REVIEW

Deep learning in cancer genomics and histopathology

Michaela Unger[1]* and Jakob Nikolas Kather[1,2,3]*

Abstract

Histopathology and genomic profiling are cornerstones of precision oncology and are routinely obtained for patients with cancer. Traditionally, histopathology slides are manually reviewed by highly trained pathologists. Genomic data, on the other hand, is evaluated by engineered computational pipelines. In both applications, the advent of modern artificial intelligence methods, specifically machine learning (ML) and deep learning (DL), have opened up a fundamentally new way of extracting actionable insights from raw data, which could augment and potentially replace some aspects of traditional evaluation workflows. In this review, we summarize current and emerging applications of DL in histopathology and genomics, including basic diagnostic as well as advanced prognostic tasks. Based on a growing body of evidence, we suggest that DL could be the groundwork for a new kind of workflow in oncology and cancer research. However, we also point out that DL models can have biases and other flaws that users in health-care and research need to know about, and we propose ways to address them.

Keywords Deep learning, Precision oncology, Histopathology, Genomics, Multimodality

Background

Precision oncology is based on diagnostic histopathological and genomic methods, which enable the application of a suitable therapy to patients [1]. Histopathology investigates the morphology, or phenotype, of a tumor and is indispensable to diagnose and subtype cancer. One of the most general and widely used methods in histopathology is staining of tissue slides with hematoxylin and eosin (H&E) [2]. To complement the phenotypic information, genomic biomarkers are routinely used for patients with advanced or metastatic cancer since they exhibit a predictive power for the patient's survival or for the effectiveness of a cancer drug. Thus, in many cases, genomics allows a more personalized form of therapy [3]. Given these advancements, it is not surprising that precision oncology could improve clinical outcomes in the last decades [4, 5]. However, precision oncology is inherently data-intensive: to support treatment decisions, a wide range of data is required, including general patient information such as age, biological sex, medical history, patient preferences, radiological imaging, histopathology, and molecular and genetic assays. At the same time, the amount of available information beyond patient data is extensive as well. For example, in 2021, the US Food and Drug Administration (FDA) had approved a total of 243 cancer drugs for patient therapy [6]. Combined, the quantity of patient-specific data and the number of treatment options create a vast decision tree which is becoming more complex to navigate for patients and physicians. Therefore, there is a need for tools to support cancer care by efficiently utilizing and analyzing all available information.

*Correspondence:
Michaela Unger
michaela.unger@tu-dresden.de
Jakob Nikolas Kather
jakob_nikolas.kather@tu-dresden.de
[1] Else Kroener Fresenius Center for Digital Health, Medical Faculty Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany
[2] Department of Medicine I, University Hospital Dresden, Dresden, Germany
[3] Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany

One solution for this growing demand could be the application of computer-aided methods. Improvements in computer hardware and algorithms have multiplied our abilities to process large-scale data since the late 20th century. Today, artificial intelligence (AI) methods have become ubiquitous tools in our everyday life. AI can solve complex tasks at the level of human experts, such as in language translation and object detection [7, 8]. This is also true for biomedical research, where AI is able to solve complex problems like predicting protein folding from amino acid sequences [9] or analyzing and interpreting radiology imaging data [10]. As a potential advantage over human skills, AI methods are scalable and can process vast amounts of data in a relatively short time.

One most fundamental component of AI is machine learning (ML). There are three main approaches to ML: reinforcement, unsupervised, supervised learning. In reinforcement learning, the model is rewarded for making correct decisions. In unsupervised learning, the model is tasked to learn from data, but is given no additional information about it. For example, clustering methods can identify similar instances in a given dataset, without being provided with explicit labels on each instance. Supervised learning, in contrast, can use human-labeled data and tasks the model with automating the labeling process. A portion of this data is given to the model to predict labels, and the model is penalized when it gives the wrong output. Model architectures used for supervised learning include support vector machines (SVMs), decision trees and artificial neural networks. These models can vary greatly in size, with the number of parameters ranging from hundreds of parameters to billions of parameters in neural networks [11]. Whenever ML is applied to image or text data, deep artificial neural networks, also known as deep learning (DL) [12], are the favored models due to their robustness and effectiveness in handling complex data structures. In precision oncology, AI with DL can process large amounts of histopathologic and genomic data (Fig. 1) [1, 13, 14]. Notably, some studies even adopted multimodal models that apply ML and DL to several data types simultaneously, such as combining histopathological images with genetic data [15–17]. This approach of multimodal data integration could potentially improve model performance by incorporating additional patient information and leveraging synergistic effects between complementary data types.

Here, we provide a high-level overview of DL's role in pathology, genomics, and multimodal data analysis. To bring structure to the diversity in the academic literature, we establish a guiding framework. In our analysis, we divide our investigation into six fields of clinically focused application, as established by previous studies [18]. Three "basic" applications are as follows: predicting the diagnosis (cancer detection), subtype, and grading of a tumor; and three "advanced" applications are as follows: predicting prognosis (survival probability of the patient), patterns of genetic alterations (such as the detection of driver mutations), or treatment response to a specific treatment scheme or a single medicine [18–20]. Furthermore, we discuss the potential limitations of DL approaches in clinical routines and provide insights into future trajectories of these fields. Altogether, this review should not only inform about the most recent developments in the area but also inspire researchers to further contribute to this topic and close its existing gaps.

DL in histopathology

Histopathology is a fundamental part of precision oncology. Virtually all solid tumor entities must be diagnosed by histopathology or cytology. In essence, all clinical decisions based on treatment and follow-up depend on histopathological information. In digital pathology, tissue slides are digitally captured as whole slide images (WSI) in high resolution, yielding images with billions of pixels, or "gigapixel images." AI can process such digital information and has emerged as the default tool to automate diagnostic processes and identify new biomarkers in WSIs (Fig. 1).

Most AI studies in histopathology employ supervised DL. Of particular relevance are "weakly" supervised approaches, in which the objective of the system is to predict a "label" for the WSI in its entirety [13, 21, 22]. A "label" can refer to any of the basic and advanced categories, including properties of slides (presence of tumor), properties of tumors (subtype or genetic alterations), and of patients (survival or response) [13]. During training, a weakly supervised tumor detection system only has access to a label on a slide level. For example, the label could denote: "does this slide contain a tumor, yes or no?". An alternative approach is "strongly" supervised learning. Here, the objective is to delineate tumor tissue or detect cell types based on accurate, manual annotations. Weakly supervised approaches obviate the need for manual annotation and, hence, are more scalable to large image archives. In addition, weakly supervised approaches allow us to predict more abstract properties of tumors, such as the presence of mutations or the survival of patients [13, 22–25].

DL for basic histopathological tasks

One of the earliest studies on weakly supervised DL in histopathology was conducted by Ertosun and Rubin in 2015 (Fig. 2a) [26], in which the authors automated histological grading in primary brain tumors using a convolutional neural network (CNN). CNNs are a type of neural network commonly used in image analysis, containing
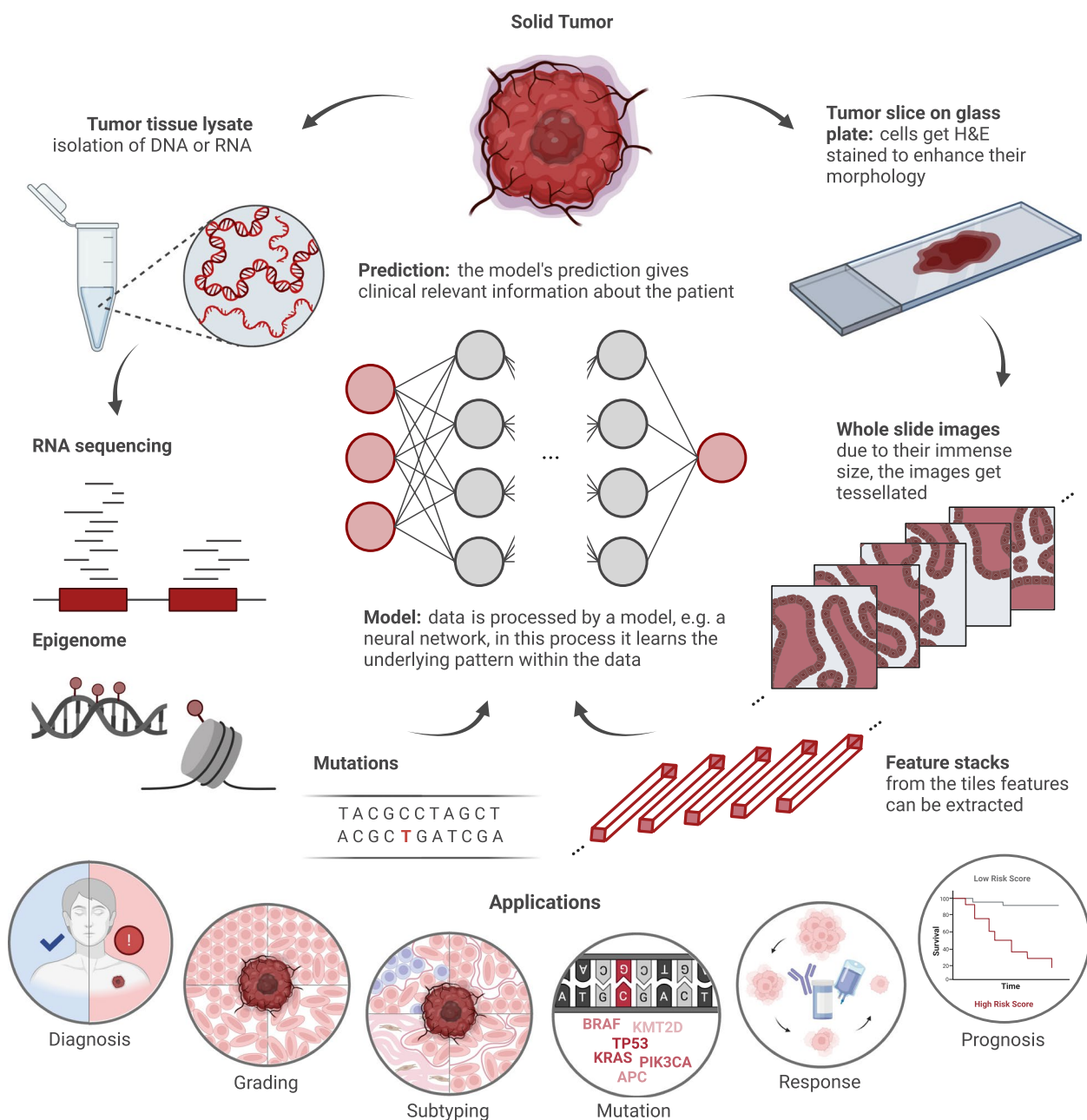
Fig. 1 Workflow of AI in histopathology and clinical genomics. In this simplified workflow, a tissue of a solid tumor is harvested via surgery or biopsy. One part is sequenced in the genomics facility to obtain molecular data about, for instance, RNA, epigenetics, or mutations, while another part is sent to the pathology department. There, tumor slices are captured on glass slides and stained with hematoxylin and eosin (H&E). Images of these glass slides can then be taken. Tabular and image data are used to train models, e.g., neural networks to provide a prediction. In this review, we describe six distinct medical application tasks (Diagnosis, Grading, Subtyping, Mutation, Response, and Survival) for these models

so-called convolutional layers. Vividly speaking, layers of convolution find basic structures like corners and edges in the original image which are then concatenated by the neural network to higher hierarchies, and with this, determine global patterns shared between images. Ertosun and Rubin were among the earliest to move from

handcrafted features with simple ML classifiers to DL. This enabled them to address a clinically relevant classification task in computational pathology.

Prior to tumor grading or any other step, the diagnosis must take place. Hence, diagnosis is one of the most obvious and most common applications of DL in
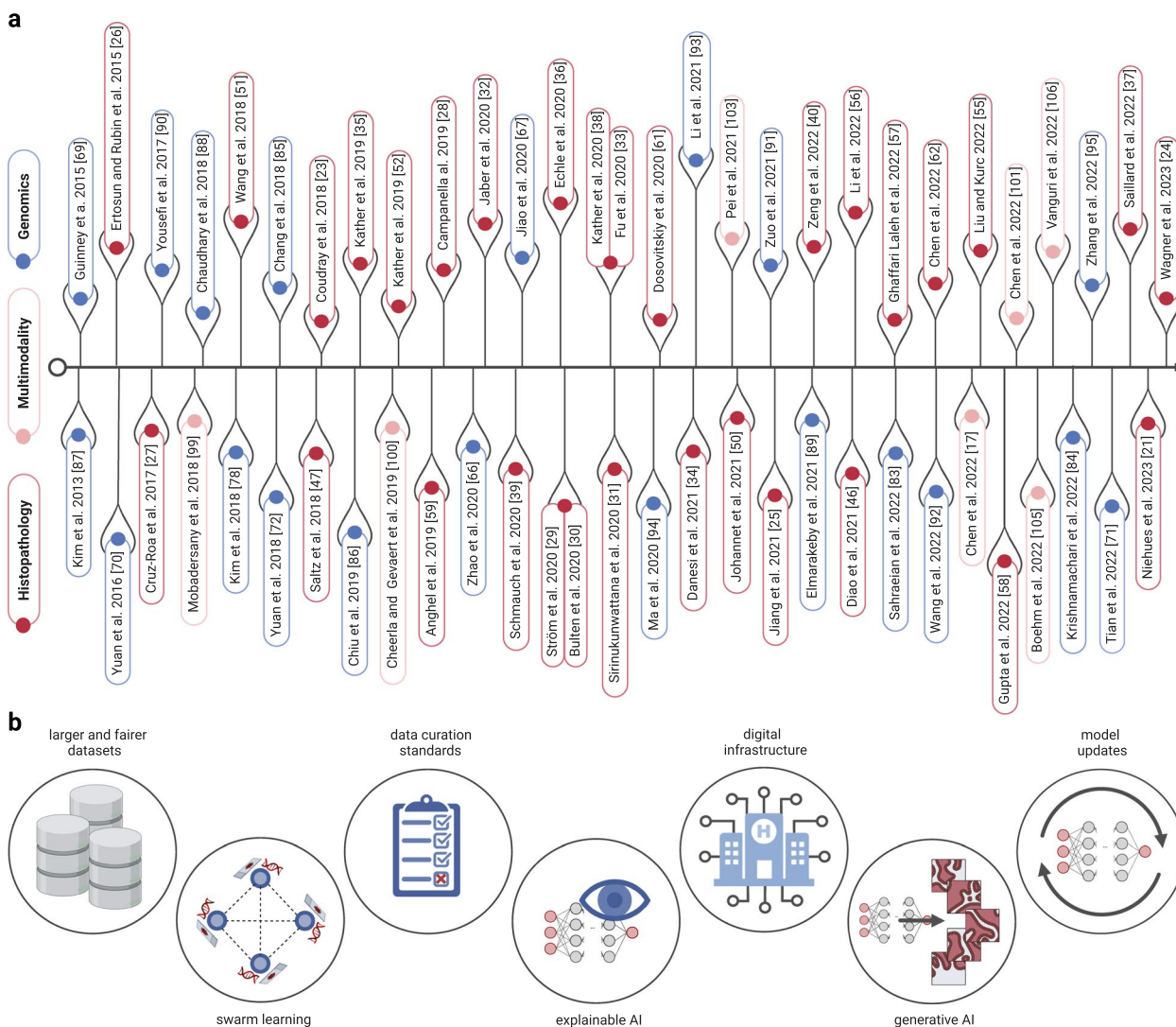
a

b

larger and fairer datasets

data curation standards

digital infrastructure

model updates

swarm learning

explainable AI

generative AI

Fig. 2 Timeline and outlook. **a** The timeline of milestone papers mentioned in this review. Articles are colored by research area (blue — genomics, rose — multimodal, red — histopathology). **b** Future perspectives AI will face in the next years to be applied in clinical routines

histopathology. In this task, models need to differentiate tumor tissue and healthy tissue on WSIs in a strongly or weakly supervised manner. One of the first studies which employed DL for tumor detection was carried out by Cruz-Roa et al. [27] (Fig. 2a) in 2017. The authors diagnosed breast cancer by using a CNN which was trained on almost 400 WSIs. Their model reached a high performance for tumor detection. At this time, essential preprocessing steps were already established, e.g. making large WSIs usable by tesselating them (Fig. 1). In 2019, the field of cancer detection with weakly supervised DL was markedly changed as a result of a large-scale seminal work by Campanella et al. [28] (Fig. 2a), whose multiple-instance learning model outperformed strongly supervised models with an area under the receiver operating

characteristic (AUROC) curve as high as 0.986. DL models could therefore probably assist pathologists in the future by pre-labeling samples, potentially reducing the load of confirmatory molecular assays.

One year later, Ström et al. [29] and Bulten et al. [30] (Fig. 2a) demonstrated that DL was able to solve a subtyping task in solid tumors, another important application of DL. Their approaches did not only include tumor segmentation, but also prediction of Gleason grade in prostate cancer with weakly supervised learning. Complementary to these diagnostic tasks, the most influential recent study in digital pathology was published by Coudray et al. [23] (Fig. 2a) in 2018. Coudray et al. established weakly-supervised methods for the slide-level prediction of histological subtype of non-small-cell lung

cancer and, importantly, showed that genetic alterations in targetable genes are predictable from histopathology slides [23]. Although straightforward in hindsight, these studies were the first large-scale evidence that weakly supervised DL could differentiate between morphologies of cancer subtypes and link the cancer genotype from morphology alone. In the subsequent years, many studies extended this methodology to other subtypes of solid tumors. A notable example is the Consensus Molecular Subtypes (CMS) of colorectal cancer, which were shown to be predictable from routine pathology slides by Sirinukunwattana et al. [31] (Fig. 2a) in 2021. Similarly, in breast cancer, Jaber et al. [32] (Fig. 2a) presented a model that classified the five molecular subtypes of breast cancer (luminal A, luminal B, *HER2*-enriched, basal-like, normal-like) from histopathology slides with high accuracy. All these studies indicate that DL could potentially streamline diagnostic workflows by automating basic diagnostic processes, like subtyping and grading. Additionally, in a broader sense, these studies show that the ground truth for DL-based predictions can be obtained from any source as long as there is a phenotypic change the model can detect.

DL for advanced histopathological tasks

Of similar importance to the DL method that is used, is the data a model is trained on. One of the largest studies in recent years was conducted by Fu et al. [33] (Fig. 2a) incorporating more than 17,000 WSIs from the TCGA. Important to note is that the performance of DL models is dependent on the size and quality of the input. Therefore, it was not surprising that such an immense dataset led to an AUROC of 0.98 when distinguishing cancer types. However, not only did they classify cancer tissues, but they also predicted genome duplications, driver mutations like *TP53* or *BRAF*, and tumor-infiltrating lymphocyte (TIL) scores, setting the stage for a broad application of AI in creating pathology biomarkers. Genetic alterations in cancer, as predicted by Fu et al., can be drug targets, biomarkers, or both. For example, the presence of certain *BRAF* mutations in many tumor types is a direct target for treatment with *BRAF* inhibitors. A concrete biomarker is microsatellite instability (MSI), which acts as a biomarker for immune checkpoint inhibitors [34]. Some of these targets and biomarkers can be predicted with DL from pathology slides. In 2019 Kather et al. [35] (Fig. 2a) were able to predict MSI in colorectal, gastric, and endometrial cancers. As a following publication, Echle et al. [36] (Fig. 2a) trained models to predict MSI in colorectal cancer, along with the driver mutations *BRAF* and *KRAS*, in larger patient cohorts. Today, some of these approaches have been implemented by commercial entities and are being marketed as algorithms for

routine clinical use in Europe [37]. In addition to predicting single gene mutations or molecular subtypes, several studies have shown that it is also possible to extract expression levels of individual genes, or panel expression profiles directly from WSIs [38–40]. Consequently, AI could in principle be used to pre-screen for a wide range of molecular alterations and suggest which targets should be further analyzed.

Another alternative for receiving information about the patient status is investigating the tumor microenvironment. The interactions between the patient's immune system and the cancer can be relevant for overall survival [41, 42] or therapy response [43]. For example, patient outcomes can be predicted by the number of TILs [44]. Moreover, the importance of spatial biology was already known as early as 2006; however, it has not been translated to clinical routines yet [45]. On this account, DL models emerged that detect TILs and catalog cell types [46, 47] in a specimen annotation-free and in an end-to-end approach. Therefore, DL could offer an easier path to clinical application of still unused knowledge.

As mentioned before, the prediction of genomic or morphologic biomarkers from routine histology slides is clinically relevant for the patient. However, biomarkers are just proxies for clinical outcomes—survival or treatment response. Direct prediction of treatment response to specific drugs from histopathology images could theoretically even outperform the predictive power of genomic biomarkers. Thus, drug response prediction is one of the latest advanced applications in digital pathology. In 2020, a study on predicting the response to chemotherapy in nasopharyngeal cancer was published by Liu et al. [48] (Fig. 2a). Similarly, Li et al. [49] (Fig. 2a) trained a DL model to predict a pathological complete response after neoadjuvant chemotherapy. Furthermore, immunotherapy, as another form of cancer treatment, was under investigation by Johannet et al. [50] (Fig. 2a) in 2021. The fact that DL captures underlying connections between tissue morphology and treatment response shows that the predictive capabilities of such models reach far beyond human expertise. However, these studies need many comparable cases and treatment data with a consecutive target score which is why drug response is one of the most difficult applications to establish a large dataset with good quality ground truth. Therefore, the current state of DL in treatment response suggests that direct predictions require more extensive studies in the future.

The second clinical endpoint being directly predicted by DL in histopathology is the prognosis of cancer patients, i.e., forecasting patient survival. To elucidate the prognosis of a patient is from fundamental interest since therapy decisions and patient care are directly dependent on it. In DL research, early publications used,

for example, shape and boundary [51] or tissue proportions [52] of tumors as features that can be linked to patient outcomes. Today, DL models construct predictive risk scores in a straightforward manner. Information about absolute survival times is collected and combined with the censoring data of each patient. Afterwards, the model can learn which pattern to connect with a longer or shorter lifespan of a patient [53, 54]. The success of this application type could also lay in its potential to reveal yet unknown relationships between survival and phenotype.

Similarly to clinical targets getting more refined over years of research, model architectures changed as well. For most early studies, CNNs were applied as the model of choice. Later, feature extraction, a process in which pretrained DL models reduce the dimensionality of input images to smaller matrices or vectors, became the state-of-the-art method [25, 55–59] (Fig 1). Another change in model design was introduced after 2017, in which transformer neural networks [60, 61] were developed. These models can weigh parts of their input differently based on an attention mechanism and parallelize the processing of multiple parts of the input data in a computationally efficient way. In 2022, Chen et al. [62] (Fig. 2a) predicted survival through the use of vision transformers, which were able to outperform convolution-based models in many cancer types.

In summary, during the last years, AI in pathology underwent many changes and trends. Starting with simple diagnostic tools the field was soon able to outperform trained pathologists in tumor detection. Subsequently, research demonstrated that patterns in WSIs can be used for prognostic tasks as well, facilitating therapy decisions based on mutational status, drug response, or overall survival. Nevertheless, rapid changes in the model landscape of DL make it challenging for companies to develop these technologies into static products. To put this into perspective, in 2023, only four AI-based tools were FDA-approved and applied in pathology [63]. Therefore, it would be clearly desirable to increase this number and move more DL tools into diagnostic routine in precision oncology.

DL in clinical genomics

Unique molecular characteristics of a tumor are encoded in its genome [64]. Thus, research in clinical genomics is a key to delivering precision oncology since it studies the human genome with a focus on a disease genotype. Thereby, genotypic properties such as genomic instability or mutation status of the tumor complement the phenotypic and spatial changes addressed in histopathology. Clinical genomics not only employs classical genomic data from whole genome or exome sequencing, but also

RNA-sequencing, methylation assays, copy number variation analyses, and more as information sources (Fig. 1). With this, it supports the identification of the patient's exact type of cancer, its potential primary site, responsiveness to certain drugs, or the patient's prognosis.

Previously, analyzing genomic data was only conducted by classical bioinformatics, which employed algorithms to perform tasks such as sequence alignment, variant calling, or differential expression analysis. However, these algorithms are highly hand-engineered and focus on finding patterns which are predefined by human experts. The potential utility of AI for clinical genomics is to expand this toolkit by offering the possibility of deeper data analysis than previously attainable. Patterns that are unknown or undetectable to humans, such as the way a protein folds into its final shape or the signature left by a mutagenic process in our DNA, were discovered through the use of ML [9, 65]. Revealing novel paradigms with AI could contribute to innovations in clinical genomics that are otherwise not possible for standard bioinformatics approaches.

DL for basic genomic tasks

DL applications in genomics have developed differently than those in histopathology. Usually, genomic information is extracted after a cancer has been diagnosed and followed up histologically. As a result, DL in clinical genomics is more involved in the advanced tasks, e.g., finding biomarkers for certain therapies or drug-response, rather than streamlining workflows by diagnosing cancer. Nevertheless, DL can be utilized in patient cases where the diagnosis is not straightforward. For example, in 2020, Zaoh et al. [66] (Fig. 2a) used a DL model to predict the original tumor tissue for patients with cancer of unknown primary from RNA-sequencing data. Similarly, in the same year, Jiao et al. [67] (Fig. 2a) found that DL can be used on passenger mutation patterns to distinguish primary from metastatic tumors. Even though these studies are not focused on cancer detection, they can provide valuable insights for the downstream decision-making process.

One basic DL application that is more prominent for clinical genomics is subtyping. Articles such as Sienkiewicz et al. [68] (Fig. 2a) utilized classical unsupervised ML in the form of non-negative matrix factorization to cluster omics data of cancer patients to discover molecular subtypes. In order to refine these classes, more sophisticated models such as random forests or DL can also be employed [69–71]. DeepGene, a model developed by Yuan et al. [70] (Fig. 2a) in 2016, used somatic mutations as their information source, whereas two years later, they published another study performing the same task, this time with copy number alterations and chromatin

structure data [72]. Despite these advancements, the state-of-the-art to detect major cancer subtypes remains the morphological evaluation in most cases, with some exceptions being the recently introduced classifications of brain tumors. High costs and standardization issues associated with sequencing are limitations that prevent molecular subtypes from clinical adoption [73]. Furthermore, while some molecular subtypes such as the CMS in colorectal cancer can partially be correlated to relevant clinical outcomes, a more extensive data exploration and validation is needed to provide clinical evidence and hence foster a broader acceptance in the community.

DL for advanced genomic tasks

The task of mutation prediction from genomic data might seem contradictory, since detecting driver mutations from it forms the ground truth for DL predictions. Classical variant calling algorithms spot nucleotide changes in the cancer genome compared to a reference, with additional tools subsequently determining if the respective mutation affects a cancer-driving gene [74–77]. In these tasks, employing DL is not a necessity. Therefore, the approaches towards mutation prediction with DL differ between those for histopathology and genomics. One example for this paradigm shift is the DL-supported discovery of gene mutations previously unrelated to cancer. In 2018, Kim et al. [78] (Fig. 2a) used what are known as skip-gram networks to visualize mutations and discover novel cancer drivers. Mutations in genes such as *CRLF2*, *TFE3*, or *DUSP22* were positive hits of their method but were previously not described as driver mutations in literature. Nevertheless, to make this knowledge clinically actionable, wet lab validation studies are needed to elucidate their mechanism of action. Besides conventional driver mutations, the whole mutational spectrum of a cancer genome, including general somatic mutations, can additionally provide important insights [79, 80]. Furthermore, variant calling must be performed as a baseline to detect driver mutations. Today, there are different bioinformatic tools that process whole genome or exome sequencing data to first align reads to a reference genome and then find changes in the donor sample compared to the reference [81, 82]. Due to the complexity of this problem, research also developed DL-based methods to improve variant calling. For example, in 2022 Sahraeian et al. [83] (Fig. 2a) used CNNs to process matched tumor and normal reads to catalog somatic mutations. A similar approach was used by Krishnamachari et al. [84] (Fig. 2a) three years later. Both methods displayed superior accuracy compared to conventional bioinformatic tools. Nevertheless, the large amount of training data and high computing power needed for DL could hinder its broad adoption. Despite these challenges, our examples

demonstrate that DL has the potential to detect genomic variations at diverse scales with promising results.

Drug response predictions in clinical genomics often rely on data generated via cancer cell line cultures rather than solid tumors. In pharmacogenomics, genome-wide association studies enable the simultaneous screening of a broad number of cancer-drug pairs and therefore build the foundation for many DL applications. In 2018, Chang et al. [85] (Fig. 2a) predicted drug efficacy from genomic information of cancer cell lines and drug structural information, whereas Chiu et al. [86] (Fig. 2a) relied on mutation and expression data, without incorporating information about the drug's chemical properties. This contrasts computational pathology since cell line-based approaches are massive simplifications of human tumors. Cancer cell lines are often genetically altered to achieve immortality introducing genotypic and phenotypic biases which eventually make them less biologically comparable to primary cancer cells. Moreover, drug screens conducted in cell lines contain no other representative elements of their original tumor microenvironment. As a result, DL approaches to evaluate drug-cancer interactions come into question and call for more practical data sources.

In contrast to current genomic drug response models, DL approaches for prognosis predictions could offer a more direct integration into clinical workflows. One of the first publications regarding DL in clinical genomics predicted cancer outcomes of ovarian cancer from DNA methylation, miRNA and bulk-RNA expression, and copy number alterations (CNAs). The software package ATHENA, developed by Kim et al. [87] (Fig. 2a), incorporated this data into grammatical evolution neural networks. Here, over several iterations, sets of neural networks with varying parameters are constructed, and the best-performing networks are combined in the following iteration until the best solution is reached. Another impactful study in this area of research was carried out by Chaudhary et al. [88] in 2017, who used "-omics" data from different platforms to predict survival classes in hepatocellular carcinoma. Their model stratified patients into distinct risk groups and demonstrated comparable performance to models that additionally used clinical data, such as gender, cancer grade, and other risk factors. Furthermore, relations between survival and mutations in *TP53*, high expression of *BIRC5*, and other types of genomic alterations were shown as well. Elmarakeby et al. [89] in 2021 discovered that alterations of formerly unrelated genes such as *MDM4*, *FGFR1*, or *MALM3* are associated with prostate cancer outcomes. For this they used a neural network with specific constraints: nodes represent a biological entity and edges their relations. By doing

so, they limited the degree of connectivity in the network to incorporate prior biological knowledge and to restrict the computational complexity. The advantage of genomics in prognosis predictions lies in the ability to obtain data at multiple levels, which can range from genomic properties to its specific sequences. As a result, subtle changes in the cellular machinery can be identified as potential biomarkers. Nevertheless, compared to histopathology, many genomic biomarkers first need to be validated clinically to be translated into medical workflows.

An aspect that distinguishes AI in clinical genomics from histopathology is the diversity of model types used. Whereas in DL for histopathology basic model architectures were adapted from computer vision, DL in genomics did not find a direct analog in computer science, leading to a broader experimentation with various model types. For example, Chaudhary et al. [88] utilized an autoencoder, a form of DL, to integrate diverse omics data and then stratified liver cancer patients into risk groups. Yousefi et al. [90] deployed multi-layer perceptrons combined with a Cox survival model for prognosis predictions. Furthermore, random forests, gradient boosting, convolutional or graph-based networks, and more simple regression methods are applied in the field as well [91–94]. Today, similar to histopathology, transformer neural networks are becoming more and more prevalent in the field [95]. Taking into account the heterogeneity of genomic data, there is no single method that can be universally applied, underlining the need for continuous exploration in the future.

Cancer genomics remains a promising area for the application of DL. Many of the designated studies have shown to effectively complement bioinformatics tools and explore applications beyond them. Nevertheless, to our knowledge, DL tools for genomics have not yet received regulatory approval for clinical use. However, the cost for sequencing has dramatically decreased since the first human genome project, which indicates that genomic testing will probably become available to a broad range of cancer patients in the future [96, 97]. Therefore, we anticipate that DL in precision oncology will also benefit from more widely available genomic data. Apart from the application classes we mention in this review, DL could play numerous roles in clinical genomics in oncology. For example, DL could leverage tasks ranging from fundamental steps such as quality control or alignment to the high-level understanding of tumor evolution and timewise changes occurring in our genome. Finally, in routine clinical practice, DL could also be instrumental for screening purposes, such as in liquid biopsies for early cancer detection and disease monitoring.

Multimodality

Gathering extensive information prior to making decisions is not an exclusive trait of AI. This is also common within clinical workflows, where physicians rely on a range of data, such as basic patient information, medical records, and test results, to inform their decisions. For these reasons, the field of multimodal AI has emerged in recent years, where the inputs of the models originate from various data sources and output a single prediction. A few studies have investigated data fusion from histopathology and genomics data, capitalizing on potential synergies between these data modalities, ultimately aimed at clinical use. Histopathology images are widely available and inexpensive, but only show tissue phenotype, not necessarily underlying molecular changes. Therefore, it was shown that already the addition of clinical parameters from the patient could improve the generalizability of DL models improving the predictions [21]. Genomic methods, on the other hand, can offer a glimpse into the underlying machinery within the cells, but there is still the disadvantage that a certain amount of material is required to obtain such information, which is not always feasible. Furthermore, technical aspects also need to be considered, as in the case of DL, where the model's performance is critically dependent on the size of the input. Hence, the integration of data from different modalities could potentially allow for an increase in the information given to a model. With this, previously missing information can be completed or extended, refining the model's predictions and subsequently improving biomarkers [15, 98].

One of the first to publish a multimodal DL model combining histopathology and genomics was Mobadersany et al. [99] in 2018. They combined WSIs, *IDH* mutation, and 1p/19q codeletion status data as input of a ML model to predict survival for patients with gliomas (Fig. 2a). Furthermore, their method surpassed several clinical biomarkers for prognosis. One year later, Cheerla and Gevaert [100] utilized RNA expression data in combination with WSIs for 20 cancer types in order to improve survival predictions. The most recent evidence indicating that utilizing multiple modalities can be superior to single modalities was provided by Chen et al., who published two separate models: PathomicFusion (2019), which integrated WSIs, driver mutation, copy number variation, as well as RNA-sequencing data, and PORPOISE (2022), which added genomic profiles to WSIs [17, 101]. In terms of performance, PathomicFusion was able to reach a c-index of 0.826 in glioma and 0.72 in clear cell renal cell carcinoma survival prediction. In PORPOISE, the best performance was achieved in kidney renal clear cell carcinomas with a c-index of 0.827. However, external validation of these results might be needed before

clinically translating these models [102]. In addition to prognostication, other application types such as grading and subtyping were studied with multimodal models as well. Especially in brain cancer, many studies were carried out. For example, Pei et al. [103] predicted grading in gliomas based on the same features of Mobadersany et al. previously mentioned. This focus on brain cancer is likely due to the change in classification standards of gliomas in 2016, in which the World Health Organization added molecular features as decision standards to histopathological ones [104]. Thus, studies that would have solely relied on histopathology in the past, would now also require genomic evidence. In this way, clinical guidelines could facilitate multimodal research as well.

Adding another layer of multimodality, Boehm et al. [105] and Vanguri et al. [106] not only utilized histologic and genomic data but also expanded this repertoire by radiology images. With this, a next step towards a holistic integration of all clinically available information was taken, even though the complexity of these models would make their training and clinical deployment more difficult than single-modality models. Nevertheless, in a medical setting, having separate models for each data type will probably not be practical. Furthermore, in the future, it is possible that AI models not only incorporate patient data but also general medical information to make knowledge-based predictions. This could make them a universally applicable tool which combines predictions with practical reasoning that humans could interact with [107].

Outlook

As a result of technical advancements over the past years, DL models are continually becoming more powerful and generalizable. Given enough data and a clearly defined task, DL models can in principle outperform human observers in patient diagnosis and potentially in downstream decision-making processes [108, 109]. Nevertheless, some key limitations need to be overcome when applying DL to precision medicine [110].

In ML, models require sufficiently large amounts of data to become good at their task. Part of this requirement is for technical reasons, as many repetitions of patterns are required to force the internal model parameters into their desired state. Another reason for data requirements, however, is the variability that is present in any biological system. In particular, tumors are diverse as their genotype, phenotype, and clinical behavior differ between patients. The minimum size of any training data set is such that it can represent the biological variability. Therefore, studies which only contain a dozen participants, will usually not have sufficiently diverse data to generalize well to external datasets, particularly in clinical

routine [111]. In consequence, to make DL models available for a wide range of clinical settings, ever larger datasets need to be acquired and shared (Fig. 2b). Data collection, not model flexibility, is the main bottleneck in training DL solutions in cancer research and oncology. Histopathology, as the base of diagnosis, is more readily obtainable than genomic data, which is typically costly and not routinely acquired for all patients. Consequently, genomic cohorts are harder to establish, particularly for multi-omic approaches. Extensive clinical setups and infrastructure are required, often limiting them to well-funded research centers or large healthcare institutions. One way to address these challenges is through distributed learning such as federated or swarm learning, where peers that are prohibited from public data sharing can still jointly train models [112–114] (Fig. 2b). Furthermore, technical concepts could supplement data acquisition. Methods such as class balancing or augmenting datasets with simulated samples could aid studies with small patient numbers [115–117]. On the other hand, improved ML models could be more data-efficient and be able to sufficiently learn from even smaller datasets, potentially improving the data availability problem with a different strategy [118, 119].

In addition to limitations in dataset size, another fundamental problem of the development and deployment of DL systems in healthcare is that many datasets contain an internal bias based on the ethnicity, sex, or socio-economic circumstances of participants, or the institution in which the studies were conducted [120–122]. Consequently, this calls for fairer and more diverse data acquisition strategies for upcoming studies which, in reverse, would have a positive impact on the generalizability of DL models again (Fig. 2b). In addition, even in homogenous data, standards for data curation need to be established nationally and internationally to make data comparable between institutions in the first place (Fig. 2b). Furthermore, since changes can occur within populations AI is used upon, we will encounter the necessity for model updates and reconfigurations, a property mostly not considered in model design today (Fig. 2b) [123]. This will eventually allow obtaining DL models that dynamically learn during deployment, rather than being "frozen" after a single static training step.

Ultimately, the aim of the research presented in this review is to implement DL in actual clinical routines. Unfortunately, this is notoriously challenging, as most countries mandate a necessary but highly complex regulatory approval. Obtaining such regulatory approval is not attainable for academic teams, only for commercial enterprises with quality-controlled development workflows and the financial means to bring an algorithm to the market as a product [124]. Even after gaining

approval, there are other additional challenges to overcome. For instance, few healthcare institutions even in the most economically prosperous countries are fully digitalized. Particularly, histopathology is based on the manual handling of glass slides in the overwhelming majority of healthcare institutions in the US and the EU today [110] (Fig. 2b). Moreover, a new skillset in healthcare providers and technical assistants is also needed to ensure processes are running efficiently. In the future, substantial investments are required to make healthcare infrastructure ready for a routine deployment of DL-based biomarkers (Fig. 2b).

Finally, for DL to be adopted by practitioners, the models should ideally not be considered as a "black box", but also inherit the explainability for their decisions (Fig. 2b) [125]. This challenge is difficult to address since DL models exhibit a high degree of complexity and are often susceptible to minor changes in the input data, making it difficult to ensure reliable and consistent outputs [126]. A number of established techniques exist which are often used to make models explainable. For histopathology, these include mostly two types: "saliency maps," which highlight parts of the input data that were relevant for decision-making, and "extreme examples," i.e., extracting the instances in the dataset that are assigned the highest and lowest prediction scores by the model [127]. In clinical genomics, particularly for tabular data, explainability methods such as Local Interpretable Model-agnostic Explanations (LIME) [128] or SHapley Additive exPlanations (SHAP) [129] values can indicate to which extent features influence predictions. However, the benefit of these methods depends on the human interpretability of the features themselves [130]. Furthermore, these approaches do not necessarily infer causality which shows that we are only at the beginning of this development. In addition to the explainability of specific models, generative AI could change the way we perceive what DL actually learns by reversing the DL workflow, creating data from an input query (Fig. 2b) [131]. More importantly, generative DL models could allow us to integrate counterfactuality. Essentially, as a first step, large DL models gather large and diverse knowledge about biological processes. Then, in counterfactual methods, the generative DL part can be used by a human experimentalist to answer questions such as "what would this particular tumor look like if it had a *BRAF* mutation?", or "what would this precise tumor look like if the lymphocytes were removed?" [132, 133]. These approaches are not widely investigated in the analysis of pathology images or genomic data of cancer, but could be a useful tool for educational purposes and search for yet unknown properties.

In conclusion, the incorporation of AI into patient care is a multifaceted endeavor that requires extensive collaboration of researchers, healthcare institutions, and administrative bodies. The strategies explored in this review have the potential to enhance personalized treatments and advance precision oncology, possibly yielding cost savings and improved outcomes for patients. The rapid evolution of DL is remarkable, especially considering that just a decade ago it had virtually no role in the analysis of clinical data at all. Therefore, we anticipate that DL will become a widely used component of clinical workflows in precision oncology.

Abbreviations

AI	Artificial intelligence
AUROC	Area under the receiver operating characteristic
CNN	Convolutional neural network
FDA	Food and drug administration
H&E	Hematoxylin and eosin
ML	Machine learning
qPCR	Quantitative polymerase chain reaction
SVM	Support vector machine
TCGA	The Cancer Genome Atlas
WHO	World Health Organization
WSI	Whole slide image

Availability of data and materials
Not applicable

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
JNK declares consulting services for Owkin, France; DoMore Diagnostics, Norway; Panakeia, UK; Scailyte, Switzerland; Mindpeak, Germany; and MultiplexDx, Slovakia. Furthermore he holds shares in StratifAI GmbH, Germany, has received a research grant by GSK, and has received honoraria by AstraZeneca,

References

1. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol. 2019;16:703–15.
2. Djuric U, Zadeh G, Aldape K, Diamandis P. Precision histology: how deep learning is poised to revitalize histomorphology for personalized cancer care. NPJ Precis Oncol. 2017;1:22.
3. Liu R, Zou J. Advancing precision oncology with large, real-world genomics and treatment outcomes data. Nat Med. 2022;28:1544–5.
4. Andre F, Filleron T, Kamal M, Mosele F, Arnedos M, Dalenc F, et al. Genomics to select treatment for patients with metastatic breast cancer. Nature. 2022;610:343–8.
5. Kato S, Kim KH, Lim HJ, Boichard A, Nikanjam M, Weihe E, et al. Real-world data from a molecular tumor board demonstrates improved outcomes with a precision N-of-One strategy. Nat Commun. 2020;11:4965.
6. Pantziarka P, Capistrano IR, De Potter A, Vandeborne L, Bouche G. An Open Access Database of Licensed Cancer Drugs. Front Pharmacol. 2021;12:627574.
7. BigScience Workshop, Le Scao T, Fan A, Akiki C, Pavlick E, et al. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. 2022. http://arxiv.org/abs/2211.05100
8. Zhao Z-Q, Zheng P, Xu S-T, Wu X. Object Detection With Deep Learning: A Review. IEEE Trans Neural Netw Learn Syst. 2019;30:3212–32.
9. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.
10. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. Nat Rev Clin Oncol. 2021;19(2):132–46.
11. Hecht-Nielsen R. Theory of the backpropagation neural network. International 1989 Joint Conference on Neural Networks. 1989;1:593–605.
12. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44.
13. Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. Nat Cancer. 2022;3:1026–38.
14. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med. 2021;13:152.
15. Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, et al. Artificial intelligence for multimodal data integration in oncology. Cancer Cell. 2022;40:1095–110.
16. Sammut S-J, Crispin-Ortuzar M, Chin S-F, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature. 2022;601:623–9.
17. Chen RJ, Lu MY, Williamson DFK, Chen TY, Lipkova J, Noor Z, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell. 2022;865–78.e6. https://doi.org/10.1016/j.ccell.2022.07.004.
18. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer. 2021;124:686–96.
19. Cifci D, Foersch S, Kather JN. Artificial intelligence to identify genetic alterations in conventional histopathology. J Pathol. 2022;257(4):430–44. https://doi.org/10.1002/path.5898.
20. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17:195.
21. Niehues JM, Quirke P, West NP, Grabsch HI, van Treeck M, Schirris Y, et al. Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: A retrospective multi-centric study. Cell Rep Med. 2023;4(4):100980.
22. Ilse M, Tomczak J, Welling M. Attention-based Deep Multiple Instance Learning. In: Dy J, Krause A, editors. Proceedings of the 35th International Conference on Machine Learning. PMLR; 2018. p. 2127–2136.
23. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med. 2018;24:1559–67.
24. Wagner SJ, Reisenbüchler D, West NP, Niehues JM, Veldhuizen GP, Quirke P, et al. Fully transformer-based biomarker prediction from colorectal cancer histology: a large-scale multicentric study. 2023. http://arxiv.org/abs/2301.09617
25. Jiang S, Zanazzi GJ, Hassanpour S. Predicting prognosis and IDH mutation status for patients with lower-grade gliomas using whole slide images. Sci Rep. 2021;11:16849.
26. Ertosun MG, Rubin DL. Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks. AMIA Annu Symp Proc. 2015;2015:1899–908.
27. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NNC, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Sci Rep. 2017;7(1):46450. https://doi.org/10.1038/srep46450.
28. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med. 2019;25:1301–9.
29. Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. Lancet Oncol. 2020;21:222–32.
30. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol. 2020;21:233–41.
31. Sirinukunwattana K, Domingo E, Richman SD, Redmond KL, Blake A, Verrill C, et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut. 2021;70:544–54.
32. Jaber MI, Song B, Taylor C, Vaske CJ, Benz SC, Rabizadeh S, et al. A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. Breast Cancer Res. 2020;22:12.
33. Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer. 2020;1:800–10.
34. Danesi R, Fogli S, Indraccolo S, Del Re M, Dei Tos AP, Leoncini L, et al. Druggable targets meet oncogenic drivers: opportunities and limitations of target-based classification of tumors and the role of Molecular Tumor Boards. ESMO Open. 2021;6:100040.
35. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med. 2019;25:1054–6.
36. Echle A, Grabsch HI, Quirke P, van den Brandt PA, West NP, Hutchins GGA, et al. Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. Gastroenterology. 2020;159:1406–16.e11.
37. Saillard C, Dubois R, Tchita O, Loiseau N, Garcia T, Adriansen A, et al. Validation of MSIntuit as an AI-based pre-screening tool for MSI detection from colorectal cancer histology slides. Nat Commun. 2023;14:6695.
38. Kather JN, Heij LR, Grabsch HI, Loeffler C, Echle A, Muti HS, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer. 2020;1:789–99.
39. Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. Nat Commun. 2020;11(1):3877. https://doi.org/10.1038/s41467-020-17678-4.
40. Zeng Q, Klein C, Caruso S, Maille P, Laleh NG, Sommacale D, et al. Artificial intelligence predicts immune and inflammatory gene

signatures directly from hepatocellular carcinoma histology. J Hepatol. 2022;77(1):116–27.

41. Angell H, Galon J. From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer. Curr Opin Immunol. 2013;25:261–7.

42. Brummel K, Eerkens AL, de Bruyn M, Nijman HW. Tumour-infiltrating lymphocytes: from prognosis to treatment selection. Br J Cancer. 2023;128:451–8.

43. Kashiwagi S, Asano Y, Goto W, Takada K, Takahashi K, Noda S, et al. Use of Tumor-infiltrating lymphocytes (TILs) to predict the treatment response to eribulin chemotherapy in breast cancer. PloS One. 2017;12:e0170634.

44. Sharma P, Shen Y, Wen S, Yamada S, Jungbluth AA, Gnjatic S, et al. CD8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. Proc Natl Acad Sci U S A. 2007;104:3967–72.

45. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. 2006;313:1960–4.

46. Diao JA, Wang JK, Chui WF, Mountain V, Gullapally SC, Srinivasan R, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. Nat Commun. 2021;12:1613.

47. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, et al. Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. Cell Rep. 2018;23:181–93.e7.

48. Liu K, Xia W, Qiang M, Chen X, Liu J, Guo X, et al. Deep learning pathological microscopic features in endemic nasopharyngeal cancer: Prognostic value and protentional role for individual induction chemotherapy. Cancer Med. 2020;9:1298–306.

49. Li F, Yang Y, Wei Y, He P, Chen J, Zheng Z, et al. Deep learning-based predictive biomarker of pathological complete response to neoadjuvant chemotherapy from histological images in breast cancer. J Transl Med. 2021;19:348.

50. Johannet P, Coudray N, Donnelly DM, Jour G, Illa-Bochaca I, Xia Y, et al. Using Machine Learning Algorithms to Predict Immunotherapy Response in Patients with Advanced Melanoma. Clin Cancer Res. 2021;27:131–40.

51. Wang S, Chen A, Yang L, Cai L, Xie Y, Fujimoto J, et al. Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. Sci Rep. 2018;8:10393.

52. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis C-A, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS Med. 2019;16:e1002730.

53. Wessels F, Schmitt M, Krieghoff-Henning E, Kather JN, Nientiedt M, Kriegmair MC, et al. Deep learning can predict survival directly from histology in clear cell renal cell carcinoma. PloS One. 2022;17:e0272656.

54. Yao J, Zhu X, Jonnagaddala J, Hawkins N, Huang J. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Med Image Anal. 2020;65:101789.

55. Liu H, Kurc T. Deep learning for survival analysis in breast cancer with whole slide image data. Bioinformatics. 2022;38:3629–37.

56. Li X, Jonnagaddala J, Yang S, Zhang H, Xu XS. A retrospective analysis using deep-learning models for prediction of survival outcome and benefit of adjuvant chemotherapy in stage II/III colorectal cancer. J Cancer Res Clin Oncol. 2022;148:1955–63.

57. Ghaffari Laleh N, Muti HS, Loeffler CML, Echle A, Saldanha OL, Mahmood F, et al. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. Med Image Anal. 2022;79:102474.

58. Gupta L, Klinkhammer BM, Seikrit C, Fan N, Bouteldja N, Gräbel P, et al. Large-scale extraction of interpretable features provides new insights into kidney histopathology - A proof-of-concept study. J Pathol Inform. 2022;13:100097.

59. Anghel A, Stanisavljevic M, Andani S, Papandreou N, Rüschoff JH, Wild P, et al. A High-Performance System for Robust Stain Normalization of Whole-Slide Images in Histopathology. Front Med. 2019;6:193.

60. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. 2017. http://arxiv.org/abs/1706.03762

61. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 2020. http://arxiv.org/abs/2010.11929

62. Chen RJ, Chen C, Li Y, Chen TY, Trister AD, Krishnan RG, et al. Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning. 2022. http://arxiv.org/abs/2206.02647

63. Center for Devices, Radiological Health. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. U.S. Food and Drug Administration. FDA; 2023. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

64. Kumar-Sinha C, Chinnaiyan AM. Precision oncology in the age of integrative genomics. Nat Biotechnol. 2018;36:46–60.

65. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013;3:246–59.

66. Zhao Y, Pan Z, Namburi S, Pattison A, Posner A, Balachander S, et al. CUP-AI-Dx: A tool for inferring cancer tissue of origin and molecular subtype using RNA gene-expression data and artificial intelligence. EBioMedicine. 2020;61:103030.

67. Jiao W, Atwal G, Polak P, Karlic R, Cuppen E, PCAWG Tumor Subtypes and Clinical Translation Working Group, et al. A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns. Nat Commun. 2020;11:728.

68. Sienkiewicz K, Chen J, Chatrath A, Lawson JT, Sheffield NC, Zhang L, et al. Detecting molecular subtypes from multi-omics datasets using SUMO. Cell Rep Methods. 2022;2(1) https://doi.org/10.1016/j.crmeth.2021.100152.

69. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21:1350–6.

70. Yuan Y, Shi Y, Li C, Kim J, Cai W, Han Z, et al. DeepGene: an advanced cancer type classifier based on deep learning and somatic point mutations. BMC Bioinformatics. 2016;17:476.

71. Tian J, Zhu M, Ren Z, Zhao Q, Wang P, He CK, et al. Deep learning algorithm reveals two prognostic subtypes in patients with gliomas. BMC Bioinformatics. 2022;23:417.

72. Yuan Y, Shi Y, Su X, Zou X, Luo Q, Feng DD, et al. Cancer type prediction based on copy number aberration and chromatin 3D structure with convolutional neural networks. BMC Genomics. 2018;19:565.

73. Zhao L, Lee VHF, Ng MK, Yan H, Bijlsma MF. Molecular subtyping of cancer: current status and moving toward clinical applications. Brief Bioinform. 2019;20:572–84.

74. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.

75. de Ligt J, Boone PM, Pfundt R, Vissers LELM, Richmond T, Geoghegan J, et al. Detection of clinically relevant copy number variants with whole-exome sequencing. Hum Mutat. 2013;34:1439–48.

76. Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. Nat Rev Cancer. 2020;20:555–72.

77. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biol. 2016;17:128.

78. Kim S, Lee H, Kim K, Kang J. Mut2Vec: distributed representation of cancerous mutations. BMC Med Genomics. 2018;11:33.

79. Luzzatto L. Somatic mutations in cancer development. Environ Health. 2011;10(Suppl 1):S12.

80. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

81. Koboldt DC. Best practices for variant calling in clinical sequencing. Genome Med. 2020;12:91.

82. Barbitoff YA, Abasov R, Tvorogova VE, Glotov AS, Predeus AV. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. BMC Genomics. 2022;23:155.

83. Sahraeian SME, Fang LT, Karagiannis K, Moos M, Smith S, Santana-Quintero L, et al. Achieving robust somatic mutation detection with deep

learning models derived from reference data sets of a cancer sample. Genome Biol. 2022;23:12.

84. Krishnamachari K, Lu D, Swift-Scott A, Yeraliyev A, Lee K, Huang W, et al. Accurate somatic variant detection using weakly supervised deep learning. Nat Commun. 2022;13:4248.

85. Chang Y, Park H, Yang H-J, Lee S, Lee K-Y, Kim TS, et al. Cancer Drug Response Profile scan (CDRscan): A Deep Learning Model That Predicts Drug Effectiveness from Cancer Genomic Signature. Sci Rep. 2018;8:8857.

86. Chiu Y-C, Chen H-IH, Zhang T, Zhang S, Gorthi A, Wang L-J, et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. BMC Med Genomics. 2019;12:18.

87. Kim D, Li R, Dudek SM, Ritchie MD. ATHENA: Identifying interactions between different levels of genomic data associated with cancer clinical outcomes using grammatical evolution neural network. BioData Min. 2013;6:23.

88. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. Clin Cancer Res. 2018;24:1248–59.

89. Elmarakeby HA, Hwang J, Arafeh R, Crowdis J, Gang S, Liu D, et al. Biologically informed deep neural network for prostate cancer discovery. Nature. 2021;598(7880):348–52. https://doi.org/10.1038/s41586-021-03922-4.

90. Yousefi S, Amrollahi F, Amgad M, Dong C, Lewis JE, Song C, et al. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. Sci Rep. 2017;7:11707.

91. Zuo Z, Wang P, Chen X, Tian L, Ge H, Qian D. SWnet: a deep learning model for drug response prediction from cancer genomic signatures and compound chemical structures. BMC Bioinformatics. 2021;22:434.

92. Wang S, Zhang H, Liu Z, Liu Y. A Novel Deep Learning Method to Predict Lung Cancer Long-Term Survival With Biological Knowledge Incorporated Gene Expression Images and Clinical Data. Front Genet. 2022;13:800853.

93. Li M-X, Sun X-M, Cheng W-G, Ruan H-J, Liu K, Chen P, et al. Using a machine learning approach to identify key prognostic molecules for esophageal squamous cell carcinoma. BMC Cancer. 2021;21:906.

94. Ma B, Meng F, Yan G, Yan H, Chai B, Song F. Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. Comput Biol Med. 2020;121:103761.

95. Zhang T-H, Hasib MM, Chiu Y-C, Han Z-F, Jin Y-F, Flores M, et al. Transformer for Gene Expression Modeling (T-GEM): An Interpretable Deep Learning Model for Gene Expression-Based Phenotype Predictions. Cancers. 2022;14(19):4763. https://doi.org/10.3390/cancers14194763.

96. Cai SF, Levine RL. 15 years after a giant leap for cancer genomics. Nature. 2023;623:920–1.

97. Pritchard D, Goodman C, Nadauld LD. Clinical Utility of Genomic Testing in Cancer Care. JCO Precis Oncol. 2022;6:e2100349.

98. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer. 2022;22:114–26.

99. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Vega JEV, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. PNAS. 2018;115(13):E2970–9. https://doi.org/10.1101/198010.

100. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. Bioinformatics. 2019;35(14):i446–54. https://doi.org/10.1093/bioinformatics/btz342.

101. Chen RJ, Lu MY, Wang J, Williamson DFK, Rodig SJ, Lindeman NI, et al. Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis. IEEE Trans Med Imaging. 2022;41(4):757–70. https://doi.org/10.1109/tmi.2020.3021387.

102. Howard FM, Kather JN, Pearson AT. Multimodal deep learning: An improvement in prognostication or a reflection of batch effect? Cancer Cell. 2023;41:5–6.

103. Pei L, Jones KA, Shboul ZA, Chen JY, Iftekharuddin KM. Deep Neural Network Analysis of Pathology Images With Integrated Molecular Data for Enhanced Glioma Classification and Grading. Front Oncol. 2021;11:668694. https://doi.org/10.3389/fonc.2021.668694.

104. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 2016;131:803–20.

105. Boehm KM, Aherne EA, Ellenson L, Nikolovski I, Alghamdi M, Vázquez-García I, et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. Nat Cancer. 2022;3:723–33.

106. Vanguri RS, Luo J, Aukerman AT, Egger JV, Fong CJ, Horvat N, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. Nat Cancer. 2022;3:1151–64.

107. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616:259–65.

108. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577:89–94.

109. Hung J-Y, Chen K-W, Perera C, Chiu H-K, Hsu C-R, Myung D, et al. An Outperforming Artificial Intelligence Model to Identify Referable Blepharoptosis for General Practitioners. J Pers Med. 2022;12(2):283. https://doi.org/10.3390/jpm12020283.

110. Reis-Filho JS, Kather JN. Overcoming the challenges to implementation of artificial intelligence in pathology. J Natl Cancer Inst. 2023;115(6):608–12. https://doi.org/10.1093/jnci/djad048.

111. Eche T, Schwartz LH, Mokrane F-Z, Dercle L. Toward Generalizability in the Deployment of Artificial Intelligence in Radiology: Role of Computation Stress Testing to Overcome Underspecification. Radiol Artif Intell. 2021;3:e210097.

112. Warnat-Herresthal S, Schultze H, Shastry KL, Manamohan S, Mukherjee S, Garg V, et al. Swarm Learning for decentralized and confidential clinical machine learning. Nature. 2021;594:265–70.

113. Lu MY, Chen RJ, Kong D, Lipkova J, Singh R, Williamson DFK, et al. Federated learning for computational pathology on gigapixel whole slide images. Med Image Anal. 2022;76:102298.

114. Saldanha OL, Quirke P, West NP, James JA, Loughrey MB, Grabsch HI, et al. Swarm learning for decentralized artificial intelligence in cancer histopathology. Nat Med. 2022;28:1232–9.

115. Ding K, Zhou M, Wang H, Gevaert O, Metaxas D, Zhang S. A Large-scale Synthetic Pathological Dataset for Deep Learning-enabled Segmentation of Breast Cancer. Sci Data. 2023;10:231.

116. Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst J-M, Ciompi F, et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med Image Anal. 2019;58:101544.

117. Lee NK, Tang Z, Toneyan S, Koo PK. EvoAug: improving generalization and interpretability of genomic deep neural networks with evolution-inspired data augmentations. Genome Biol. 2023;24:105.

118. Panayides AS, Amini A, Filipovic ND, Sharma A, Tsaftaris SA, Young A, et al. AI in Medical Imaging Informatics: Current Challenges and Future Directions. IEEE J Biomed Health Inform. 2020;24:1837–57.

119. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng. 2021;5:555–70.

120. Dehon E, Weiss N, Jones J, Faulconer W, Hinton E, Sterling S. A Systematic Review of the Impact of Physician Implicit Racial Bias on Clinical Decision Making. Acad Emerg Med. 2017;24:895–904.

121. Schulman KA, Berlin JA, Harless W, Kerner JF, Sistrunk S, Gersh BJ, et al. The effect of race and sex on physicians' recommendations for cardiac catheterization. N Engl J Med. 1999;340:618–26.

122. Howard FM, Dolezal J, Kochanny S, Schulte J, Chen H, Heij L, et al. The impact of site-specific digital histology signatures on deep learning model accuracy and bias. Nat Commun. 2021;12:4423.

123. Schnellinger EM, Yang W, Kimmel SE. Comparison of dynamic updating strategies for clinical prediction models. Diagn Progn Res. 2021;5:20.

124. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. Lancet Digit Health. 2021;3:e195–203.

125. Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. Proc Natl Acad Sci U S A. 2019;116:22071–80.

126. Ghaffari Laleh N, Truhn D, Veldhuizen GP, Han T, van Treeck M, Buelow RD, et al. Adversarial attacks and adversarial robustness in computational pathology. Nat Commun. 2022;13:5711.
127. Evans T, Retzlaff CO, Geißler C, Kargl M, Plass M, Müller H, et al. The explainability paradox: Challenges for xAI in digital pathology. Future Gener Comput Syst. 2022;133:281–96.
128. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. 2016. http://arxiv.org/abs/1602.04938
129. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. 2017. http://arxiv.org/abs/1705.07874
130. Yap M, Johnston RL, Foley H, MacDonald S, Kondrashova O, Tran KA, et al. Verifying explainability of a deep learning tissue classifier trained on RNA-seq data. Sci Rep. 2021;11:2641.
131. Jose L, Liu S, Russo C, Nadort A, Di Ieva A. Generative Adversarial Networks in Digital Pathology and Histopathological Image Processing: A Review. J Pathol Inform. 2021;12:43.
132. Mertes S, Huber T, Weitz K, Heimerl A, André E. GANterfactual-Counterfactual Explanations for Medical Non-experts Using Generative Adversarial Learning. Front Artif Intell. 2022;5:825565.
133. Wang C, Li J, Zhang F, Sun X, Dong H, Yu Y, et al. Bilateral Asymmetry Guided Counterfactual Generating Network for Mammogram Classification. IEEE Trans Image Process. 2021;30:7980–94.

Publisher's Note