raffaele Ferrari. "Genome, Transcriptome and Proteome: The Rise of Omics Data and Their Integration in Biomedical Sciences." *Briefings in Bioinformatics* 19, no. 2 (March 1, 2018): 286-302.
- 50. Creixell, Pau, Jüri Reimand, Syed Haider, Guanming Wu, Tatsuhiko Shibata, Miguel Vazquez, Ville Mustonen, et al. "Pathway and Network Analysis of Cancer Genomes." *Nature Methods* 12, no. 7 (July 2015): 615-621.
- 51. Ngiam, Kee Yuan, and Ing Wei Khor. "Big Data and Machine Learning Algorithms for Health-Care Delivery." *The Lancet. Oncology* 20, no. 5 (May 2019): e262-e273.
- 52. Reyna, Matthew A., David Haan, Marta Paczkowska, Lieven P. C. Verbeke, Miguel Vazquez, Abdullah Kahraman, Sergio Pulido-Tamayo, et al. "Pathway and Network Analysis of More than 2500 Whole Cancer Genomes." *Nature Communications* 11, no. 1 (February 5, 2020): 729.
- 53. Luo, Ping, Yulian Ding, Xiujuan Lei, and Fang-Xiang Wu. "DeepDriver: Predicting Cancer Driver Genes Based on Somatic Mutations Using Deep Convolutional Neural Networks." *Frontiers in Genetics* 10 (2019): 13.
- 54. Jiao, Wei, Gurnit Atwal, Paz Polak, Rosa Karlic, Edwin Cuppen, PCAWG Tumor Subtypes and Clinical Translation Working Group, Alexandra Danyi, et al. "A Deep Learning System Accurately Classifies Primary and Metastatic Cancers Using Passenger Mutation Patterns." *Nature Communications* 11, no. 1 (February 5, 2020): 728.
- 55. Chaudhary, Kumardeep, Olivier B. Poirion, Liangqun Lu, and Lana X. Garmire. "Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 24, no. 6 (March 15, 2018): 1248-1259.
- 56. Gao, Feng, Wei Wang, Miaomiao Tan, Lina Zhu, Yuchen Zhang, Evelyn Fessler, Louis Vermeulen, and Xin Wang. "DeepCC: A Novel Deep Learning-Based Framework for Cancer Molecular Subtype Classification." *Oncogenesis* 8, no. 9 (August 16, 2019): 44.
- 57. Zeng, Xiangxiang, Siyi Zhu, Xiangrong Liu, Yadi Zhou, Ruth Nussinov, and Feixiong Cheng. "DeepDR: A Network-Based Deep Learning Approach to in Silico Drug Repositioning." *Bioinformatics (Oxford, England)* 35, no. 24 (December 15, 2019): 5191-5198.
- 58. Issa, Naiem T., Vasileios Stathias, Stephan Schürer, and Sivanesan Dakshanamurthy. "Machine and Deep Learning Approaches for Cancer Drug Repurposing." *Seminars in Cancer Biology* 68 (January 2021): 132-142.
- 59. Park, Youngjun, Dominik Heider, and Anne-Christin Hauschild. "Integrative Analysis of Next-Generation Sequencing for Next-Generation Cancer Research toward Artificial Intelligence." *Cancers* 13, no. 13 (June 24, 2021): 3148.

60. Zhang, Li, Chenkai Lv, Yaqiong Jin, Ganqi Cheng, Yibao Fu, Dongsheng Yuan, Yiran Tao, Yongli Guo, Xin Ni, and Tieliu Shi. "Deep Learning-Based Multi-Omics Data Integration Reveals Two Prognostic Subtypes in High-Risk Neuroblastoma." *Frontiers in Genetics* 9 (2018): 477.
61. Francescatto, Margherita, Marco Chierici, Setareh Rezvan Dezfooli, Alessandro Zandonà, Giuseppe Jurman, and Cesare Furlanello. "Multi-Omics Integration for Neuroblastoma Clinical Endpoint Prediction." *Biology Direct* 13, no. 1 (April 3, 2018): 5.
62. Huang, Zhi, Xiaohui Zhan, Shunian Xiang, Travis S. Johnson, Bryan Helm, Christina Y. Yu, Jie Zhang, et al. "SALMON: Survival Analysis Learning With Multi-Omics Neural Networks on Breast Cancer." *Frontiers in Genetics* 10 (2019): 166.
63. Xie, Gangcai, Chengliang Dong, Yinfei Kong, Jiang F. Zhong, Mingyao Li, and Kai Wang. "Group Lasso Regularized Deep Learning for Cancer Prognosis from Multi-Omics and Clinical Features." *Genes* 10, no. 3 (March 21, 2019): 240.
64. Chen, Luonan, and Jiarui Wu. "Bio-Network Medicine." *Journal of Molecular Cell Biology* 7, no. 3 (June 2015): 185–186.
65. Song, Hong, Lei Chen, Yutao Cui, Qiang Li, Qi Wang, Jingfan Fan, Jian Yang, and Le Zhang. "Denoising of MR and CT Images Using Cascaded Multi-Supervision Convolutional Neural Networks with Progressive Training." *Neurocomputing* 469 (January 16, 2022): 354–365. Accessed May 3, 2023. <https://www.sciencedirect.com/science/article/pii/S0925231221011061>.
66. Zhang, Le, Lei Zhang, Yue Guo, Ming Xiao, Lu Feng, Chengcan Yang, Guan Wang, and Liang Ouyang. "MCDB: A Comprehensive Curated Mitotic Catastrophe Database for Retrieval, Protein Sequence Alignment, and Target Prediction." *Acta Pharmaceutica Sinica. B* 11, no. 10 (October 2021): 3092–3104.
67. Zhou, Yadi, Fei Wang, Jian Tang, Ruth Nussinov, and Feixiong Cheng. "Artificial Intelligence in COVID-19 Drug Repurposing." *The Lancet. Digital Health* 2, no. 12 (December 2020): e667–e676.
68. Suhail, Yasir, Margo P. Cain, Kiran Vanaja, Paul A. Kurywchak, Andre Levchenko, Raghu Kalluri, and null Kshitiz. "Systems Biology of Cancer Metastasis." *Cell Systems* 9, no. 2 (August 28, 2019): 109–127.
69. Barabási, Albert-László, and Zoltán N. Oltvai. "Network Biology: Understanding the Cell's Functional Organization." *Nature Reviews. Genetics* 5, no. 2 (February 2004): 101–113.
70. Ali, Mehreen, Suleiman A. Khan, Krister Wennerberg, and Tero Aittokallio. "Global Proteomics Profiling Improves Drug Sensitivity Prediction: Results from a Multi-Omics, Pan-Cancer Modeling Approach." *Bioinformatics* (Oxford, England) 34, no. 8 (April 15, 2018): 1353–1362.
71. Sharifi-Noghabi, Hossein, Olga Zolotareva, Colin C. Collins, and Martin Ester. "MOLI: Multi-Omics Late Integration with Deep Neural Networks for Drug Response Prediction." *Bioinformatics* (Oxford, England) 35, no. 14 (July 15, 2019): i501–i509.
72. Kwon, Min-Seok, Yongkang Kim, Seungyeoun Lee, Junghyun Namkung, Taegyun Yun, Sung Gon Yi, Sangjo Han, et al. "Integrative Analysis of Multi-Omics Data for Identifying Multi-Markers for Diagnosing Pancreatic Cancer." *BMC genomics* 16 Suppl 9, no. Suppl 9 (2015): S4.
73. Gautam, Prson, Alok Jaiswal, Tero Aittokallio, Hassan Al-Ali, and Krister Wennerberg. "Phenotypic Screening Combined with Machine Learning for Efficient Identification of Breast Cancer-Selective Therapeutic Targets." *Cell Chemical Biology* 26, no. 7 (July 18, 2019): 970–979.e4.
74. Peng, Chen, Yang Zheng, and De-Shuang Huang. "Capsule Network Based Modeling of Multi-Omics Data for Discovery of Breast Cancer-Related Genes." *IEEE/ACM transactions on computational biology and bioinformatics* 17, no. 5 (2020): 1605–1612.
75. Mazzone, Peter J., Catherine Rufatto Sears, Doug A. Arenberg, Mina Gaga, Michael K. Gould, Pierre P. Massion, Vish S. Nair, 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." *American Journal of Respiratory and Critical Care Medicine* 196, no. 7 (October 1, 2017): e15–e29.
76. Seijo, Luis M., Nir Peled, Daniel Ajona, Mattia Boeri, John K. Field, Gabriella Sozzi, Ruben Pio, et al. "Biomarkers in Lung Cancer Screening: Achievements, Promises, and Challenges." *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 14, no. 3 (March 2019): 343–357.
77. Tsvetkov, Philipp O., Rémi Eyraud, Stéphane Ayache, Anton A. Bougaev, Soazig Malesinski, Hamed Benazha, Svetlana Gorokhova, et al. "An AI-Powered Blood Test to Detect Cancer Using NanoDSF." *Cancers* 13, no. 6 (March 15, 2021): 1294.
78. Parikh, Aparna R., Ignaty Leshchiner, Liudmila Elagina, Lipika Goyal, Chaya Levovitz, Giulia Siravegna, Dimitri Livitz, et al. "Liquid versus Tissue Biopsy for Detecting Acquired Resistance and Tumor Heterogeneity in Gastrointestinal Cancers." *Nature Medicine* 25, no. 9 (September 2019): 1415–1421.
79. Lu, Tian, and Jinming Li. "Clinical Applications of Urinary Cell-Free DNA in Cancer: Current Insights and Promising Future." *American Journal of Cancer Research* 7, no. 11 (2017): 2318–2332.
80. Heitzer, Ellen, Imran S. Haque, Charles E. S. Roberts, and Michael R. Speicher. "Current and Future Perspectives of Liquid Biopsies in Genomics-Driven Oncology." *Nature Reviews. Genetics* 20, no. 2 (February 2019): 71–88.

81. Savage, Richard, Mike Messenger, Richard D. Neal, Rosie Ferguson, Colin Johnston, Katherine L. Lloyd, Matthew D. Neal, 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* 12, no. 4 (April 1, 2022): e053590.
82. Cohen, Joshua D., Lu Li, Yuxuan Wang, Christopher Thoburn, Bahman Afsari, Ludmila Danilova, Christopher Douville, et al. "Detection and Localization of Surgically Resectable Cancers with a Multi-Analyte Blood Test." *Science (New York, N.Y.)* 359, no. 6378 (February 23, 2018): 926–930.
83. Cree, Ian A., Lesley Uttley, Helen Buckley Woods, Hugh Kikuchi, Anne Reiman, Susan Harnan, Becky L. Whiteman, et al. "The Evidence Base for Circulating Tumour DNA Blood-Based Biomarkers for the Early Detection of Cancer: A Systematic Mapping Review." *BMC cancer* 17, no. 1 (October 23, 2017): 697.
84. Aerts, Hugo J. W. L., Emmanuel Rios Velazquez, Ralph T. H. Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, et al. "Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach." *Nature Communications* 5 (June 3, 2014): 4006.
85. Hosny, Ahmed, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz, and Hugo J. W. L. Aerts. "Artificial Intelligence in Radiology." *Nature Reviews. Cancer* 18, no. 8 (August 2018): 500–510.
86. Chan, Heang-Ping et al. "Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography-a review." *Academic radiology* vol. 15,5 (2008): 535-55.
87. Rasch, C., I. Barillot, P. Remeijer, A. Touw, M. van Herk, and J. V. Lebesque. "Definition of the Prostate in CT and MRI: A Multi-Observer Study." *International Journal of Radiation Oncology, Biology, Physics* 43, no. 1 (January 1, 1999): 57–66.
88. Chen, Weijie, Maryellen L. Giger, Hui Li, Ulrich Bick, and Gillian M. Newstead. "Volumetric Texture Analysis of Breast Lesions on Contrast-Enhanced Magnetic Resonance Images." *Magnetic Resonance in Medicine* 58, no. 3 (September 2007): 562–571.
89. Zhu, Bin, Nan Song, Ronglai Shen, Arshi Arora, Mitchell J. Machiela, Lei Song, Maria Teresa Landi, et al. "Integrating Clinical and Multiple Omics Data for Prognostic Assessment across Human Cancers." *Scientific Reports* 7, no. 1 (December 5, 2017): 16954.
90. Liu, Saifeng, Huaixiu Zheng, Yesu Feng, and Wei Li. "Prostate Cancer Diagnosis Using Deep Learning with 3D Multiparametric MRI." 1013428, 2017. Accessed May 3, 2023. <http://arxiv.org/abs/1703.04078>.
91. Chen, Quan, Xiang Xu, Shiliang Hu, Xiao Li, Qing Zou, and Yunpeng Li. "A Transfer Learning Approach for Classification of Clinical Significant Prostate Cancers from MpMRI Scans" 10134 (March 1, 2017): 101344F. Accessed May 3, 2023. <https://ui.adsabs.harvard.edu/abs/2017SPIE10134E..4FC>.
92. Seah, Jarrel C. Y., Jennifer S. N. Tang, and Andy Kitchen. "Detection of Prostate Cancer on Multiparametric MRI." edited by Samuel G. Armato and Nicholas A. Petrick, 1013429. Orlando, Florida, United States, 2017. Accessed May 3, 2023. <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2277122>.
93. Mehrtash, Alireza, Alireza Sedghi, Mohsen Ghafoorian, Mehdi Taghipour, Clare M. Tempany, William M. Wells, Tina Kapur, Parvin Mousavi, Purang Abolmaesumi, and Andriy Fedorov. "Classification of Clinical Significance of MRI Prostate Findings Using 3D Convolutional Neural Networks." *Proceedings of SPIE--the International Society for Optical Engineering* 10134 (February 11, 2017): 101342A.
94. Azizi, Shekoofeh, Sharareh Bayat, Pingkun Yan, Amir Tahmasebi, Guy Nir, Jin Tae Kwak, Sheng Xu, et al. "Detection and Grading of Prostate Cancer Using Temporal Enhanced Ultrasound: Combining Deep Neural Networks and Tissue Mimicking Simulations." *International Journal of Computer Assisted Radiology and Surgery* 12, no. 8 (August 2017): 1293–1305.
95. Al Bakir, Maise, Ariana Huebner, Carlos Martínez-Ruiz, Kristiana Grigoriadis, Thomas B. K. Watkins, Oriol Pich, David A. Moore, et al. "The Evolution of Non-Small Cell Lung Cancer Metastases in TRACERx." *Nature* 616, no. 7957 (April 20, 2023): 534–542. Accessed May 3, 2023. <https://www.nature.com/articles/s41586-023-05729-x>.
96. Martínez-Ruiz, Carlos, James R. M. Black, Clare Puttick, Mark S. Hill, Jonas Demeulemeester, Elizabeth Larose Cadieux, Kerstin Thol, et al. "Genomic-Transcriptomic Evolution in Lung Cancer and Metastasis." *Nature* 616, no. 7957 (April 20, 2023): 543–552. Accessed May 3, 2023. <https://www.nature.com/articles/s41586-023-05706-4>.
97. Chen, Chen, Zehua Wang, Yi Ding, Lei Wang, Siyuan Wang, Haonan Wang, and Yanru Qin. "DNA Methylation: From Cancer Biology to Clinical Perspectives." *Frontiers in Bioscience-Landmark* 27, no. 12 (December 20, 2022): 326. Accessed May 3, 2023. <https://www.imrpress.com/journal/FBL/27/12/10.31083/j.fbl2712326>.

98. Olaronke, Iroju, and Ojerinde Oluwaseun. "Big Data in Healthcare: Prospects, Challenges and Resolutions." 2016 Future Technologies Conference (FTC) (2016): 1152–1157. Accessed May 3, 2023. <http://ieeexplore.ieee.org/document/7821747/>.
99. Esteva, Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks." *Nature* 542, no. 7639 (February 2, 2017): 115–118.

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.