

Application of digital pathology–based advanced analytics of tumour microenvironment organisation to predict prognosis and therapeutic response

Xiao Fu^{1,2*}, Erik Sahai¹ and Anna Wilkins^{1,3,4*} 

¹ Tumour Cell Biology Laboratory, The Francis Crick Institute, London, UK

² Biomolecular Modelling Laboratory, The Francis Crick Institute, London, UK

³ Division of Radiotherapy and Imaging, Institute of Cancer Research, London, UK

⁴ Royal Marsden Hospitals NHS Trust, London, UK

*Correspondence to: X Fu or A Wilkins, Tumour Cell Biology Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT, UK. E-mail: xiao.fu@crick.ac.uk (X Fu) or anna.wilkins@cr.ac.uk (A Wilkins)

Abstract

In recent years, the application of advanced analytics, especially artificial intelligence (AI), to digital H&E images, and other histological image types, has begun to radically change how histological images are used in the clinic. Alongside the recognition that the tumour microenvironment (TME) has a profound impact on tumour phenotype, the technical development of highly multiplexed immunofluorescence platforms has enhanced the biological complexity that can be captured in the TME with high precision. AI has an increasingly powerful role in the recognition and quantitation of image features and the association of such features with clinically important outcomes, as occurs in distinct stages in conventional machine learning. Deep-learning algorithms are able to elucidate TME patterns inherent in the input data with minimum levels of human intelligence and, hence, have the potential to achieve clinically relevant predictions and discovery of important TME features. Furthermore, the diverse repertoire of deep-learning algorithms able to interrogate TME patterns extends beyond convolutional neural networks to include attention-based models, graph neural networks, and multimodal models. To date, AI models have largely been evaluated retrospectively, outside the well-established rigour of prospective clinical trials, in part because traditional clinical trial methodology may not always be suitable for the assessment of AI technology. However, to enable digital pathology–based advanced analytics to meaningfully impact clinical care, specific measures of ‘added benefit’ to the current standard of care and validation in a prospective setting are important. This will need to be accompanied by adequate measures of explainability and interpretability. Despite such challenges, the combination of expanding datasets, increased computational power, and the possibility of integration of pre-clinical experimental insights into model development means there is exciting potential for the future progress of these AI applications.

© 2023 The Authors. *The Journal of Pathology* published by John Wiley & Sons Ltd on behalf of The Pathological Society of Great Britain and Ireland.

Keywords: advanced analytics; digital pathology; tumour microenvironment; artificial intelligence; biomarker

Received 10 March 2023; Revised 25 May 2023; Accepted 7 June 2023

Conflict of interest statement: ES receives research funding from Merck Sharp Dohme, AstraZeneca, consults for Theolytics, and is on the scientific advisory board of Phenomic AI. AW receives funding from AstraZeneca and imCORE.

Introduction

Visual interpretation of H&E-stained tissue sections by specialist pathologists using an optical microscope has been the cornerstone of diagnostic oncology since the late 19th century [1]. Over several decades, detailed descriptions of grading systems, such as the Gleason score in prostate cancer or histological grade in breast cancer, have been developed and refined, often involving international working groups of pathologists establishing consensus statements [2–6]. Similarly, histological variants within tumour types and tumour

hallmarks known to represent indolent versus aggressive behaviour have been described and validated, collectively representing a huge wealth of pathological knowledge [7–9]. Yet in the last 10 years, the application of advanced analytics, especially artificial intelligence (AI), to digital H&E images and other histological image types has begun to radically change how histological images are used in the clinic.

Innovation in advanced analytics, especially AI, has typically occurred outside of oncology but has been rapidly transferred from other disciplines to diagnostic, prognostic or predictive use in oncology. The automated

extraction of quantitative data and the elucidation of image patterns not detectable by the human eye are both important areas of additional benefit over microscopic assessment by a pathologist. Recent years have seen a number of AI-based pathology products obtain regulatory approval, including from the FDA [10–12]. In a large clinical study across more than 100 institutions, use of a prostate cancer diagnostic AI algorithm was shown to increase the sensitivity of cancer detection and reduce both false positives and false negatives [10]. For a cancer that is very common globally, and for which diagnosis requires meticulous examination of large tissue areas to avoid missing small tumour foci, AI approaches have a number of benefits. In the developed world, this includes a reduction in pathologist workload and fatigue, thus enabling pathologists to focus on the description of tumour foci identified by the AI detection system. In the developing world, the combination of innovative smartphone technology and appropriate validated AI algorithms could help to overcome the substantial clinically important lack in pathology expertise and microscope facilities.

The tumour microenvironment (TME) encompasses the wider cellular and acellular milieu in which tumour cells reside. It primarily consists of innate and adaptive immune cells alongside the tumour vasculature and fibroblasts, which, together with tumour cells, are supported by an extensive ECM. The TME has a pivotal role in shaping tumour phenotype, evolutionary dynamics, and therapy responses [13]. As a result, characterising the spatial relationships between specific cell populations in the TME using digital pathology tools and analytical approaches that are able to capture the TME complexity is a research priority.

This review will present the wide range of digital pathology images to which advanced analytics are currently being applied, as well as the methodology for such analytics, especially those based on machine learning. Finally, we will discuss the feasibility and challenges of widespread incorporation of AI approaches to clinical practice.

Image types for advanced analytics

There is a wide range of image types to which AI algorithms are increasingly being applied. These include H&E-stained sections – a simple and inexpensive part of standard diagnostic procedures that is widely available in routine pathology laboratories worldwide but lacks unambiguous cell type information. Considerable progress has been made in AI algorithms using H&E images, as discussed earlier [11,14,15]. Here, computational extraction of cell and nuclear size, as well as their shape, is fairly straightforward and precise identification of tumour cells versus fibroblast versus leukocyte is increasingly possible, e.g. using digital pathology image analysis tools such as QuPath [16] or convolutional neural networks (CNNs) [17]. The haematoxylin stain also enables the detection of specific chromatin structures [18,19]. Furthermore, it is increasingly

possible to link H&E images with the genome using AI algorithms, as discussed later. Immunohistochemical stains, e.g. for CD8+ T cells or cytokeratins for tumour cells, can be helpful in validating cell detection algorithms using H&E, as well as being useful in their own right for quantitative analysis. Non-cellular components of the TME can be visualised using chemical stains such as Picrosirius red, Masson's trichrome or Gomori trichrome, which specifically detect the collagen ECM or immunofluorescence for other ECM components, such as fibronectin, as exemplified in Figure 1. Both immunohistochemistry and collagen stains are simple and widely available but are limited in that they characterise very few (typically one or two) components of the TME in each section.

With the recognition that the TME has a profound impact on tumour phenotype, construction of tissue atlases encompassing a range of cell types at subcellular resolution is increasingly important. Highly multiplexed platforms enable the detection of over 100 antigens on a single tissue section, with the obvious advantage of capturing the spatial complexity of the TME with much greater precision. Important multiplex platforms include imaging mass cytometry (IMC), Phenocycler (formerly CODEX), Akoya Biosciences (Marlborough, MA, USA), and VECTRA Polaris, Akoya Biosciences (Table 1). Of clinical relevance, formalin-fixed paraffin-embedded tissue can be used for all three platforms. IMC consists of high-resolution ($1 \mu\text{m}^2$) laser ablation and cytometry by time of flight to detect up to 40 antigens labelled with antibodies conjugated to metal tags [32]. Multiplexed ion beam imaging time of flight (MIBI-TOF) is technically very similar to IMC, including the number of targets that can be visualised, but uses a tuneable ion beam that can be adjusted for tissue depth instead of a laser for tissue ablation [33]. Phenocycler uses fluorescent oligonucleotide-based tagging of antibodies, which are sequentially hybridised and dehybridised across multiple cycles; automated microscopy is able to detect over 100 targets. In VECTRA Polaris, a secondary antibody is fused to a fluorescent opal dye and using up to six serial antigen retrieval cycles, six different targets plus a counterstain can be visualised on a single section [25]. Across platforms, there is a trade-off between the number of targets visualised and the quantity of tissue that can practically be imaged. As a result, whole-slide imaging using ultra highly multiplexed systems is considerably slower than with VECTRA Polaris; this is of practical relevance for discovery science and a very important consideration for the clinic, where inexpensive high throughput staining and scanning is essential.

Concomitant with the advances in imaging techniques, innovative imaging data analysis and machine-learning algorithms have been developed in the last couple of decades to assist in tumour diagnosis and understanding of clinically relevant tumour and microenvironmental features [1,12,34] (Figure 2). In the following sections we discuss recent progress and achievements in this area, with a focus on hand-crafted

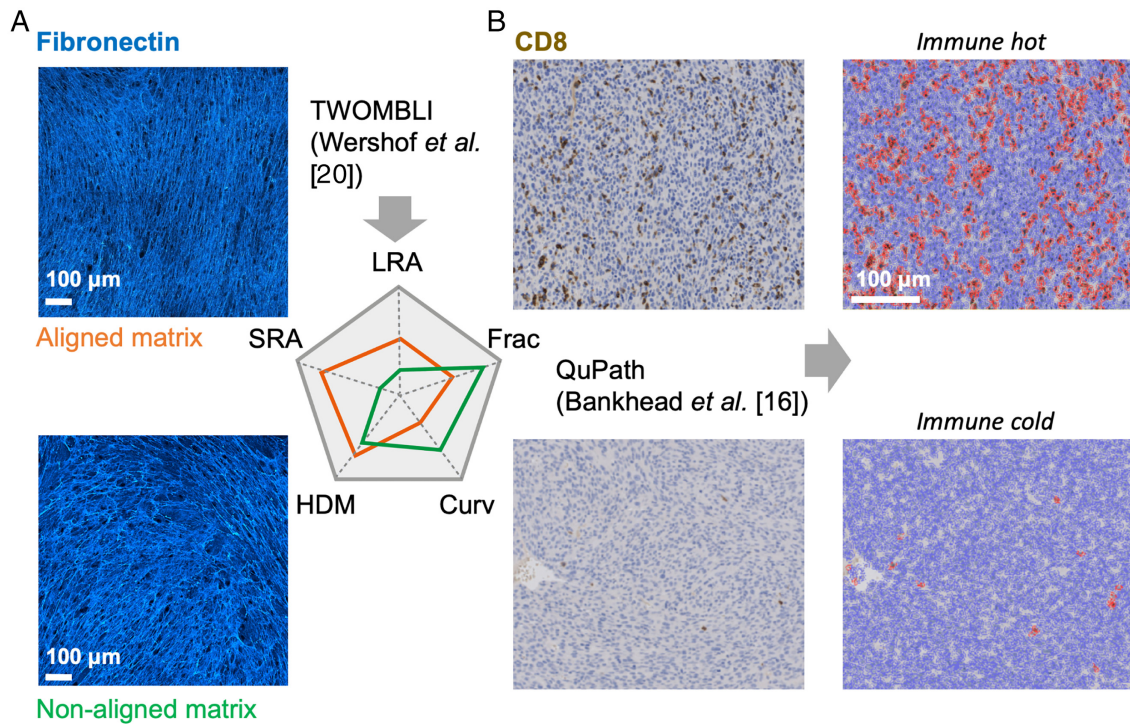


Figure 1. Exemplar analyses of quantitative feature extraction. (A) Quantitative matrix features extracted using TWOMBLI based on immunofluorescence for fibronectin. (B) Density of T cells extracted using QuPath based on CD8 staining. Curv, curvature; Frac, fractal dimension; HDM, high-density matrix; LRA, long-range alignment; SRA, short-range alignment.

feature engineering and deep-learning studies that discovered clinically relevant TME features.

What quantitative approaches are possible using digital images?

A range of machine-learning models are used in the quantitative analysis of digital images. These range from less complex regression models to considerably more complex AI approaches. AI can play a role at different stages of the process of linking images to clinical parameters. First, AI can be trained to recognise features in

images and extract metrics of those features. Second, features can be extracted using ‘rules-based’ algorithms and AI then used to relate the output metrics to clinically important features/outcomes as conventional machine learning. Finally, AI can be used to perform both steps together, i.e. there is no need for prior feature extraction, this is referred to as deep learning.

Machine-learning applications can be categorised as supervised, weakly supervised, and unsupervised, depending on the extent and type of data annotation [35]. Whereas in a supervised method the label needs to be provided for every data point (e.g. pixel-level annotation

Table 1. Types of image used in digital pathology.

Type of image	Brief description	Number of detectable targets	Examples
H&E	Routinely used stain in which haematoxylin precisely stains nuclear components, including heterochromatin and nucleoli, whereas eosin stains cytoplasmic components, including collagen and elastic fibres, muscle fibres, and red blood cells.	Unspecified	[10,11,14]
Picrosirius red staining	A widely used histological technique to visualise the distribution of collagen. The stain highlights the natural birefringence of collagen fibres when exposed to polarised light, enabling a detailed study of collagen organisation.	1	[20,21]
Immunohistochemistry	A commonly used test in which an antibody detects a specific antigen or marker in a sample of tissue. The antibody is typically linked to an enzyme or a fluorescent dye, which is activated to enable visualisation using a microscope or digital scanner.	1–2	[22–24]
VECTRA Polaris	A secondary antibody is fused to a fluorescent opal dye and using up to six serial antigen retrieval cycles, six different targets plus a counterstain can be visualised on a single section	Up to 6	[25,26]
IMC	High-resolution (1 µm ²) laser ablation and cytometry by time of flight is combined to detect up to 40 antigens labelled with antibodies conjugated to metal tags	Up to 40	[27–29]
Phenocycler (formerly CODEX)	Uses oligonucleotide-based tagging of antibodies, which are sequentially hybridised and dehybridised across multiple cycles	Over 100	[30,31]

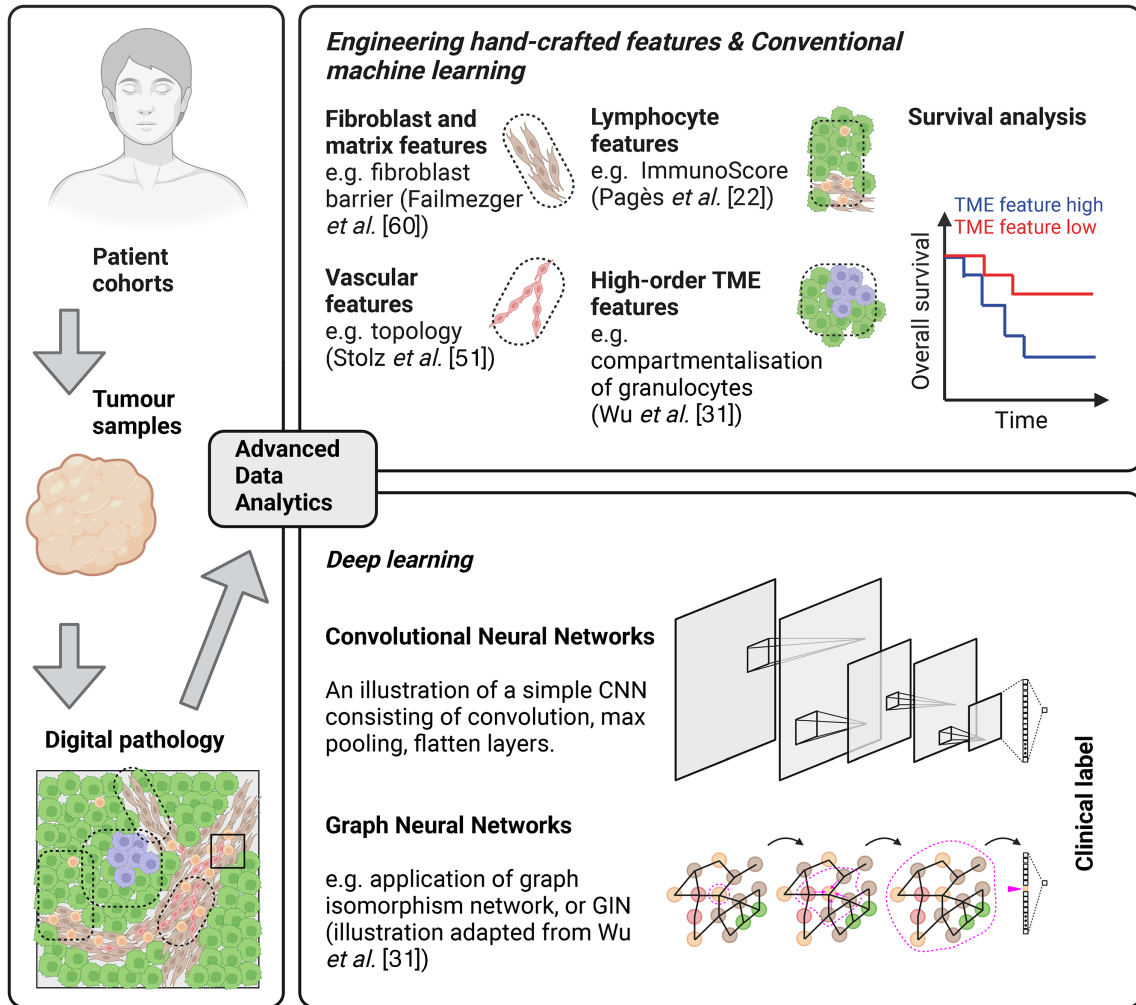


Figure 2. Prediction of clinical outcomes via conventional versus deep learning. Created with BioRender.com.

of tumour versus normal tissue), in a weakly supervised method the label is given at the patient level (e.g. the whole-slide H&E image contains areas of tumour). When the aim is to identify distinct phenotypes present in the data without labels, unsupervised methods, such as dimensionality reduction and clustering analysis, are applied.

In conventional machine learning, an image processing step prior to feature engineering is commonly segmentation to identify cell positions and classify cell types. Neural network models have been increasingly employed to perform image segmentation [36–38]. Subsequently, a set of quantitative features are extracted from a digital image based on certain mathematical descriptions, which are usually referred to as ‘hand-crafted’ features (Figure 1). These are then applied to a conventional machine-learning system, which establishes relationships between these input features and an output label, such as tumour diagnosis. In contrast, in deep-learning AI approaches, the raw image is fed into the AI model, such as a CNN, which then progressively ‘learns’ which aspects of the image are most relevant for the outcome of relevance [1]. Both conventional and deep-learning models usually require images to be

broken down into a number of equally sized ‘tiles’ for computational tractability. In addition, each image needs to be labelled for the outcome of interest, i.e. cancer or no cancer, recurrence versus non-recurrence, and this label will then be applied to all tiles in the image. As it is either laborious (cancer or no cancer) or infeasible (recurrence status) for pathologists to provide pixel-level annotations for all images, weakly supervised learning that labels tiles according to slide-level annotation has been commonly applied (e.g. [11]). The subsequent sections will discuss conventional machine-learning and deep-learning approaches to study the TME in greater depth.

Conventional machine learning to map from hand-crafted TME features to clinical outcomes

Quantitative hand-crafted TME features are frequently employed to predict patient outcomes via the use of conventional machine learning. These features are commonly defined according to mathematical descriptions of the density, spatial distribution, and

higher-order structures of cancer cells and microenvironmental components. Conventional machine-learning approaches, such as logistic regression and random forests, and survival statistical models, such as Cox proportional hazards model, are applied to map from features to clinical labels such as patient survival outcomes.

Lymphocyte features

Evasion of immunosurveillance is a hallmark of many cancers and can arise from a deficiency of anti-cancer immune cells (e.g. CD8⁺ cytotoxic T lymphocytes) or the presence of immunosuppressive elements (e.g. regulatory T cells) [39]. Consequently, features of immune contexture that represent the density, composition, and spatial organisation of immune components are shown to be correlated with cancer prognosis [40–42].

Imaging-based quantitative analysis of immune contexture has led to the identification of immune features associated with patient outcomes. An international study involving 13 countries demonstrated that Immunoscore, which quantifies the density of CD8⁺ and CD3⁺ cells in tumour core versus invasive margin based on immunohistochemistry images, was a reproducible and robust prognostic factor in patients with stage I–III colon cancer [22]. Immunoscore was also shown to predict responses to chimeric antigen receptor T cell therapy in large B cell lymphoma [23]. A similar quantitative metric applied to more immune cell types based on immunohistochemistry images showed that immune topographies involving both CD8 and CD163 were prognostic [24].

Analysis of tumour-infiltrating lymphocytes (TIL) has increasingly been facilitated by the application of CNNs, a type of deep learning that will be discussed in greater detail later. One CNN approach predicts the over- or under-representation of TILs within individual small patches of a whole-slide H&E image. Using this approach in a pan-cancer analysis revealed spatial clustering features of TILs that associate with overall survival in some tumour types, including breast cancer and melanoma, and identified variable enrichment of structural patterns across tumour types [15]. Another approach seeks to detect individual cell nuclei and classify cell types within a whole-slide H&E image. One such method is Spatially Constrained Convolutional Neural Network (SCCNN) [38]. This was used to reveal that intra-tumour heterogeneity in immune landscapes has prognostic value. Facilitated by a machine-learning pipeline for single-cell identification and classification in H&E images, quantified geospatial features of lymphocytes in multiple tumour regions showed that lung adenocarcinomas with more than one region characterised by low lymphocyte infiltration had worse prognosis [17]. As discussed earlier, and enabled by multiplex imaging such as IMC and CODEX, higher-order spatial immune cell neighbourhoods and interactions have been found to associate with survival outcomes and therapy outcomes in various tumour types, including cutaneous T cell

lymphomas [43], melanoma [44,45], colorectal cancer [46], and brain tumours [47].

Concomitant with the discovery of clinically relevant features based on digital pathology, experimental evidence increasingly sheds light on how immune contexture emerges from complex and dynamic immune cell behaviours. Combining immunostaining and dynamic imaging of T cells showed that stromal ECM density and orientation impacted T cell migration and localisation [48].

Vascular features

Vascular systems play a fundamental role in the distribution of oxygen and nutrients to sustain tumour growth, as well as in the delivery of therapies. Disorganised tumour vasculature is a hallmark of cancer [39]. Microvessel density and fractal dimension, which measure the complexity of microvessel networks, extracted from CD34 immunohistochemistry images, were associated with survival outcomes in clear cell renal cell carcinomas [49]. More recently, quantitative morphometric features based on CD31 immunohistochemistry identified vascular features, including endothelial density and vascular arm numbers, associated with disease-free survival in patients with clear cell renal cell carcinomas [50]. In addition to these quantitative metrics, topological data analysis, a set of methods to distil low-dimensional features from high-dimensional data, is emerging as a powerful mathematical tool for characterising vascular patterns. Multiscale topological descriptors of vascular patterns based on intravital images of mouse colorectal cancer quantitatively captured the dynamic changes in network architecture following anti-cancer therapies [51].

Fibroblast and matrix features

Fibroblasts play a variety of important roles within the TME, including deposition and organisation of ECM and complex interactions with cancer cells and different types of immune cell [52]. Imaging-based quantitative analysis of the organisation of stromal fibroblasts and matrix has led to the identification of features associated with patient outcomes. By overlaying second harmonic generation (SHG) images of collagen with H&E images, Conklin *et al* [53] devised a metric called tumour-associated collagen signature-3 (TACS-3) to describe a pattern in which bundles of aligned collagen were oriented perpendicular to the tumour boundary in breast cancer. They found that a positive TACS-3 score was correlated with unfavourable survival outcomes using a Cox proportional hazards model. Beck *et al* [54] extracted a rich set of quantitative features of epithelial and stromal compartments in breast cancer and discovered that stroma morphological features and their contextual relationships with cancer cells had prognostic value for patient survival. Yuan *et al* [14] reported that a quantitative score characterising the degree of spatial clustering in contrast to the randomly scattered

distribution of stromal cells based on H&E-stained images was associated with poor outcomes for breast cancer patients.

Various tools have been developed to facilitate the extraction of quantitative matrix features. A FIJI ImageJ plugin called TWOMBLY was developed to extract a diverse repertoire of quantitative metrics from Picosirius red or SHG images, including the number of end points and branch points, high-density matrix, curvature, alignment, and fractional dimension [20]. These metrics have shown clinical relevance, e.g. ECM alignment in the metastatic potential of prostate tumours [55] and collagen density and curvature in bladder cancer progression [56]. The TWOMBLY tool [20] can also be applied to extract features of blood vessels. A fibre segmentation and extraction MATLAB tool was developed that enabled quantitative analysis of collagen architectural features in SHG images [21]. In addition to fibre-level analysis, quantitative features derived from gaps between fibres were shown to contain biologically relevant architectural information [57]. Recently, a machine-learning approach was developed to construct SHG-like representation of collagen directly based on H&E images, which enabled non-destructive extraction of quantitative matrix features such as fibre orientation and alignment [58]. Overall, in contrast to lymphocyte features, quantitative features of vascular or matrix patterns are less developed and a consensual quantitative metric that predicts clinical outcome has not yet emerged.

High-order TME features

An increasing body of work also sought to extract higher-order features from H&E-stained images that describe the composition and spatial organisation involving multiple TME components [14,59,60]. Enabled by an automated computational pipeline to segment and classify cell nuclei, a set of features was engineered that describe the nuclear morphologies of cancer cells, stromal cells, and lymphocytes, demonstrating that integration of these histological features with genomic features improved the prediction of survival of oestrogen receptor-negative breast cancer patients [14]. An ecological diversity index to quantitatively characterise the spatial variation in the local composition of cancer cells, stromal cells, and lymphocytes revealed that high microenvironmental heterogeneity was linked with worse disease-free survival in breast cancer patients [59]. Characterising patterns of spatial cell-cell networks using a graph-based approach found that a high extent of both stromal clustering and barrier appeared to suppress lymphocyte infiltration into tumours and was associated with poor survival outcomes in melanoma patients [60]. A set of cell-level and tissue-level quantitative features of cancer cells and four TME cell types identified from H&E images of multiple tumour types indicated that these human-interpretable features were able to predict clinically relevant molecular phenotypes, such as PD-1 expression [61].

With the recent advances of spatially resolved multiplex imaging techniques, such as VECTRA Polaris, MIBI-TOF, IMC, and Phenocycler (CODEX), an increasing body of research has identified clinically relevant higher-order quantitative features characteristic of spatial cell communities and organisation of TMEs [27,28,30,62]. Quantitative characterisation of co-occurrence, interactions, and spatial enrichment of immune cell populations in MIBI-TOF images of triple-negative breast cancer revealed that spatially mixed, in contrast to compartmentalised, tumour-immune organisations were associated with poor survival outcomes [62]. Using IMC images of breast cancer, quantitative characterisation of pairwise cell neighbourhoods and higher-order cell communities showed that spatial multicellular features had superior predictive power of overall survival in comparison to clinically defined subtypes [27]. In another study, cell population composition transitioned at tissue interfaces in IMC images of breast cancer and higher-order multicellular structures were associated with genomic features and predictive of clinical outcomes [28]. Using CODEX images of advanced-stage colorectal cancer, comprehensive analysis of the organisation, functional state, and communication patterns of cell neighbourhoods uncovered spatially resolved multicellular features associated with effective antitumour immunity and survival outcomes [30].

The future development of spatially resolved imaging, such as 3D methods [63] and spatial high-plex profiling of RNA and protein expression [64], combined with advanced data analysis [65,66], is set to further provide a refined depiction and understanding of the TME. As hand-crafted feature engineering and machine learning increasingly show utility in the discovery of TME patterns associated with clinical outcomes, there remains a largely unmet need – a promising opportunity for understanding better the mechanistic underpinning of such patterns in pre-clinical laboratory research.

Deep learning to predict clinical outcomes and beyond

Deep-learning approaches have emerged as a powerful tool in digital pathology applications, such as tumour diagnosis and the inference of genotypes, without the need to explicitly engineer hand-crafted features. These approaches commonly employ algorithms capable of discovering relevant contextual patterns inherent in the input data and often require only a minimum level of human intelligence as input. Deep-learning approaches, therefore, have the potential to achieve both clinically relevant predictions and discovery of important features, which can enable further biological and experimental hypothesis generation. The development of deep-learning biomedical applications requires a large amount of data with clinical annotations and has benefited from pan-cancer, multimodality datasets curated in landmark programmes such as The Cancer Genome Atlas

(TCGA). In the following sections, we discuss a variety of digital pathology applications using deep learning.

Convolutional neural networks

CNNs are the most widely applied deep-learning method in digital histopathology applications (Table 2). Digital pathology applications of CNNs, and deep learning in general, benefit from the success of transfer learning, namely reuse of neural network models extensively trained for image classification problems with abundant training data, without the need to retrain all layers. CNNs have been extensively used to predict patient outcomes in recent years and demonstrate model performance that is comparable or superior to human experts, as discussed earlier and below. Despite the ‘black-box’ nature, researchers have developed methods to visualise domains within images associated with model prediction and therefore inform the clinically relevant tumour and microenvironmental features.

Campanella *et al* [11] developed a deep-learning framework that performed well without pixel-level annotation of tumour areas in the diagnosis of prostate cancer, basal cell carcinoma, and breast cancer based on

whole-slide H&E images. Courtiol *et al* [67] developed a CNN model predicting patient outcomes based on whole-slide H&E images of mesothelioma and found that histological stromal features were associated with poor survival outcomes. Kather *et al* [68] trained CNN models to form internal representations of different tissue classes based on histological images of colorectal cancer and showed that a ‘deep stroma score’ related to the representation of the stromal compartment was associated with survival outcomes.

In addition to the classification of clinical outcomes, multiple studies have shown that CNNs are also capable of predicting genomic alternations in various cancer contexts. A CNN model was able to classify subtypes of non-small cell lung cancer (NSCLC) and predict the mutational status of six of 10 most frequently mutated genes from histological images [69], although as discussed later, not with the precision seen using next-generation sequencing. CNN models were also able to predict the status of microsatellite instability in colorectal cancer based on H&E images [70]. Two concomitantly published studies by Kather *et al* [71] and Fu *et al* [72] showed that CNN-based deep-learning algorithms could predict diverse molecular alterations, including gene mutations and transcriptional profiles, directly from histology in a pan-cancer context. Both studies also found an association of TME features with CNN prediction of genomic alterations. Kather *et al* [71] also reported that the enrichment of stroma was associated with CNN prediction of consensus molecular subtype 4 (CMS4) in colorectal cancers. Another CNN model accurately classified CMS based on H&E images of rectal cancer [81]. Image-based CMS, associated the predictions of molecularly determined classes with histological features of TME organisation, such as lymphocytic infiltrates and desmoplastic stromal reaction, and demonstrated value in investigating intratumoural transcriptional heterogeneity. Overall, it remains unclear in these studies to what extent TME features were associated with CNN predictions of genomic alterations. Although these methods innovatively demonstrated the feasibility of predicting genomic alternations based on digital pathology images, the current precision is inferior to detection using next-generation sequencing, as discussed later. Further improvements are required to enable clinical translation.

Recent studies also sought to apply CNN models to image data beyond H&E slides, concomitant with the advances of spatially resolved imaging. By combining matched H&E staining and spatial transcriptomics data for model training, He *et al* [73] developed a CNN-based algorithm to predict expression profiles of 250 genes in a proof-of-concept study based on a dataset of 23 patients with breast cancer. Using IMC images as training data, Sorin *et al* [29] developed a CNN-based framework that processes and integrates information of individual marker stains of IMC images – this is outlined in detail later but an important finding was that a combination of five markers could predict survival outcomes of patients with NSCLC.

Table 2. Types of deep-learning model.

Type of model	Brief description	Examples
CNNs	In its core part, a CNN applies a mathematical operation called a convolution to pixel intensities within an input image and is hierarchically structured with layers of operations to represent features at varying scales within the image. Variants of the basic CNN, including Inception-V3 and ResNet, are among the best models in digital pathology applications.	[11,29,67–73]
AMs	An important feature of AMs, compared to CNNs, is the explicit representation of a non-uniform contribution of information in different parts of the input data as a trainable property of the neural network. Therefore, AMs make biological interpretation convenient, e.g. by outputting the relative importance of subdomains within the input image(s) for predicting patient outcomes.	[74–76]
GNNs	GNNs work on graphs constructed based on pre-processed biological landmarks, such as different types of cell, in contrast to pixel intensities. Therefore, GNNs can explicitly build in the structure of multicellular communities and cell–cell communications.	[31,77]
MMs	MMs integrates, as streams of model input, multiple modalities of data, such as digital pathology images, genomic sequencing data and clinical annotations, such as tumour grade. MMs can enable the assessment of the relative contribution of input modalities to predictions.	[78–80]

Attention-based models

Attention-based models (AMs) have transformed other machine-learning fields, such as language translation, and, recently, have started to have footprints in digital histopathology applications to cancer research (Table 2). AMs achieved good performance in the subtyping of renal cell carcinomas and NSCLC and were able to highlight characteristic morphological features on the whole-slide images [74]. In a follow-up study, an AM-based framework was developed that could simultaneously predict whether a tumour is a primary or metastasis and the site of its origin across multiple tumour types [75]. More recently, an AM was developed that predicted prognosis and therapy response in colorectal cancer based on immunohistochemistry images of four immune cell markers and explained the relative importance of markers via the built-in attention module [76].

Graph neural networks

Graph neural networks (GNNs) have been widely applied in other biological research fields, such as protein structure predictions [82]. Recent methodology developments led to promising applications to digital histopathology images in cancer research (Table 2). In application to imaging data, GNNs are commonly applied to graphs constructed based on pre-processed biological landmarks, such as different types of cells, in contrast to pixel intensities.

Applying a GNN to a spatial graph connecting segmented cells from H&E images of tissue microarrays of prostate cancer showed that the algorithm was able to classify Gleason scores [77]. A GNN based on CODEX images of head and neck and colorectal cancers demonstrated that the model was able to predict survival outcomes and identify disease-relevant cellular communities, such as compartmentalisation of granulocytes [31].

Multimodal models

In addition to approaches based on digital histopathology images alone, multimodal models (MMs) that integrate multiple modalities of data as streams of model input are emerging to become an exciting avenue of deep-learning application. Mobadersany *et al* [78] developed a MM framework integrating both histopathology images and genomic markers and achieved superior prediction of overall survival of glioma patients. They further showed that the model linked higher risk with histological features, including microvascular proliferation and high tumour cell density. Esteva *et al* [79] developed a MM based on both histopathology images and clinical variables such as Gleason score and tumour stage and showed that the model improved prognostication of patients with prostate cancer in randomised clinical trials with long-term follow-up. In a recent pan-cancer study, Chen *et al* [80] developed a multimodal AM that integrates digital histopathology and molecular profile data to predict patient

outcomes across 14 cancer types. Their approach led to the identification of prognostic morphological and molecular features correlated with outcomes. They also found that their MM attributed attention to areas of tumour-associated stroma in high-risk cases of pancreatic adenocarcinoma, suggesting a role for TME features in the model prediction of survival outcomes.

Other deep-learning approaches

A variety of exciting deep-learning applications have focused on other biologically and clinically important research areas, including real-time AI to assist in intraoperative diagnosis [83–85], biologically inspired AI to improve interpretability [86], efficient search of archival histopathology images to facilitate decision-making [87,88] and federated learning to encourage cross-centre collaboration and protect data privacy [89].

Challenges for the implementation of digital image AI – clinical and methodological

What can digital pathology advanced analytics achieve in the clinic?

An important clinical question is: what can AI algorithms as biomarkers realistically achieve to improve patient care? Biomarkers are typically thought of as diagnostic, prognostic or predictive of response to any specific therapy. Predictive biomarkers are particularly useful to direct personalised therapy choices but to date have been challenging to develop robustly and are infrequently available.

Progress in the use of AI in diagnostics and prognostication

In terms of diagnostic AI algorithms it is clear that in specific tumours, e.g. prostate cancer, there has been considerable progress in diagnosis, some of which are now available for diagnostic clinical use [10,11]. Cancers of unknown primary are tumours where it is difficult to define the primary site of origin. Tumour Origin Assessment via Deep Learning (TOAD) is a recently published high-throughput interpretable deep learning-based solution that uses H&E whole-slide images to predict whether a tumour is primary or metastatic and to ascribe a differential diagnosis for the primary site of origin [75]. Transfer learning and weakly supervised multitask learning were combined to train a unified predictive model. In addition, attention-based learning located slide regions of particular diagnostic relevance, which were validated by pathologists. On an external test cohort of cancer of unknown primary, TOAD showed an accuracy of 79.9% in diagnosing a primary site of origin. For distinguishing between a metastasis and a primary tumour, the model had an impressive area under the curve (AUC) of 0.919 in the external test cohort. TOAD illustrates how AI

algorithms using inexpensive routinely available diagnostic tissue can help pathologists with complex diagnostic decisions and reduce diagnostic workup times.

In terms of prognostic biomarkers, once again AI algorithms based on digital AI images have already shown impressive results. The Gleason score in prostate cancer is a measure of glandular differentiation and a well-established prognostic factor that is subject to inter-pathologist variation in scoring [90]. In recent years, several AI algorithms have been described using H&E images that attribute the Gleason score with an accuracy equivalent to that of specialist pathologists [91,92]. A further important growth in AI approaches to prognosis has been in extracting clinically relevant spatial relationships between different TME features from highly multiplex immunofluorescence images. Sorin *et al* [29] recently used IMC with 35 markers on 1 mm tissue microarray cores and a pre-trained neural network model that combined routine clinical parameters plus spatial cell information, including measures of cellular TME communities, to predict recurrence following lung cancer surgery. Intriguingly, the deep-learning model substantially improved in predictive ability with the specific incorporation of spatial information, but not with just cell frequencies, and achieved 95.9% accuracy in predicting recurrence in stage I NSCLC. The size of tissue used in this study corresponds well to the small diagnostic biopsies available in the clinic. Furthermore, the authors were able to reduce the number of markers and still achieve a predictive accuracy of 90.8% using CD14, CD16, CD94, α SMA, and CD117, suggesting that more clinically practical lower plex methods involving important TME targets add clinical benefit.

Despite the promise of AI algorithms in diagnosis and prognostication, potential risks of widespread incorporation into clinical practice include the deskilling of pathologists, e.g. in the detection of less common tumour patterns. Furthermore, it is not yet clear how well digital pathology-based AI algorithms perform in the detection of atypical tumour variants, e.g. prostate sarcoma rather than adenocarcinoma. In specific cases, pathologists will decide to section deeper and perform additional tests in the face of diagnostic uncertainty; it is possible that AI algorithms are unable to recommend the optimal management of such ambiguous cases.

Limitations in the use of AI for therapy selection

AI algorithms based on digital images to predict benefit from a specific therapy have not yet achieved the same accuracy as the previously described diagnostic and prognostic AI biomarkers. This is partly because detection of a specific genetic aberration is often required from digital images. Here, AI models provide little added value if the key information can be obtained by sequencing, except potentially reducing cost in some contexts. Using CNNs based on modified inception v3 architecture and whole-slide images of H&E-stained NSCLC from the TCGA lung cohort, the authors trained a model to identify 10 commonly mutated genes

in adenocarcinoma of the lung, which they compared with next-generation sequencing data [69]. The resulting AUC values for the detection of mutations in serine/threonine protein kinase 11 (*STK11*), *EGFR*, FAT atypical cadherin 1 (*FAT1*), SET binding protein 1 (*SETBP1*), *KRAS*, and *TP53* were between 0.733 and 0.856. Testing these predictive models in an independent dataset showed an AUC of 0.687 (CI 0.554–0.811), with a higher AUC (0.75; CI 0.500–0.966) in samples validated by sequencing than in those tested by immunohistochemistry (AUC 0.659; CI 0.485–0.826). These AUC values are similar to those seen in prostate cancer, where a deep learning-based predictive model was able to identify SPOP mutations with an AUC of 0.71 [93]. The above results are important, not least because *KRAS* and *EGFR* have specific drug candidates in the clinic. At present the accuracy of these models is not sufficient to guide the choice of therapy, but emerging larger datasets will improve model training and potentially reach a precision that is acceptable for clinical use. As discussed later, the deep-learning AI methodology used in these studies means it is difficult to ascertain how much the TME informs the eventual decision regarding mutation status.

In other personalised therapy contexts, there is no gene mutation to guide treatment selection but better predictive biomarkers are still urgently needed. The selection of patients who might benefit from immunotherapy is a relevant example where TME constituents play an important role in driving the response or resistance to the immunotherapy. A second example is the prediction of which patients with localised prostate cancer need additional androgen deprivation therapy alongside radical radiotherapy. Encouragingly, an AI-based marker has recently shown promise in filling this gap. Using self-supervised learning, a Resnet-50 feature extraction model was trained on image patches from H&E images of prostate biopsies from 5,000 patients recruited to five phase III randomised trials of radiotherapy plus or minus androgen deprivation therapy [94]. When applied to the validation or test set of 1,594 patients, there was a significantly positive biomarker treatment test for interaction (p -interaction = 0.01). These results require further validation, and it is relevant that the size of the datasets were small compared with AI studies outside of oncology, e.g. those used in animal recognition. Nonetheless, the findings are exciting because the test is the first predictive marker generated in this context. In addition, H&E images are practically easier and less expensive to acquire than the gene expression signatures currently used as prognostic biomarkers in localised prostate cancer [95].

Linking digital pathology-based advanced analytics with clinical imaging modalities

A further largely unexplored research opportunity with the potential to enhance clinical decision-making is linking digital pathology-based advanced analytics with routinely used clinical imaging modalities, such as

ultrasound, CT, and MRI. AI approaches using CNNs to link radiomics data from CT and MRI with clinical outcomes, e.g. in the diagnosis of liver metastasis from colorectal tumours, have recently expanded [96,97]. Chen *et al* [98] demonstrated the utility of CNN in predicting the methylation status of the O⁶-methylguanine methyltransferase (MGMT) gene promoter in glioblastoma multiforme, a prognostic biomarker, based on fluid-attenuated inversion recovery (FLAIR) MRI images. Furthermore, the relationship between digital pathology-based TME features, such as collagen morphology and mammographic features, is well established [99]. Beyond this, there is considerable scope to link more complex digital imaging features with widely available clinical images. AI algorithms have the computational capacity to overcome the challenges in combining such diverse data types using MMs. In addition, clinical images often visualise an entire tumour longitudinally, which enables detailed mapping of temporal and spatial heterogeneity in the TME.

Challenges of validation and qualification

Robust statistical validation and qualification is essential for novel AI algorithms based on digital pathology to translate to the clinic. The Reporting recommendations for Tumour MARKer prognostic studies (REMARK) guidelines were developed following the observation that ‘...despite years of research and hundreds of reports on tumour biomarkers in oncology, the number of markers that have emerged as clinically useful is pitifully small...’ [100–102]. The REMARK recommendations include adjustment for multiple testing and use of adequate measures of discrimination and calibration of any novel biomarker, which is particularly relevant in predicting prognosis. Both AUC and the concordance index (c-index) are useful measures of discrimination and an important judge of clinical utility is the improvement in AUC or c-index seen with the addition of the novel biomarker to standard clinical factors versus such factors alone [103,104].

Previous reporting guidelines including Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) [105] and Consolidated Standards of Reporting Trials (CONSORT) [106] are not readily applicable to clinical trials based on AI systems. Recently, SPIRIT-AI [107–109] and CONSORT-AI [110–112] have been developed to provide standards for designing and conducting clinical trials based on AI systems [113]. These guidelines provide AI-specific items in addition to the core items described in SPIRIT and CONSORT. Alternative guidelines are currently being developed [114,115]. Independently, SPIRIT-Path is another guideline extending from SPIRIT to recommend items on reporting of cellular and molecular pathology content [116]. With the proposal of multiple guidelines for the implementation of AI-facilitated clinical trials, a challenge emerges in the choice of which guideline to apply in specific clinical settings. A further challenge is how to enable the ongoing training and thus improvement

in AI algorithms versus needing to ‘lock’ an AI algorithm at a defined point in time prior to approvals for clinical use.

The explainability of AI applications is an important consideration for clinical use and there is ongoing debate about whether inherently interpretable machine-learning models should be advocated more than black-box models assisted by *post hoc* explanation [117,118]. Some of the core discussion points in the debate include whether black-box models necessarily have superior predictive performance and whether *post hoc* explanation of black-box models, such as saliency maps in CNNs or attention maps in AMs, may be misleading or insufficient. Future AI applications will need to encourage attempts to develop biologically inspired interpretable machine-learning models [86], as well as critically assessing the biological/clinical relevance of *post hoc* explanations of black-box models.

A further challenge for the clinical implementation of digital pathology-based AI algorithms is the need for prospective evaluation [119]. For predictive biomarkers, the gold standard is a prospective randomised controlled trial where patients are randomised according to the current standard of care or the novel biomarker score [119]. This expensive and lengthy prospective qualification is rarely carried out and, in reality, gene expression profiling signatures, e.g. Oncotype DX in breast cancer, have often entered routine clinical use following repeated retrospective validation and before prospective qualification is complete [120,121]. Digital pathology-based AI algorithms have almost entirely been evaluated retrospectively and, to our knowledge, have yet to be reported in a prospective setting. Unless this is addressed, there is the danger of a growing misfit between AI methodology and the well-established rigour of clinical trials. Equally, traditional randomised controlled trial methodology is not always suitable for the evaluation of AI technology; addressing a potential misfit requires clinical trialists to identify novel approaches to trial design that can accelerate approval of clinically beneficial digital pathology-based AI algorithms.

Prostate cancer diagnosis in a prospective setting is being tested in an ongoing multidisciplinary study called ARTICULATE PRO by integrating pathologists’ decisions with the use of Paige Prostate. Arguably, one critical element of successful AI implementations in prospective, in contrast to retrospective, settings is the human–AI interaction, as experts need to conduct live appraisal of recommendations by AI algorithms and make decisions on clinical practice. Guidelines like DECIDE-AI [122–124] have been developed to provide standards for reporting on the evaluation of AI systems in live clinical settings of patient care.

Summary and conclusions

The combination of highly multiplex platforms to profile the TME and increasing powerful computational models means there is exciting potential for enhanced clinical decision-making using digital pathology-based advanced

analytical methods. In order to realise this potential, relevant studies need to incorporate meaningful measures of 'added value' to the current standard of care, incorporate prospective validation and meet the high standards of reliability and robustness set out in guidelines such as REMARK [100].

Engineering hand-crafted features combined with conventional machine-learning approaches has the advantage of delivering clear interpretability in terms of which key features correlate with clinical outcomes. However, these approaches are often less successful than deep-learning methods, partly because the set of hand-crafted features can under-represent the complexity in digital histopathology. Therefore, there is a trade-off between model performance and interpretability. Deep-learning models have achieved good performance in the prediction of patient outcomes and demonstrated promising results in predicting genotypes from digital pathology images. However, there remains a need for interpretation, led by pathologists, of biological patterns within the input image captured by the model.

Of note, in both types of model framework and digital pathology application, interpretability remains largely limited from the perspective of mechanistic principles underpinning the formation of clinically relevant patterns. Whereas model interpretation sheds light on what kinds of TME organisation inform clinical outcomes, understanding why, and how, such patterns underlie disease progression will require pre-clinical mechanistic laboratory research. Integration of experimental insights into the analysis of histopathology images during model development will be an exciting avenue for the future progress of AI applications.

Acknowledgements

XF acknowledges funding from Prostate Cancer Research and the European Research Council. AW acknowledges funding from AstraZeneca and imCORE, the RMH/ICR Cancer Research UK RadNet Centre, and Royal Marsden NHS Foundation Trust funding to the NIHR Biomedical Research Centre at The Royal Marsden and The Institute of Cancer Research. ES acknowledges funding from the Crick and the European Research Council.

Author contributions statement

XF and AW drafted the original manuscript, tables and figures, ES provided advice on the structure and critical review of the manuscript. All authors agreed the final version to submit for publication.

References

1. Shmatko A, Ghaffari Laleh N, Gerstung M, et al. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Cancer* 2022; **3**: 1026–1038.

2. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. *Am J Surg Pathol* 2016; **40**: 244–252.
3. Humphrey PA, Moch H, Cubilla AL, et al. The 2016 WHO classification of tumours of the urinary system and male genital organs-part B: prostate and bladder tumours. *Eur Urol* 2016; **70**: 106–119.
4. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. the value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991; **19**: 403–410.
5. Page DL, Ellis IO, Elston CW. Histologic grading of breast cancer. Let's do it. *Am J Clin Pathol* 1995; **103**: 123–124.
6. Rakha EA, El-Sayed ME, Lee AH, et al. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *J Clin Oncol* 2008; **26**: 3153–3158.
7. Gobbi H, Simpson JF, Borowsky A, et al. Metaplastic breast tumors with a dominant fibromatosis-like phenotype have a high risk of local recurrence. *Cancer* 1999; **85**: 2170–2182.
8. Ho BC, Tan HW, Lee VK, et al. Preoperative and intraoperative diagnosis of low-grade adenosquamous carcinoma of the breast: potential diagnostic pitfalls. *Histopathology* 2006; **49**: 603–611.
9. van der Kwast TH, van Leenders GJ, Berney DM, et al. ISUP consensus definition of cribriform pattern prostate cancer. *Am J Surg Pathol* 2021; **45**: 1118–1126.
10. Raciti P, Sue J, Retamero JA, et al. Clinical validation of artificial intelligence-augmented pathology diagnosis demonstrates significant gains in diagnostic accuracy in prostate cancer detection. *Arch Pathol Lab Med* 2022; DOI: 10.5858/arpa.2022-0066-OA.
11. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019; **25**: 1301–1309.
12. Echle A, Rindtorff NT, Brinker TJ, et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer* 2021; **124**: 686–696.
13. Hirata E, Sahai E. Tumor microenvironment and differential responses to therapy. *Cold Spring Harb Perspect Med* 2017; **7**: a026781.
14. Yuan Y, Failmezger H, Rueda OM, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Transl Med* 2012; **4**: 157ra143.
15. Saltz J, Gupta R, Hou L, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep* 2018; **23**: 181–193.e7.
16. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep* 2017; **7**: 16878.
17. AbdulJabbar K, Raza SEA, Rosenthal R, et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat Med* 2020; **26**: 1054–1062.
18. Bedin V, Adam RL, de Sá BC, et al. Fractal dimension of chromatin is an independent prognostic factor for survival in melanoma. *BMC Cancer* 2010; **10**: 260.
19. Metzke K. Fractal dimension of chromatin: potential molecular diagnostic applications for cancer prognosis. *Expert Rev Mol Diagn* 2013; **13**: 719–735.
20. Wershof E, Park D, Barry DJ, et al. A Fiji macro for quantifying pattern in extracellular matrix. *Life Sci Alliance* 2021; **4**: e202000880.
21. Bredfeldt JS, Liu Y, Pehlke CA, et al. Computational segmentation of collagen fibers from second-harmonic generation images of breast cancer. *J Biomed Opt* 2014; **19**: 16007.
22. Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus immunoscore for the classification of colon

- cancer: a prognostic and accuracy study. *Lancet* 2018; **391**: 2128–2139.
23. Scholler N, Perbost R, Locke FL, *et al.* Tumor immune contexture is a determinant of anti-CD19 CAR T cell efficacy in large B cell lymphoma. *Nat Med* 2022; **28**: 1872–1882.
 24. Kather JN, Suarez-Carmona M, Charoentong P, *et al.* Topography of cancer-associated immune cells in human solid tumors. *Elife* 2018; **7**: e36967.
 25. Taube JM, Roman K, Engle EL, *et al.* Multi-institutional TSA-amplified multiplexed immunofluorescence reproducibility evaluation (MITRE) study. *J Immunother Cancer* 2021; **9**: e002197.
 26. Viratham Pulsawatdi A, Craig SG, Bingham V, *et al.* A robust multiplex immunofluorescence and digital pathology workflow for the characterisation of the tumour immune microenvironment. *Mol Oncol* 2020; **14**: 2384–2402.
 27. Jackson HW, Fischer JR, Zanotelli VRT, *et al.* The single-cell pathology landscape of breast cancer. *Nature* 2020; **578**: 615–620.
 28. Danenberg E, Bardwell H, Zanotelli VRT, *et al.* Breast tumor microenvironment structures are associated with genomic features and clinical outcome. *Nat Genet* 2022; **54**: 660–669.
 29. Sorin M, Rezaejad M, Karimi E, *et al.* Single-cell spatial landscapes of the lung tumour immune microenvironment. *Nature* 2023; **614**: 548–554.
 30. Schürch CM, Bhate SS, Barlow GL, *et al.* Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* 2020; **182**: e1319.
 31. Wu Z, Trevino AE, Wu E, *et al.* Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. *Nat Biomed Eng* 2022; **6**: 1435–1448.
 32. Giesen C, Wang HA, Schapiro D, *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* 2014; **11**: 417–422.
 33. Keren L, Bosse M, Thompson S, *et al.* MIBI-TOF: a multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci Adv* 2019; **5**: eaax5851.
 34. Bera K, Schalper KA, Rimm DL, *et al.* Artificial intelligence in digital pathology – new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* 2019; **16**: 703–715.
 35. Lipkova J, Chen RJ, Chen B, *et al.* Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* 2022; **40**: 1095–1110.
 36. Graham S, Vu QD, Raza SEA, *et al.* Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* 2019; **58**: 101563.
 37. Mahmood F, Borders D, Chen RJ, *et al.* Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Trans Med Imaging* 2020; **39**: 3257–3267.
 38. Sirinukunwattana K, Ahmed Raza SE, Yee-Wah T, *et al.* Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imaging* 2016; **35**: 1196–1206.
 39. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; **144**: 646–674.
 40. Fridman WH, Zitvogel L, Sautès-Fridman C, *et al.* The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol* 2017; **14**: 717–734.
 41. Fridman WH, Pagès F, Sautès-Fridman C, *et al.* The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer* 2012; **12**: 298–306.
 42. Bruni D, Angell HK, Galon J. The immune contexture and immunoscore in cancer prognosis and therapeutic efficacy. *Nat Rev Cancer* 2020; **20**: 662–680.
 43. Phillips D, Matusiak M, Gutierrez BR, *et al.* Immune cell topography predicts response to PD-1 blockade in cutaneous T cell lymphoma. *Nat Commun* 2021; **12**: 6726.
 44. Moldoveanu D, Ramsay L, Lajoie M, *et al.* Spatially mapping the immune landscape of melanoma using imaging mass cytometry. *Sci Immunol* 2022; **7**: eabi5072.
 45. Hoch T, Schulz D, Eling N, *et al.* Multiplexed imaging mass cytometry of the chemokine milieu in melanoma characterizes features of the response to immunotherapy. *Sci Immunol* 2022; **7**: eabk1692.
 46. Lin JR, Wang S, Coy S, *et al.* Multiplexed 3D atlas of state transitions and immune interaction in colorectal cancer. *Cell* 2023; **186**: 363–381.e19.
 47. Karimi E, Yu MW, Maritan SM, *et al.* Single-cell spatial immune landscapes of primary and metastatic brain tumours. *Nature* 2023; **614**: 555–563.
 48. Salmon H, Franciszkiewicz K, Damotte D, *et al.* Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. *J Clin Invest* 2012; **122**: 899–910.
 49. Sabo E, Boltenko A, Sova Y, *et al.* Microscopic analysis and significance of vascular architectural complexity in renal cell carcinoma. *Clin Cancer Res* 2001; **7**: 533–537.
 50. Ing N, Huang F, Conley A, *et al.* A novel machine learning approach reveals latent vascular phenotypes predictive of renal cancer outcome. *Sci Rep* 2017; **7**: 13190.
 51. Stolz BJ, Kaeppler J, Markelc B, *et al.* Multiscale topology characterizes dynamic tumor vascular networks. *Sci Adv* 2022; **8**: eabm2456.
 52. Sahai E, Astsaturov I, Cukierman E, *et al.* A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer* 2020; **20**: 174–186.
 53. Conklin MW, Eickhoff JC, Riching KM, *et al.* Aligned collagen is a prognostic signature for survival in human breast carcinoma. *Am J Pathol* 2011; **178**: 1221–1232.
 54. Beck AH, Sangoi AR, Leung S, *et al.* Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med* 2011; **3**: 108ra113.
 55. Penet MF, Kakkad S, Pathak AP, *et al.* Structure and function of a prostate cancer dissemination-permissive extracellular matrix. *Clin Cancer Res* 2017; **23**: 2245–2254.
 56. Brooks M, Mo Q, Krasnow R, *et al.* Positive association of collagen type I with non-muscle invasive bladder cancer progression. *Oncotarget* 2016; **7**: 82609–82619.
 57. Acton SE, Farrugia AJ, Astarita JL, *et al.* Dendritic cells control fibroblastic reticular network tension and lymph node expansion. *Nature* 2014; **514**: 498–502.
 58. Keikhosravi A, Li B, Liu Y, *et al.* Non-disruptive collagen characterization in clinical histopathology using cross-modality image synthesis. *Commun Biol* 2020; **3**: 414.
 59. Natrajan R, Sailem H, Mardakheh FK, *et al.* Microenvironmental heterogeneity parallels breast cancer progression: a histology-genomic integration analysis. *PLoS Med* 2016; **13**: e1001961.
 60. Failmezger H, Muralidhar S, Rullan A, *et al.* Topological tumor graphs: a graph-based spatial model to infer stromal recruitment for immunosuppression in melanoma histology. *Cancer Res* 2020; **80**: 1199–1209.
 61. Diao JA, Wang JK, Chui WF, *et al.* Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nat Commun* 2021; **12**: 1613.
 62. Keren L, Bosse M, Marquez D, *et al.* A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell* 2018; **174**: 1373–1387.e19.
 63. Kuett L, Catena R, Özcan A, *et al.* Three-dimensional imaging mass cytometry for highly multiplexed molecular and cellular mapping of tissues and the tumor microenvironment. *Nat Cancer* 2022; **3**: 122–133.
 64. He S, Bhatt R, Brown C, *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat Biotechnol* 2022; **40**: 1794–1806.

65. Zhang W, Li I, Reticker-Flynn NE, et al. Identification of cell types in multiplexed in situ images by combining protein expression and spatial information using CELESTA. *Nat Methods* 2022; **19**: 759–769.
66. Kim J, Rustam S, Mosquera JM, et al. Unsupervised discovery of tissue architecture in multiplexed imaging. *Nat Methods* 2022; **19**: 1653–1661.
67. Courtiol P, Maussion C, Moarii M, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat Med* 2019; **25**: 1519–1525.
68. Kather JN, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med* 2019; **16**: e1002730.
69. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018; **24**: 1559–1567.
70. Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019; **25**: 1054–1056.
71. Kather JN, Heij LR, Grabsch HI, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Cancer* 2020; **1**: 789–799.
72. Fu Y, Jung AW, Torne RV, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer* 2020; **1**: 800–810.
73. He B, Bergensträhle L, Stenbeck L, et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* 2020; **4**: 827–834.
74. Lu MY, Williamson DFK, Chen TY, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021; **5**: 555–570.
75. Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. *Nature* 2021; **594**: 106–110.
76. Foersch S, Glasner C, Woerl AC, et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. *Nat Med* 2023; **29**: 430–439.
77. Wang JCR, Lu M, Baras A, et al. Weakly supervised prostate Tma classification via graph convolutional networks. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. Iowa City, 2020; 239–243.
78. Mobadersany P, Yousefi S, Amgad M, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A* 2018; **115**: E2970–E2979.
79. Esteva A, Feng J, van der Wal D, et al. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *NPJ Digit Med* 2022; **5**: 71.
80. Chen RJ, Lu MY, Williamson DFK, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* 2022; **40**: e866.
81. Sirinukunwattana K, Domingo E, Richman SD, et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* 2021; **70**: 544–554.
82. Li MM, Huang K, Zitnik M. Graph representation learning in biomedicine and healthcare. *Nat Biomed Eng* 2022; **6**: 1353–1369.
83. Chen PC, Gadepalli K, MacDonald R, et al. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nat Med* 2019; **25**: 1453–1457.
84. Ozyoruk KB, Can S, Darbaz B, et al. A deep-learning model for transforming the style of tissue images from cryosectioned to formalin-fixed and paraffin-embedded. *Nat Biomed Eng* 2022; **6**: 1407–1419.
85. Hollon TC, Pandian B, Adapa AR, et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med* 2020; **26**: 52–58.
86. Elmarakeby HA, Hwang J, Arafeh R, et al. Biologically informed deep neural network for prostate cancer discovery. *Nature* 2021; **598**: 348–352.
87. Kalra S, Tizhoosh HR, Shah S, et al. Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence. *NPJ Digit Med* 2020; **3**: 31.
88. Komura D, Kawabe A, Fukuta K, et al. Universal encoding of pan-cancer histology by deep texture representations. *Cell Rep* 2022; **38**: 110424.
89. Ogier du Terrail J, Leopold A, Joly C, et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. *Nat Med* 2023; **29**: 135–146.
90. Harbias A, Salmo E, Crump A. Implications of observer variation in Gleason scoring of prostate cancer on clinical management: a collaborative audit. *Gulf J Oncolog* 2017; **1**: 41–45.
91. Bulten W, Kartasalo K, Chen PC, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med* 2022; **28**: 154–163.
92. Ström P, Kartasalo K, Olsson H, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020; **21**: 222–232.
93. Schaumberg Andrew RM, Thomas F, Fuchs TJ. H&E-stained whole slide image deep learning predicts SPOP mutation state in prostate cancer. *bioRxiv* 2016. [Accessed 29 June 2023]. Available from: <http://biorxiv.org/content/early/2016/07/17/064279>. [Not peer reviewed].
94. Spratt Daniel SY, Van der Wal D, Huang S-C, et al. An AI-Derived digital pathology-based biomarker to predict the benefit of androgen deprivation therapy in localized prostate cancer with validation in NRG/RT0G 9408. *J Clin Oncol* 2022; **40**: 223.
95. Jairath NK, Dal Pra A, Vince R Jr, et al. A systematic review of the evidence for the decipher genomic classifier in prostate cancer. *Eur Urol* 2021; **79**: 374–383.
96. Lee S, Choe EK, Kim SY, et al. Liver imaging features by convolutional neural network to predict the metachronous liver metastasis in stage I-III colorectal cancer patients based on preoperative abdominal CT scan. *BMC Bioinformatics* 2020; **21**: 382.
97. Taghavi M, Trebeschi S, Simões R, et al. Machine learning-based analysis of CT radiomics model for prediction of colorectal metachronous liver metastases. *Abdom Radiol (NY)* 2021; **46**: 249–256.
98. Chen X, Zeng M, Tong Y, et al. Automatic prediction of MGMT status in glioblastoma via deep learning-based MR image analysis. *Biomed Res Int* 2020; **2020**: 9258649.
99. McConnell JC, O'Connell OV, Brennan K, et al. Increased periductal collagen micro-organization may contribute to raised mammographic density. *Breast Cancer Res* 2016; **18**: 5.
100. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Eur J Cancer* 2005; **41**: 1690–1696.
101. Altman DG, McShane LM, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med* 2012; **10**: 51.
102. Sauerbrei W, Taube SE, McShane LM, et al. Reporting recommendations for tumor marker prognostic studies (REMARK): an abridged explanation and elaboration. *J Natl Cancer Inst* 2018; **110**: 803–811.
103. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; **15**: 361–387.

104. Royston P, Altman DG. External validation of a cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013; **13**: 33.
105. Chan AW, Tetzlaff JM, Gøtzsche PC, *et al.* SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 2013; **346**: e7586.
106. Moher D, Hopewell S, Schulz KF, *et al.* CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; **340**: c869.
107. Cruz Rivera S, Liu X, Chan AW, *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digit Health* 2020; **2**: e549–e560.
108. Cruz Rivera S, Liu X, Chan AW, *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* 2020; **26**: 1351–1363.
109. Rivera SC, Liu X, Chan AW, *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *BMJ* 2020; **370**: m3210.
110. Liu X, Cruz Rivera S, Moher D, *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health* 2020; **2**: e537–e548.
111. Liu X, Cruz Rivera S, Moher D, *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020; **26**: 1364–1374.
112. Liu X, Rivera SC, Moher D, *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ* 2020; **370**: m3164.
113. Ibrahim H, Liu X, Rivera SC, *et al.* Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. *Trials* 2021; **22**: 11.
114. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019; **393**: 1577–1579.
115. Sounderajah V, Ashrafian H, Aggarwal R, *et al.* Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: the STARD-AI steering group. *Nat Med* 2020; **26**: 807–808.
116. Kendall TJ, Robinson M, Brierley DJ, *et al.* Guidelines for cellular and molecular pathology content in clinical trial protocols: the SPIRIT-path extension. *Lancet Oncol* 2021; **22**: e435–e445.
117. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019; **1**: 206–215.
118. Theodore Evans COR, Geißler C, Kargl M, *et al.* The explainability paradox: challenges for xAI in digital pathology. *Future Generation Computer Systems* 2022; **133**: 281–296.
119. Mandrekar SJ, Grothey A, Goetz MP, *et al.* Clinical trial designs for prospective validation of biomarkers. *Am J Pharmacogenomics* 2005; **5**: 317–325.
120. Sparano JA, Gray RJ, Makower DF, *et al.* Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med* 2018; **379**: 111–121.
121. Dowsett M, Sestak I, Lopez-Knowles E, *et al.* Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol* 2013; **31**: 2783–2790.
122. Vasey B, Novak A, Ather S, *et al.* DECIDE-AI: a new reporting guideline and its relevance to artificial intelligence studies in radiology. *Clin Radiol* 2023; **78**: 130–136.
123. Vasey B, Nagendran M, Campbell B, *et al.* Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med* 2022; **28**: 924–933.
124. Vasey B, Nagendran M, Campbell B, *et al.* Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* 2022; **377**: e070904.