Medical Image Analysis 93 (2024) 103071

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Social network analysis of cell networks improves deep learning for prediction of molecular pathways and key mutations in colorectal cancer

Neda Zamanitajeddin [a], Mostafa Jahanifar [a], Mohsin Bilal [a], Mark Eastwood [a], Nasir Rajpoot [a,b,*]

[a] Tissue Image Analytics Centre, Department of Computer Science, University of Warwick, Coventry, UK
[b] Histofy Ltd., Birmingham, UK

ARTICLE INFO

Keywords:
Social network analysis
Deep learning
IDaRS
CLAM
Colorectal cancer
Key mutation prediction

ABSTRACT

Colorectal cancer (CRC) is a primary global health concern, and identifying the molecular pathways, genetic subtypes, and mutations associated with CRC is crucial for precision medicine. However, traditional measurement techniques such as gene sequencing are costly and time-consuming, while most deep learning methods proposed for this task lack interpretability. This study offers a new approach to enhance the state-of-the-art deep learning methods for molecular pathways and key mutation prediction by incorporating cell network information. We build cell graphs with nuclei as nodes and nuclei connections as edges of the network and leverage Social Network Analysis (SNA) measures to extract abstract, perceivable, and interpretable features that explicitly describe the cell network characteristics in an image. Our approach does not rely on precise nuclei segmentation or feature extraction, is computationally efficient, and is easily scalable. In this study, we utilize the TCGA-CRC-DX dataset, comprising 499 patients and 502 diagnostic slides from primary colorectal tumours, sourced from 36 distinct medical centres in the United States. By incorporating the SNA features alongside deep features in two multiple instance learning frameworks, we demonstrate improved performance for chromosomal instability (CIN), hypermutated tumour (HM), *TP53* gene, *BRAF* gene, and Microsatellite instability (MSI) status prediction tasks (2.4%–4% and 7–8.8% improvement in AUROC and AUPRC on average). Additionally, our method achieves outstanding performance on MSI prediction in an external PAIP dataset (99% AUROC and 98% AUPRC), demonstrating its generalizability. Our findings highlight the discrimination power of SNA features and how they can be beneficial to deep learning models' performance and provide insights into the correlation of cell network profiles with molecular pathways and key mutations.

1. Introduction

Realizing trustworthy computer-assisted diagnostic systems for histopathology applications is now feasible because of developments in digital pathology and the emergence of powerful computational techniques like Deep Learning (DL). In particular, Convolutional Neural Networks (CNNs) (LeCun et al., 2015) have shown promising results in solving fundamental processing tasks in Computational Pathology (CPath) such as nuclei instance segmentation (Graham et al., 2019; Alemi Koohbanani et al., 2019, 2020), tissue segmentation (Graham et al., 2022b; Shephard et al., 2021; Jahanifar et al., 2021), and tissue classification (Vuong et al., 2022). Several studies have reviewed numerous tasks in which DL methods can help to improve task performance, pathologist efficiency, and patient care (Srinidhi et al., 2021; Hong and Fenyö, 2022; Madabhushi and Lee, 2016; Janowczyk and Madabhushi, 2016).

With advances in Whole-Slide Image (WSI) scanning and increasing adoption of digital pathology, it is now possible to design machine learning (ML) algorithms that can capture information from the whole tissue slide and predict complex entities such as cancer diagnosis labels (Campanella et al., 2019; Lu et al., 2021b), tumour grading (Oliveira et al., 2021), and patient prognosis (Kather et al., 2019a; Chen et al., 2022). Usually, these methods are trained on large-scale datasets of WSIs for which only patient-level or slide-level labels are available, i.e. weakly-labelled datasets. To leverage these kinds of datasets, a specific type of weakly-supervised learning approach, called Multiple Instance Learning (MIL), has been widely utilized (Carbonneau et al., 2018). Usually, DL-based MIL approaches for CPath assume a WSI as a labelled "bag" of image tile "instances". MIL methods have become popular in CPath because they only require bag-level (WSI-level) labels to train a model and there is no need to exhaustively annotate numerous instances (tiles) inside each bag (Carbonneau et al., 2018;

* Corresponding author at: Tissue Image Analytics Centre, Department of Computer Science, University of Warwick, Coventry, UK.
E-mail addresses: neda.zamanitajeddin@warwick.ac.uk (N. Zamanitajeddin), n.m.rajpoot@warwick.ac.uk (N. Rajpoot).

https://doi.org/10.1016/j.media.2023.103071
Received 14 March 2023; Received in revised form 14 November 2023; Accepted 19 December 2023
Available online 5 January 2024

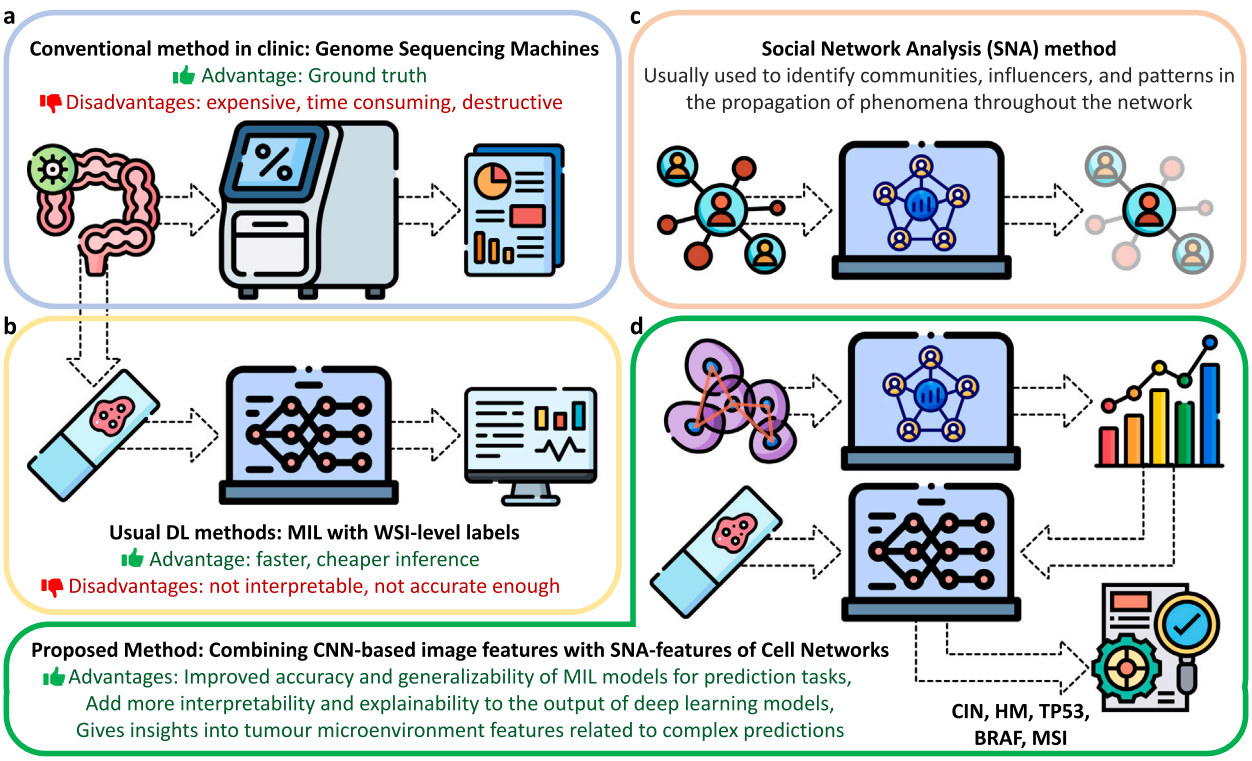

Fig. 1. Usual method in the clinic for molecular pathways and key mutation prediction is through gene sequencing and immunohistochemistry staining (a). While DL offers cheap and fast alternatives, they are not good enough (b). Enhanced with SNA features (c), the proposed MIL approach can improve the prediction accuracy (d).

Campanella et al., 2019). In comparison to instance-level labels, WSI-level labels are relatively easier and faster to extract from pathology reports, patient follow-ups, or associated genomics data. Furthermore, a property of MIL methods is that they allow for a variable number of instances in each bag which corresponds well with the properties of the WSIs.

Colorectal cancer (CRC) is the third most common malignancy and the second most lethal cancer worldwide (Xi and Xu, 2021). CRC carcinogenesis, arising through multiple genetic events, provides the basis for precision medicine by explaining the genetic and epigenetic diversity in disease progression and tumour growth among patients. Determining CRC molecular pathways, genetic subtypes, and mutations along with TNM staging are key to the shift in paradigm from "one drug fits all" to precision medicine (Al-Sohaily et al., 2012), though limited to only a subset of patients (Bilal et al., 2022a). Microsatellite instability (MSI) status of a CRC tumour serves as an approved biomarker for the first-line immunotherapy treatment (FDA et al., 2020). Other molecular pathways including chromosomal instability (CIN), CPG island methylator phenotype (CIMP) and genetic mutations (such as *TP53*, *BRAF*, *NRAS*, and *KRAS*) further help to select adjuvant or anti-EGFR therapies (Singh et al., 2021). Immunohistochemistry (IHC) (Boland and Goel, 2010) or Polymerase Chain Reaction (PCR) (Pritchard and Grady, 2011), In-Situ Hybridization (ISH) or Next Generation Sequencing (NGS) of the human genome, are required to know patients molecular and genetic statuses (Luchini et al., 2019; El-Deiry et al., 2019) (Fig. 1a).

These workflows require extra processing time, human expertise, and expenses. Moreover, some tests for measuring these attributes can be destructive to the testing sample. An interesting use case of MIL and weakly-supervised learning can be predicting such entities directly from Haematoxylin and Eosin (H&E) stained slides (Fig. 1b). To this end, DL-based methods have shown great potential to predict different molecular pathways or key mutations with considerable accuracy using only H&E slides and WSI-level labels (Kather et al., 2019b, 2020; Echle et al., 2020; Bilal et al., 2021; Laleh et al., 2021; Couture, 2022)

Although predicting these important entities through DL methods is gaining increasing attention (Couture, 2022) and MIL-based methods have shown promising results on many of these complex prediction tasks (Bilal et al., 2021), most existing methods are bound by intrinsic limitations of CNNs relying on only pixel-level information. In other words, most of the tested DL-based or MIL-based methods in CPath use only feature representations extracted by a pretrained CNN model from input images (Laleh et al., 2021). Despite studies showing that using auxiliary information from different modalities can help improve models' performance in difficult prediction tasks such as survival prediction (Chen et al., 2020; Mobadersany et al., 2018), no method has ever tried incorporating auxiliary information for molecular alteration or key mutation prediction tasks. This is perhaps because of the fact that widely used auxiliary information comes from genomics data (Boehm et al., 2022) whereas, for the aforementioned tasks, we are trying to avoid using expensive IHC staining or gene sequencing and instead predict these important entities directly from H&E images. Furthermore, access to auxiliary data from other imaging modalities, such as MRI, is not always available.

Another important drawback of DL-based methods is the lack of interpretability. In recent studies, different DL-based methods try to increase the explainability of their approach by visualizing the attention score (Lu et al., 2021b) or the prediction score (Bilal et al., 2021) of different tiles of the WSI as a heatmap overlay. However, even with heatmap overlays, it is still unclear how tissue morphology or pathological primitive structures in a histology image relate to deep CNN features. Often, it is required that pathologists review the most attentive regions (Lu et al., 2021b) or study the cellular compositions in those regions (Bilal et al., 2021) to gain some insights into the prediction.

In this paper, we propose to address these challenges by incorporating cell-to-cell relationship information as auxiliary inputs in weakly-supervised DL frameworks to boost their performance for the prediction of molecular pathways and key mutations in colorectal cancer. This is done by introducing cell graphs with nuclei as nodes and nuclei

connections as edges of the network. These networks help us to identify statistics of cellular communities or influential cells, which may be impossible to locate by manual inspection but may contain diagnostically significant information about the tissue microenvironment. Because cell nuclei are vital histology primitives and are always present in H&E stained images, this kind of auxiliary information is always available at hand without incurring extra costs.

Our proposed approach leverages Social Network Analysis (SNA) measures (Newman, 2018) to capture cell connections in the tumour microenvironment. SNA is a field of research that has been widely used to investigate social relationships among entities (Fig. 1c), such as social media analysis (Hagen et al., 2018) and disease transmission (Krause et al., 2007). By considering cells in the tumour microenvironment as individuals and their connections as social relationships, we can extract SNA-based features from cell graphs, which provide abstract yet interpretable information about the spatial distribution and relationships of cells (Zamanitajeddin et al., 2021). These features can be used as auxiliary information alongside CNN-based image representations to improve prediction performance for key mutations and molecular pathways.

The proposed framework offers several advantages, including computational efficiency and scalability. First, the extraction of SNA features from cell networks, which does not rely on graph neural networks and GPU computation, eliminates the need for accurate nuclei segmentation or morphological feature extraction. Second, it explicitly captures cell relationships within the image, which CNN-based methods do not explicitly model. Third, it allows for the identification of tissue phenotypes and the highlighting of biologically important regions in the image. By incorporating SNA-based features in recent deep learning frameworks (Fig. 1d), we show that our approach outperforms other methods and achieves the state-of-the-art (SOTA) performance for CIN, HM, *TP53*, *BRAF*, and MSI prediction tasks. To the best of our knowledge, this is the first study to extract auxiliary information about cell relationships from the histology image itself to enhance DL models' performance for high-level prediction tasks. In summary, the main contributions of this work are as below:

- We propose a novel weakly supervised DL framework that integrates deep image features and cell relationship information to enhance performance.
- We utilize efficient social network analysis to capture cell relationship information.
- The proposed framework achieves SOTA performance for predicting CIN, HM, *TP53*, *BRAF*, and MSI for colorectal cancer.
- Unveil insights into the association between cell distributions and relationships with molecular pathways and key mutation status.

2. Related works

In recent years, deep learning has been applied to the field of CPath to extract valuable information from histopathological images. Various studies have demonstrated the ability to predict molecular biomarkers, genomic subtypes, and protein biomarkers from WSIs using DL algorithms (Couture, 2022; Ghaffari Laleh et al., 2022; Srinidhi et al., 2021; Ahmedt-Aristizabal et al., 2021). These algorithms are often weakly supervised, meaning that they only require slide-level labels rather than pixel-level or tile-level annotations (Ghaffari Laleh et al., 2022; Carbonneau et al., 2018). These methods have been applied to a range of cancer types and have shown promising results in improving the accuracy of classifying molecular pathways and key mutations. However, challenges such as small training sets, rigorous validation, and model interpretability still persist.

2.1. MSI prediction in CRC

Two of the first studies that investigated the use of DL methods for molecular alteration detection directly from H&E images are (Kather et al., 2019b; Echle et al., 2020). Echle et al. (2020) proposed a DL method for the detection of MSI and mismatch-repair deficiency (dMMR) in colorectal tumours. They collected data from 8836 colorectal tumours from various locations and trained a DL classifier using this data. Kather et al. (2019b) performed a similar study where they claimed that using DL-based methods, MSI testing would be more widely available, as DL method only requires H&E histology which is commonly available in clinical practice, rather than additional genetic or immunohistochemical tests.

Cao et al. (2020) propose an ensemble MIL model to predict MSI status in CRC based on histopathology images. The method uses a residual convolutional neural network (ResNet) (He et al., 2016) to compute patch likelihoods in a MIL paradigm and then aggregates the patch likelihoods using two independent MIL methods, Patch Likelihood Histogram (PALHI) and Bag of Words (BoW), to obtain the WSI-level prediction. The model presented in Cao et al. (2020) has been tested on TCGA and Asian-CRC datasets where it was able to capture the relationship between pathological phenotype and MSI and identify pathological imaging signatures that are associated with mutation burden and DNA damage repair-related genotype in the genomic profiles, and anti-tumour immunity activated pathway in the transcriptomic profiles (Cao et al., 2020). Yamashita et al. (2021) proposed a DL-based system called MSINet to predict MSI in CRC directly from H&E-stained WSIs. The study not only reported high performance on internal and external datasets but also compared the model's performance with that of five pathologists, finding that the model's performance was significantly higher than the pathologists (Yamashita et al., 2021).

Lv et al. (2022) proposed a joint region-attention and multi-scale transformer (RAMST) network for MSI detection in colorectal cancer. Their proposed method used a region-attention mechanism and a feature weight uniform sampling (FWUS) method to learn a representative subset of image patches from WSIs and introduced a transformer architecture to fuse multi-scale histopathology features for slide-level MSI detection. Prediction of MSI from H&E images has attracted attention and a comprehensive review of the recent methods can be found in Echle et al. (2021) where authors systematically summarized and compared the existing methods on this topic. Echle et al. (2021) noted that DL-based methods have the potential to facilitate the identification of patients eligible for treatment with immune checkpoint inhibitors by pre-screening or replacement of current methods.

2.2. Prediction of other genomic markers

One of the pioneering works in gene mutation prediction from H&E images was from Coudray et al. (2018) where they trained CNN network to predict the ten most commonly mutated genes in lung cancer, with six of them (*STK11, EGFR, FAT1, SETBP1*, *KRAS* and *TP53*) showing promising results on TCGA-LUAD dataset with AUCs ranging from 0.733 to 0.856. Another influential work is by Kather et al. (2020) where they applied a DL method to more than 5000 patients with various types of cancer and found that in multiple major cancer types, the genotype of point mutations was predictable directly from H&E images. Kather et al. (2020) were able to successfully identify various actionable genetic alterations such as *TP53* and *EGFR* mutational status in lung adenocarcinoma, *BRAF* and *KRAS* mutations in colon and rectal cancer, *TP53*, *MAP2K4*, and *PIK3CA* mutations in breast cancer, and *CDK12* and *PIK3CA* mutations in prostate cancer.

Building on the previous successful studies, Teichmann et al. (2022) presented an end-to-end learning pipeline for WSI classification that predicts MSI in colorectal tumours and specific mutations in colon, lung, and breast cancer cases without requiring any auxiliary annotations. A feature recalibration-based MIL method was proposed

by Chikontwe et al. (2022), which recalibrates the distribution of a WSI bag (instances) by using the statistics of the max-instance (critical) feature and employs a position encoding module (PEM) to model spatial/morphological information. DeepSMILE is a deep learning method proposed by Schirris et al. (2022) for analysing WSIs of H&E-stained tumour tissue using self-supervised pre-training and heterogeneity-aware deep MIL, which improves the performance of genomic label classification (MSI in CRC and HRD score in breast cancer) with fewer labelled data, reaching similar performance as the baseline with only 40% of the labelled data. Similarly, Li et al. (2022) proposed a novel contrastive representation learning framework called Lesion-Aware Contrastive Learning (LACL) for histopathology WSI analysis. The authors evaluated their proposed method on the EGFR dataset, comprising lung adenocarcinoma WSIs, for epidermal growth factor receptor (EGFR) gene mutation identification. The results show that LACL outperforms other methods in histopathology image representation learning on EGFR mutation detection (Li et al., 2022). Furthermore, Zhao et al. (2022) proposed SETMIL, a novel Spatial Encoding Transformer-based MIL method, that aims to learn an informative and effective representation of each WSI to predict EGDR gene mutation prediction in lung cancer.

Utilizing graphs and graph neural networks (GNNs) for solving complex problems in CPath is becoming popular as well (Jaume et al., 2021; Ahmedt-Aristizabal et al., 2021). Authors in Reisenbüchler et al. (2022), proposed a new method called Local Attention Graph-Based Transformer for Multi-target Genetic Alteration Prediction (LA-MIL) that utilizes local self-attention to model dependencies in large-scale WSIs for predicting multiple biomarkers in gastrointestinal cancer. Reisenbüchler et al. (2022) compared their method to existing models and showed competitive results for predicting gene mutation (such as MSI, TMB, *BRAF*, and *TP53*) in CRC. The authors also show that their approach can leverage information from multiple targets to achieve better overall performance (Reisenbüchler et al., 2022). Lu et al. (2022) introduced a novel graph neural network-based model (SlideGraph) to predict HER2 status in breast cancer directly from H&E WSIs.

Hu et al. (2022) proposed a self-interactive MIL method for predicting molecular traits from tissue morphology in WSIs. Authors argued that previous methods, such as CLAM (Lu et al., 2021b), which rely on tile-level multi-instance learning and a fixed pretrained model for feature extraction, are not optimal for capturing both fine-grained features at the tile level and global features at the slide level. The proposed approach iteratively feedbacks training information between fine-grained and global context features. The method is evaluated on four subtyping tasks: EMT status in ovarian cancer, *KRAS* mutation in colon and lung cancer, EGFR mutation in colon cancer, and HER2 status in breast cancer and the results show an average improvement of 7.05–8.34% in terms of AUC over the baseline (Hu et al., 2022). Apart from H&E images, Jiménez-Sánchez et al. (2022) introduced NaroNet, a deep learning framework that models the multi-scale tumour microenvironment from multiplex-stained cancer tissue images and provides patient-level interpretable predictions using a seamless end-to-end learning pipeline. They utilized their method to predict POLE mutation status in Endometrial carcinomas.

2.3. SOTA weakly-supervised DL

Lu et al. (2021b) introduced a weakly supervised DL method called clustering-constrained-attention multiple-instance learning (CLAM) which is presented for data-efficient WSI processing. The method uses attention-based learning to identify subregions of high diagnostic value to classify whole slides and instance-level clustering to constrain and refine the feature space. It is applied to the subtyping of renal cell carcinoma and non-small-cell lung cancer, and detection of lymph node metastasis (Lu et al., 2021b). Bilal et al. (2021) proposed a novel MIL training strategy called iterative draw-and-rank sampling (IDaRS) that was used to predict the status of molecular pathways and key mutations in CRC. The original IDaRS was able to predict the status of the three main CRC molecular pathways (MSI, CIN, CIMP) and detect *BRAF* and *TP53* mutations as well as predict hypermutated (HM) CRC tumours with high accuracy. IDaRS was able to achieve SOTA AUROC values and also revealed novel histological features associated with each molecular pathway and key mutation through data-driven analysis of the cellular composition of image tiles. Among the recently proposed DL methods for WSI processing, IDaRS and CLAM have shown SOTA results for various tasks (Bilal et al., 2022b,a; Graham et al., 2023; Lu et al., 2021a; Chen et al., 2022). Ghaffari Laleh et al. (2022) aimed to systematically compare the performance of such DL methods for WSI processing in CPath. Six clinically relevant prediction tasks were selected, including the classification of morphological subtypes in renal cell carcinoma, prediction of microsatellite instability in colorectal and gastric cancer, and prediction of mutations in the *BRAF* and FGFR3 genes in colorectal and bladder cancer respectively. Six commonly used methods were implemented and compared, including both classical weakly-supervised approaches and MIL-based approaches. The results of the study showed that in mutation prediction tasks, classical weakly-supervised methods outperformed MIL-based methods, and these findings motivate the development of new methods which combine the assumptions of MIL with the higher performance of classical weakly-supervised approaches (Ghaffari Laleh et al., 2022). Couture (2022) recently reviewed the current advancements in using DL for predicting molecular biomarkers from H&E images in various cancer types, highlighting the leading trends, challenges, and potential applications in biomarker screening and patient outcomes prediction.

3. Methodology

3.1. Overview

In Fig. 2, we present an overview of the proposed method where we use social network analysis techniques to improve DL-based MIL methods in the prediction of molecular pathways and key mutations in CRC. The proposed method requires four main steps to operate, each of which will be explained in the following sections. First, we explain how tumour tiles are extracted from WSIs to form a tile dataset. Then, we describe the construction of the cell social networks from tumour tiles, which fill the cell network dataset. After that, we introduce social network analysis (SNA) measures and their statistical features as descriptors of cell relationships in tumour tiles. Finally, we explain how we incorporate social cell relationship information in conjunction with visual (deep) features into DL-based MIL pipelines (such as CLAM and IDaRS) to improve predictions.

3.2. Tumour tile extraction and preprocessing

The first step in our method is to extract tumour tiles from WSIs in the dataset. This is because only tumour regions carry information and potential signals related to key mutations and molecular pathways. This approach has been utilized previously in DL-based approaches (Bilal et al., 2021) and replicates the DNA sequencing or bio-marker profiling experiments that are done in the lab, where only tumour regions and cells are studied, to obtain ground truth labels.

To this end, the tissue region in the slides is segmented through Otsu thresholding. Tiles are then extracted from the segmented tissue regions only, following the protocol outlined in Kather et al. (2019b, 2020) and Bilal et al. (2021). More precisely, each WSI is divided into square tiles of 256 μm edge length at 20× magnification (0.5 μm/pixel resolution), which results in tile images with a size of 512 × 512 pixels. In our experiments, slides with less than ten tiles were excluded. The tile images are then resized to 1.14 μm/pixel resolution, resulting in square tiles of size 224 × 224 pixels, which is a standard size for ResNet18 (He et al., 2016). To ensure the best results for downstream

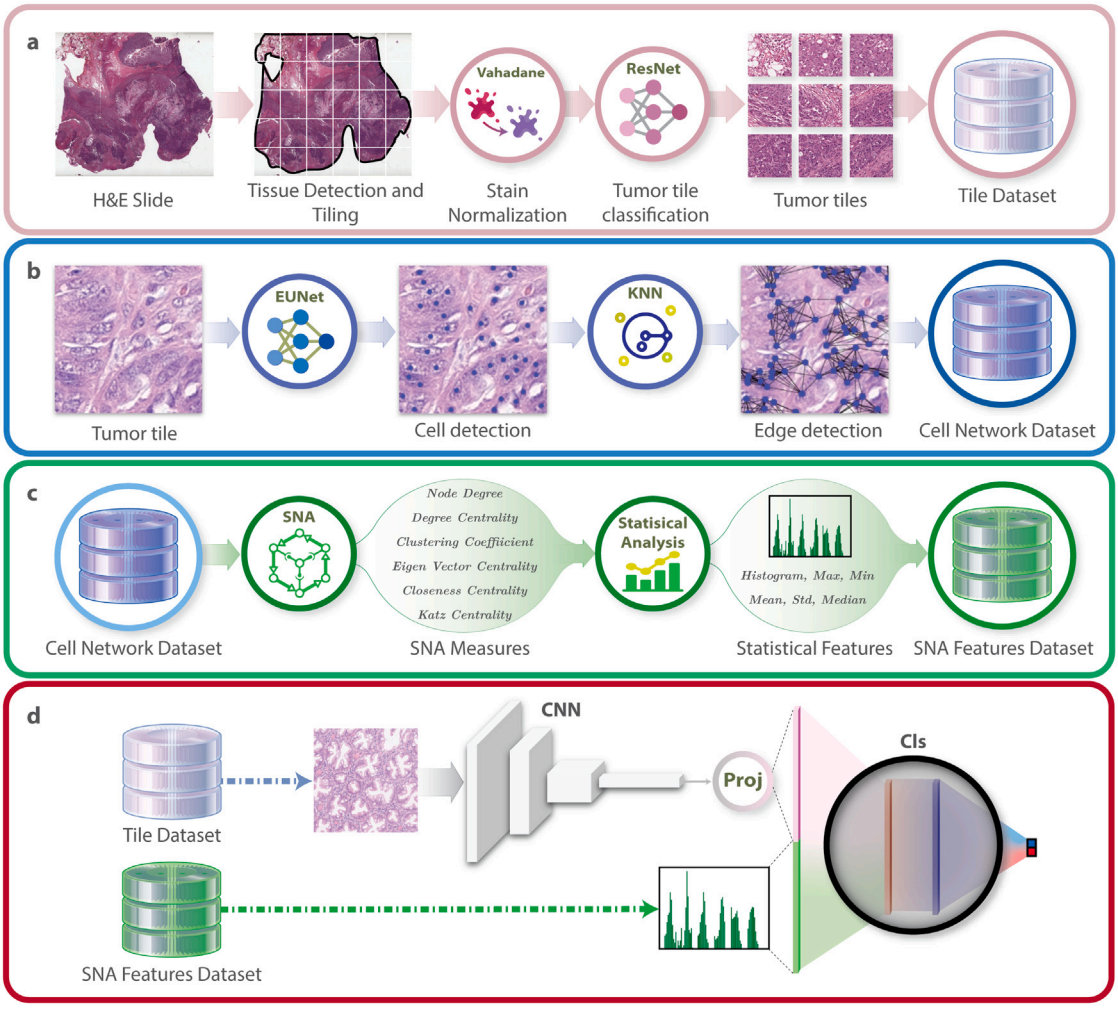

Fig. 2. Overview of the proposed method for improving deep learning methods for prediction of key mutations using social network analysis: (a) tumour tiles database creation, (b) cell network generation, (c) SNA-based statistical feature extraction from cell networks, and (d) training DL model using both deep and SNA features (Proj and Cls stand for Projection layer and Classifier MLP, respectively, see Section 3.6).

analysis through deep learning models and avoid domain-shift problems, the stain information of all tiles is normalized using (Vahadane et al., 2016) structure preserving stain normalization algorithm. These steps are depicted in Fig. 2a.

Normalized tiles are classified in tumour and non-tumour groups using the same tumour classifier model which is introduced in Bilal et al. (2021), a ResNet34 model and trained on 35,436 tiles taken from seven TCGA slides and datasets taken from Kather et al. (2019b) and Shaban et al. (2020). The resulting tumour tiles form 'Tile Dataset', $\mathbf{S_T}$, which comprises of tile sets, $\mathbf{T}_I \in \mathbf{S_T}$, of the images ($I$) in the dataset, where tile set of each image contains only tumour tiles $\mathbf{t}$ of that image, i.e. $\mathbf{S_T} = \{\mathbf{T}_I | \mathbf{T}_I = \{\mathbf{t}_1, \dots, \mathbf{t}_k\}, I = 1, \dots, K\}$, where K is the total number of WSIs in the dataset and k is the number of tumour tiles in image I. All the tile extraction, stain normalization and tumour tile classification steps can be done using TIAToolbox (Pocock et al., 2022).

3.3. Cell network generation

In our method, the potential cell-to-cell relationships are represented by a meaningful network (graph $G = (\mathbf{V}, \mathbf{L})$) constructed from a histological image. In such a network, nuclei are considered as network nodes ($\mathbf{V}$) and their potential relationships as network edges ($\mathbf{L}$). Accordingly, we use these two procedures to create cell networks for each tumour tile in the dataset (as shown in Fig. 2b): (1) node identification by nuclei centroid detection and (2) finding the network's edges configuration using K-Nearest Neighbours (KNN) algorithm.

3.3.1. Detection of nuclei as network's nodes

Because we do not incorporate nuclear morphological features in our cell network, our approach does not require exact nuclei segmentation. We are only interested in nuclei relationships, therefore, detecting the centroid position of each nucleus would suffice. To achieve cell detection, we train Efficient-Unet (Jahanifar et al., 2021) with nuclei inverse distance maps ($\tilde{M}_d$) as ground truth. In particular, we first extract the centroid of each instance and place a point on an empty canvas to form a centroid map, M_c. Then, euclidean distance map M_d is generated where the value of each pixel (i, j) is equal to its distance to the closest centroid point in M_c, (i_c, j_c). In other words, $M_d(i, j) = \sqrt{(i - i_c)^2 + (j - j_c)^2}$. Having M_d distance map, we can calculate the inverse distance map for an image as below:

$$\tilde{M}_d = s \cdot \frac{e^{\sigma(1 - M_d/\tau)} - 1}{e^{\sigma} - 1}, \tag{1}$$

where $\tau = 15$ is the distance threshold to the nucleus centroid, which is used to clip distances more than τ to 0, and σ is a smoothing parameter set to 3. The scaling parameter, $s = 3$, is also used to scale up all the original values from range $[0, 1]$ into range $[0, 3]$ so that the loss function, mean squared error (MSE) in our case, can better learn from the ground-truth and penalize more for mistakes. We incorporate MSE as the loss function because inverse-distance map prediction is a pixel-wise regression task and MSE proved to perform well with this kind of problem. A sample input and output prediction (inverse distance map) are shown in Fig. 3.

Fig. 3. Nuclei detection by distance map prediction using Efficient-UNet.

For post-processing at the inference time, we use a simple peak detection algorithm which finds local maximums in the prediction map that are at least 11 pixels apart. We also neglect weak peaks that have prediction values less than a third of the highest peak in the prediction map. The final collection of refined peaks forms the graph node set: $\mathbf{V} = \{v_i = (x_i, y_i) \,|(x_i, y_i) \in \text{Peaks}(\tilde{M}_d)\}$.

We use the Lizard dataset (Graham et al., 2021a,b) to train our nuclei detection model which is the largest and the most versatile publicly available nuclei instance segmentation and classification dataset, containing nearly 500,000 labelled nuclei in H&E stained colon tissue.

3.3.2. Finding network edges

We consider an edge between two nuclei if their Euclidean distance is less than a particular radius because we believe that nearby cells in a histology sample are more likely to interact with one another. Therefore, in our cell graph, G, we consider an edge between two nodes (v_i, v_j) if their euclidean distance, $\text{D}(\text{v}_\text{i}, \text{v}_\text{j})$, is smaller than a certain radius, r. However, in order to prevent over-connecting a node, we further confine the number of edges to the k-nearest neighbours. Therefore, the (i, j) cell of our graph's adjacency matrix, $\mathbf{A}_{ij}$, for two nodes (v_i, v_j) is set to 1 (indicating an edge between the nodes) if those nodes are k-nearest neighbours to each other and have their Euclidean distance less than a radius. In other words, our graph edges can be defined as follows: $\mathbf{E} = \{(v_i, v_j) \,|v_i \in \text{KNN}(v_j) \wedge \text{D}(v_i, v_j) \leq r\}$. The whole process of cell network generation is depicted in Fig. 2b where results of cell detection and network edge formation are shown for a sample tumour tile.

3.4. SNA measures

Raw data from a large cell network is not easily understandable, therefore, we need some measures to capture the informative characteristics of the network structures and describe them quantitatively. In the tumour microenvironment, we assume that cells are social actors, and we derive a number of metrics from the SNA field (Newman, 2018; Scott, 1988) to describe their relationships. In particular, the following six SNA measures are incorporated into this study:

Node Degree (ND): The number of edges that connect a node to other nodes in a network is known as its node degree, $\deg(v)$. For instance, a cell's ND in a cell social network represents how many other cells it is directly associated with. If nuclei co-exist in dense packs (for example in tumour regions), one can expect high values of ND for all the nodes in that region.

Clustering Coefficient (CL): The percentage of triangles $\text{Tri}(v)$ that are actually formed across a node is known as a clustering coefficient, which quantifies the density of triangles in a network:

$$\text{CL}(v_i) = 2\text{Tri}(v_i)/(\deg(v_i)(\deg(v_i) - 1)) \tag{2}$$

The clustering coefficient calculates how many cells in a social network of cells tend to group together. For example, this measure can highlight inflammatory clusters where cells tend to form large clusters.

Closeness Centrality (CC): The closeness centrality of a node v_i is the reciprocal of the average of the shortest-path distances $\text{d}(.,.)$ between that node and all other nodes in the network:

$$\text{CC}(v_i) = \frac{N-1}{\sum_{j,j\neq i} \text{d}(v_i, v_j)}, \tag{3}$$

where N is the total number of cells in the network. Nodes that can easily access other nodes are highlighted by closeness centrality. In the analogy of a cell social network, a cell that has a higher closeness centrality is closer (in network terms) to every other cell.

Degree Centrality (DC): The node degree is normalized to determine the network's simplest centrality definition or degree centrality:

$$\text{DC}(v_i) = \frac{\sum_{j=1}^{N} \mathbf{A}_{ij}}{N-1} \tag{4}$$

This measure includes network size information in addition to node degree.

Eigen Vector Centrality (EC): This centrality gauges a node's influence within the network. EC is a variation on degree centrality in which each node's score is derived from the scores of the nodes that surround it. Therefore, if a node in the network is connected to many nodes or certain nodes with high degrees, it may have a high EC score. We define the eigenvector centrality for node v_i as x_i:

$$\text{EC}(v_i) = \lambda^{-1} \sum_{v_j \in \mathcal{N}(v_i)} \mathbf{A}_{ij} \text{EC}(v_j) \tag{5}$$

where λ is eigenvalue and $\mathcal{N}(v_i)$ is the set of neighbours of v_i (Newman, 2018). The power iteration algorithm is used to solve the above equation and find the EC of nodes. EC in cell social networks reveals the cells with significant network influence.

Katz Centrality (KC): Similar to EC, Katz centrality is a generalization of degree centrality in which each node's centrality score is determined by taking into account all linked neighbours, whether they are close-by or far-away (Katz, 1953). Each neighbouring node in KC will receive an initial constant centrality, β, while contributions from far-off nodes will be penalized with an attenuation factor, α, during the calculation of Katz centrality for the node v_i:

$$\text{KC}(v_i) = \alpha \sum_{v_j \in \mathcal{N}(v_i)} \mathbf{A}_{ij} \text{KC}(v_j) + \beta_i \tag{6}$$

Please note that, unlike other works based on cell networks (Zhou et al., 2019), nuclei sampling is not necessary before extracting SNA measures since they are not constrained to computationally intensive GNNs. To better understand the concept of each introduced network measure, you can refer to Fig. 7 where we illustrate two sample cell networks with their nodes coloured based on values of different SNA measures.

3.5. Statistical SNA features

All the introduced SNA measures in the previous section are calculated per node in the network. However, it is challenging to directly incorporate nodes' SNA measures in the classical ML or DL pipeline because the number of nodes is often too high and always varies from one sample to another (the number of nuclei in different WSIs is not the same). Therefore, following Zamanitajeddin et al. (2021), we extract statistical features of all SNA measures from each slide to distil the cell relationship information of those WSIs into fixed-size and relatively low-dimensional feature vectors.

In particular, we find the distribution of values of every SNA measure by calculating their histogram (with 10 specified bins) and then concatenate the resulting histogram counts with the maximum, minimum, mean, standard deviation, and median of SNA measure to build a feature vector, $\mathbf{e}$. In total, for each cell social network, we have 6 sets of SNA measures and for each of those we extract 15 statistical features which will result in a feature vector of size 90 elements to describe the cell relationships in that social cell network i,e, $\mathbf{e} \in \mathbb{R}^{\mathbf{1\times 90}}$. The collection of SNA feature vectors forms 'SNA Feature Dataset', $\mathbf{S_E}$, which contains sets of feature vectors $(\mathbf{E}_I)$ for K images in the dataset i.e., $\mathbf{S_E} = \{\mathbf{E}_I | \mathbf{E}_I = \{\mathbf{e}_1, \ldots, \mathbf{e}_k\}, I = 1, \ldots, K\}$. Here, k is the number of tumour tiles in image I and for every tumour tile exists one SNA feature vector $\mathbf{e}$.

It is worth mentioning that SNA features collect information from the cell network by focusing exclusively on cell-to-cell connections and recording spatial relationships between them. They are not reliant on exact nucleus segmentation or feature extraction. Furthermore, SNA-based features are easily scaleable since SNA measure extraction is computationally efficient and is not restricted by GPU memory. The process of calculating SNA measures and their statistical features is illustrated in Fig. 2c. We provide an extensive list of key SNA features and their interpretation in cell networks in Table S3 of supplementary material Section S2, offering insights into their significance and implications in the context of our study.

3.6. Embedding SNA features in DL methods

The extracted feature vectors for cell networks can be directly compared, used to train a classical ML model (Zamanitajeddin et al., 2021), or incorporated in DL pipelines as an auxiliary input to enhance a histology image classifier that uses only image representations.

Typically, WSIs have a large number of tiles that may not be related to the slide label. It can be difficult to decide which tiles to use for training a DL model, and there is often no cellular-level annotation available. Traditional methods of training a deep neural network on all tiles (such as Kather et al., 2019b) of the training set are not efficient with only a slide-level label available. This can lead to training bias and inefficiency. Randomly selecting tiles from each WSI may not improve prediction accuracy, and selecting the most predictive tiles in a brute-force manner is computationally infeasible.

In this study, we choose to work with two SOTA DL-based weakly-supervised approaches, namely CLAM (Lu et al., 2021b) and IDaRS (Bilal et al., 2021), that proposed different ways to alleviate these challenges. Although these methods share similar principles, there are some differences in their training and implementation that make a great gap in their performances. Both methods utilize a CNN (different variants of ResNet) to extract deep image features from tiles, $\mathbf{z} = \text{CNN}(\mathbf{t})$ where $\mathbf{z} \in \mathbb{R}^{\mathbf{1\times1024}}$.

The idea of this study is to combine deep image features with SNA features to generate a better representation of the image and then feed that into either CLAM or IDaRS (or any other DL-based MIL/weakly-supervised system) instead of using only deep features. However, directly doing that will dilute SNA features as there exists a dimension imbalance between $\mathbf{z}$ and $\mathbf{e}$ (1024 vs. 90). Therefore, we introduce dimension-reduction function Proj(.) to actively reduce the dimension of $\mathbf{z}$:

$$\mathbf{h} = \text{Proj}(\mathbf{z}) = \text{ReLU}(\mathbf{W}\mathbf{z}^{\text{T}} + \mathbf{b}) \tag{7}$$

where $\mathbf{W} \in \mathbb{R}^{\mathbf{90\times1024}}$ and $\mathbf{b} \in \mathbb{R}^{\mathbf{90\times1}}$ are trainable parameters of reduction function. By doing this, projected deep features $\mathbf{h}$ and SNA features $\mathbf{e}$ can be efficiently combined (by concatenation) to form a feature vector of $\mathbf{f} = [\mathbf{h}, \mathbf{e}] \in \mathbb{R}^{\mathbf{1\times180}}$. The feature vector $\mathbf{f}$ is fed into the next layers of the network, Cls(.) function, which is either a 3-layer multilayer perceptron (with 64, 64, 2 output nodes) in IDaRS or the attention mechanism network in CLAM (Lu et al., 2021b). In either case, we can consider the training of the whole deep neural network model $\phi((\mathbf{t}, \mathbf{e}), \boldsymbol{\psi}) = \text{Cls}\left(\mathbf{f} = \left[\text{Proj}(\text{CNN}(\mathbf{t})), \mathbf{e}\right]\right)$ with trainable parameters $\boldsymbol{\psi}$ as a weakly-supervised machine learning problem with the following empirical error formulation:

$$\min_{\boldsymbol{\psi}} \frac{1}{|\mathbf{S}|} \sum_{(\mathbf{T}_I, \mathbf{E}_I) \in \{\mathbf{S_T}, \mathbf{S_E}\}} \frac{1}{|\mathbf{S_T}|} \sum_{(\mathbf{t}, \mathbf{e}) \in \{\mathbf{T}_I, \mathbf{E}_I\}} \mathcal{L}(y_I, \phi((\mathbf{t}, \mathbf{e}), \boldsymbol{\psi})) \tag{8}$$

where $\mathbf{S} = \{\mathbf{S_T}, \mathbf{S_E}\}$ is the collection of Tiles and SNA features datasets, $\mathcal{L}$ is the loss function, and $y_I \in \{-1, +1\}$ is the WSI-level label (indicating if image I belongs to the positive or negative class).

As we mentioned before, there are many differences between the CLAM and IDaRS in the implementation of Eq. (8). For example, IDaRS uses a ResNet34 model (He et al., 2016) as the CNN(.) function whose parameters are trainable and fine-tuned during the training process whereas CLAM utilizes a ResNet50 with fixed pre-trained ImageNet weights. Furthermore, IDaRS incorporates a special iterative draw-and-rank sampling strategy to choose a collection of random and previously top-ranked tiles in each iteration (Bilal et al., 2021) while CLAM includes all the tiles and weights the contribution of each tile in the loss calculation using attention score (Lu et al., 2021b) retrieved from attention layer in CLAM's Cls function. The loss functions $\mathcal{L}$ between the two methods are also different where IDaRS utilizes symmetric cross-entropy whereas CLAM uses standard cross-entropy loss function. Explaining the exact details of each algorithm is outside of the scope of the current manuscript, therefore, please refer to Bilal et al. (2021) and Lu et al. (2021b) for more information.

3.7. Inference and prediction aggregation

Our proposed method is still a tile classifier that incorporates cell relationship information in the prediction pipeline. Therefore, when we want to test the model on a new WSI, we need to pre-process the WSI according to the steps outlined in Fig. 2 to find tumour tiles and their SNA feature vectors to be fed into the predictor model to return a category probability for the task that model has been trained for. This will result in a set of prediction probabilities for all the tumour patches in the slide and we will need an aggregation approach to integrate all those scores into a WSI-level score. The aggregation method is different for IDaRS and CLAM use cases. For IDaRS, we use the average score of all the tumour tiles that have their score larger than the median score of all the tiles in the WSI. Similar to Lu et al. (2021b), for CLAM, the slide-level representation is computed by taking the average of all patches in the slide, with each patch's contribution being weighted by its attention score.

4. Material and results

4.1. Dataset and data split

In this study, we employ the TCGA-CRC-DX cohort, which includes 499 patients and 502 diagnostic slides of primary colorectal tumours provided by 36 different US centres. This cohort is the same as the one used by Bilal et al. (2021), Kather et al. (2020), Echle et al. (2020) and Kather et al. (2019b) to present fair comparisons. As a part of the cancer genome atlas (TCGA) (Weinstein et al., 2013), the TCGA-CRC-DX cohort comes with comprehensive medical data for each subject but we only use the molecular data provided by Liu et al. (2018) to define binary labels for the training of the proposed method for binary classification.

In particular, we consider training five different models each trying to predict one of the following entities where the binary labels are defined:

- Hypermutated (HM) tumour prediction: high mutation density (HMD, +) vs. low mutation density (LMD, −).
- MSI pathway prediction: microsatellite instable-High (MSI-High or MSI, +) vs. microsatellite stable and MSI-low tumours (MSS and MSI-Low combined, −) following Kather et al. (2019b) and Bilal et al. (2021)
- Chromosomal instability (CIN) pathway: CIN (+) vs. genome stable (GS, −).
- *BRAF* mutation: mutant (MUT, +) vs. Wild Type (WT, −).
- *TP53* mutation: mutant (MUT, +) vs. Wild Type (WT, −).

Although a cohort of 502 slides is used in experiments related to *BRAF* and *TP53* prediction, we can only include a smaller subset of the cohort for HM, CIN, and MSI prediction due to missing metadata or insufficient tumour tiles. A summary of the incorporated number of slides and their category is reported in Fig. 4. Nearly 450,000 tumour tiles are extracted from the entire TCGA-CRC-DX cohort using the method explained in Section 3.2.

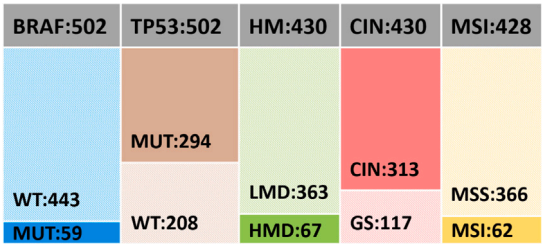

Fig. 4. Number of slides from TCGA-CRC-DX dataset used for different cross-validation experiments. In each category, the solid and shaded colours represent the positive and negative classes, respectively.

We further incorporated a cohort from Pathology Artificial Intelligence Platform (PAIP) challenge[1] for MSI prediction in colorectal cancer, which contains 35 microsatellite stable (MSS) and 12 microsatellite instable (in total 47) slides. These slides were acquired from 3 different South Korean centres where PCR tests were used to detect the microsatellite status of each tumour.

4.2. Evaluation metrics

We follow the convention used in previous studies (Bilal et al., 2021; Kather et al., 2019b, 2020; Echle et al., 2020) to evaluate the binary classification done in each experiment. First, the receiver operating characteristic curve (ROC) and precision–recall curve (PRC) are extracted by plotting recall against the false positive rate and precision against recall as the discrimination threshold is varied, respectively. ROC curve can better highlight the true positive rate (TPR) or sensitivity of the classifier while PRC can show how much the returning results are accurate. High areas under both ROC and PRC indicate high recall, low false positive rate, and high precision. Therefore, we use Area Under ROC (AUROC) and Area Under PRC (AUPRC) as two metrics for the evaluation of binary classification.

4.3. Internal cross-validation

All internal cross-validation experiments in this paper are carried out with four folds where two folds are used for training, one fold serves as a validation set for identifying the top-performing model, and a fourth fold is held out as an unseen test set on which the model's performance is assessed.

Results of cross-validation experiments for CIN, HM, *TP53*, *BRAF*, and MSI status prediction are reported in Tables 1 to 5, respectively. In all the tables, we compare the performance of the proposed Social-CLAM and Social-IDaRS methods, which benefit from SNA features as well as deep features, with the original CLAM (Lu et al., 2021b) and IDaRS (Bilal et al., 2021) methods which are SOTA models using only image (deep) features, SNA features in classical ML setting (Zamanitajeddin et al., 2021), and MIL-SNA (using only SNA features in multiple instance learning frameworks such as IDaRS).

Based on Table 1 for CIN prediction, our proposed Social-IDaRS method is able to outperform all other methods and achieve SOTA AUROC of 0.85 improving the results by 2% in comparison to the original IDaRS. Even higher improvement is seen for Social-CLAM in comparison to CLAM (more than 4%), showing the relevance and the added value of using SNA features alongside deep image features. In comparison to the previously published results in Kather et al. (2020) which used a DL classifier, performance improvement by Social-IDaRS is much higher (about 12%). Social-IDaRS also achieves the highest AUPRC value of 0.93 for CIN tumour prediction which shows both high sensitivity and precision of the proposed method for the detection of chromosomal instability pathway.

[1] Available at: https://paip2020.grand-challenge.org/.

Table 1
Results of internal cross-validation experiments on TCGA dataset for CIN status prediction.

Method	AUROC	AUPRC
Kather et al. (2020)	0.730	–
Zamanitajeddin et al. (2021)	0.670 ± 0.04	0.830 ± 0.04
MIL-SNA	0.700 ± 0.04	0.850 ± 0.03
CLAM (Lu et al., 2021b)	0.692 ± 0.09	0.852 ± 0.04
IDaRS (Bilal et al., 2021)	0.830 ± 0.02	0.920 ± 0.01
Social-CLAM (Proposed)	0.735 ± 0.06	0.867 ± 0.05
Social-IDaRS (Proposed)	**0.850** ± 0.03	**0.930** ± 0.02

Table 2
Results of internal cross-validation experiments on TCGA dataset for HM status prediction.

Method	AUROC	AUPRC
Zamanitajeddin et al. (2021)	0.680 ± 0.09	0.290 ± 0.03
MIL-SNA	0.560 ± 0.00	0.270 ± 0.04
CLAM (Lu et al., 2021b)	0.717 ± 0.09	0.405 ± 0.10
IDaRS (Bilal et al., 2021)	0.810 ± 0.03	0.570 ± 0.09
Social-CLAM (Proposed)	0.717 ± 0.11	0.517 ± 0.13
Social-IDaRS (Proposed)	**0.870** ± 0.01	**0.630** ± 0.08

Table 3
Results of internal cross-validation experiments on TCGA dataset for *TP53* status prediction.

Method	AUROC	AUPRC
Kather et al. (2020)	0.640	–
Zamanitajeddin et al. (2021)	0.580 ± 0.05	0.690 ± 0.05
MIL-SNA	0.630 ± 0.06	0.710 ± 0.05
CLAM (Lu et al., 2021b)	0.639 ± 0.03	0.697 ± 0.04
IDaRS (Bilal et al., 2021)	0.730 ± 0.02	0.780 ± 0.04
Social-CLAM (Proposed)	0.680 ± 0.04	0.775 ± 0.05
Social-IDaRS (Proposed)	**0.750** ± 0.04	**0.830** ± 0.03

For tumour HM prediction in Table 2, although SNA features in classical ML methods or MIL models cannot achieve high AUROC values (0.68 and 0.56, respectively), they can considerably improve the AUROC of HM prediction by 6% (in comparison to original IDaRS) when used as auxiliary information in the proposed Social-IDaRS method where we achieve AUROC of 0.87. Furthermore, the Social-IDaRS outperforms all other methods in terms of AUPRC, showing its capability of hypermutated tumour detection. Although the AUROC of Social-CLAM in HM prediction does not increase with the introduction of SNA features, the AUPRC increases about 11% in comparison to the original CLAM.

Based on Table 3, the *TP53* mutation prediction is also improved by 11% when comparing the Social-IDaRS with Kather et al. (2020) where the former obtains AUROC of 0.75. Once again, adding SNA features to image features in both Social-CLAM and Social-IDaRS helps improve AUROC by about 4% and 2% in comparison to the original CLAM and IDaRS methods, respectively. However, SNA features in classical ML or MIL methods or using only image features in an attention model (CLAM) do not seem very predictive for *TP53* mutation prediction, all resulting in AUROC in the range of 0.58 to 0.63. Moreover, Social-IDaRS is reported to be considerably more precise than other methods by achieving the highest AUPRC of 0.83 (5% improvement in comparison to the original IDaRS).

Although the Social-IDaRS outperforms all other methods by high margins for *BRAF* mutation prediction based on Table 4 (for example, improving AUROC by 13% and 15% over Kather et al., 2020 and Lu et al., 2021b methods, respectively), it seems that using SNA features does not boost the sensitivity of IDaRS where both IDaRS and Social-IDaRS achieve the same AUROC of 0.79. However, Social-IDaRS significantly improves the AUPRC by 6% over IDaRS and SNA features help Social-CLAM to increase *BRAF* prediction AUROC and

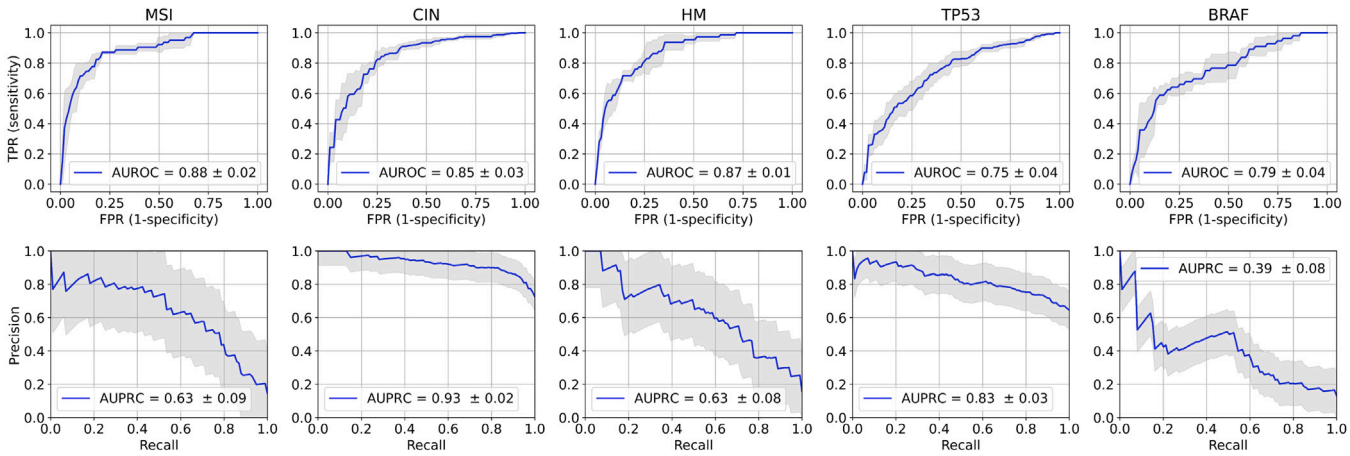

Fig. 5. The average ROC (first row) and PR (second row) curves related to Social-IDaRS performance on MSI, CIN, HM, *TP53*, and *BRAF* prediction tasks. The shaded region represents the standard deviation of results over 4-fold cross-validation.

Table 4
Results of internal cross-validation experiments on TCGA dataset for *BRAF* status prediction.

Method	AUROC	AUPRC
Kather et al. (2020)	0.660	–
Zamanitajeddin et al. (2021)	0.680 ± 0.04	0.280 ± 0.11
MIL-SNA	0.590 ± 0.06	0.190 ± 0.04
CLAM (Lu et al., 2021b)	0.638 ± 0.06	0.249 ± 0.10
Transformer (Wagner et al., 2023)	0.78	–
IDaRS (Bilal et al., 2021)	**0.790** ± 0.01	0.330 ± 0.05
Social-CLAM (Proposed)	0.738 ± 0.06	0.344 ± 0.10
Social-IDaRS (Proposed)	**0.790** ± 0.04	**0.390** ± 0.08

Table 5
Results of internal cross-validation experiments on TCGA dataset for MSI status prediction.

Method	AUROC	AUPRC
Echle et al. (2020)	0.740	–
Zamanitajeddin et al. (2021)	0.490 ± 0.07	0.300 ± 0.06
MIL-SNA	0.630 ± 0.07	0.240 ± 0.06
CLAM (Lu et al., 2021b)	0.742 ± 0.08	0.439 ± 0.12
Transformer (Wagner et al., 2023)	0.83	–
IDaRS (Bilal et al., 2021)	0.860 ± 0.04	0.620 ± 0.10
Social-CLAM (Proposed)	0.760 ± 0.01	0.487 ± 0.09
Social-IDaRS (Proposed)	**0.880** ± 0.02	**0.630** ± 0.10

AUPRC by 10%. Once again, these results indicate the importance of incorporating SNA features as additional information together with pixel-based information.

Finally, in Table 5 for MSI status prediction, the proposed Social-IDaRS method which benefits from both pixel-based image features and SNA features gains AUROC and AUPRC of 0.88 and 0.63, respectively, significantly outperforming both methods based on only SNA features (MIL-SNA by 25% and ML-SNA (Zamanitajeddin et al., 2021) 39%) and both weakly-supervised DL methods based on only deep features (IDaRS by 2% and CLAM by 13%). Comparing the attention-based method of CLAM with the DL-based method of Echle et al. (2020) shows 2% improvement indicating the importance of the attention mechanism. But when IDaRS is compared to CLAM, the performance is notably increased by 11% which highlights the importance of learning image representation from histology images instead of using ImageNet pre-trained weights as well as the efficiency of the iterative draw-and-rank sampling strategy of IDaRS in comparison to attention mechanism in Lu et al. (2021b). Furthermore, using SNA features as additional information to pixel-based features in Social-IDaRS and Social-CLAM adds an extra 2% boost in performance compared to the original IDaRS and CLAM methods.

In our experiments, we considered recent transformer-based approaches, specifically the method proposed by Wagner et al. (2023). For *BRAF* status prediction, the Transformer achieved an AUROC of 0.78, whereas our proposed Social-IDaRS produced comparable performance with an AUROC of 0.790 ± 0.04. Similarly, for MSI status prediction, the Transformer reported an AUROC of 0.83, while our Social-IDaRS method outperformed it with an AUROC of 0.880 ± 0.02. This demonstrates that while transformer-based methods like that of Wagner et al. (2023) have indeed achieved commendable results, our proposed methodology consistently matches or surpasses these benchmarks, emphasizing its efficacy and potential in the given context.

Furthermore, we have repeated similar experiments for *NRAS* and *KRAS* prediction tasks, where our analysis revealed that the predictive values are considerably low across all the methods. In particular, the best performance metrics were AUROC of 0.63 and AUPRC of 0.15 for *NRAS* status prediction which did not achieve the robustness observed for other molecular alterations. Based on these results, we decided to focus on tasks with higher predictive validity in the main manuscript to ensure that the subsequent analysis and discussions derived from the results stand robust and are not misconstrued or overinterpreted. For a detailed assessment of the models' performance on *NRAS* and *KRAS* predictions and to gain insights into the challenges of predicting *NRAS* and *KRAS* statuses, we refer the reader to the supplementary material Section S1.

On average over all prediction tasks, the Social-IDaRS method raises the AUROC and AUPRC of predictions by 2.4% and 8.8% over IDaRS (as well as about average 4% and 7% improvement in AUROC and AUPRC for Social-CLAM over the original CLAM) which indicates the additional value that SNA features can bring into DL-based multiple instance learning frameworks. Additionally, the average standard deviations of 0.026 and 0.064 for the AUROC and AUPROC of the Social-IDaRS method, respectively, represent the stability of the proposed method when making predictions.

We summarize the performance of the best model, Social-IDaRS, for MSI, CIN, HM, *TP53*, and *BRAF* prediction tasks on the TCGA dataset in Fig. 5 where ROC and PR curves are illustrated, and the average and standard deviation of areas under curves are reported. Furthermore, confusion matrices related to these experiments are provided in Section S4 of supplementary materials.

4.4. External cross-validation

For external validation of different methods on the MSI status prediction task, we first train algorithms on the TCGA-CRC-DX cohort

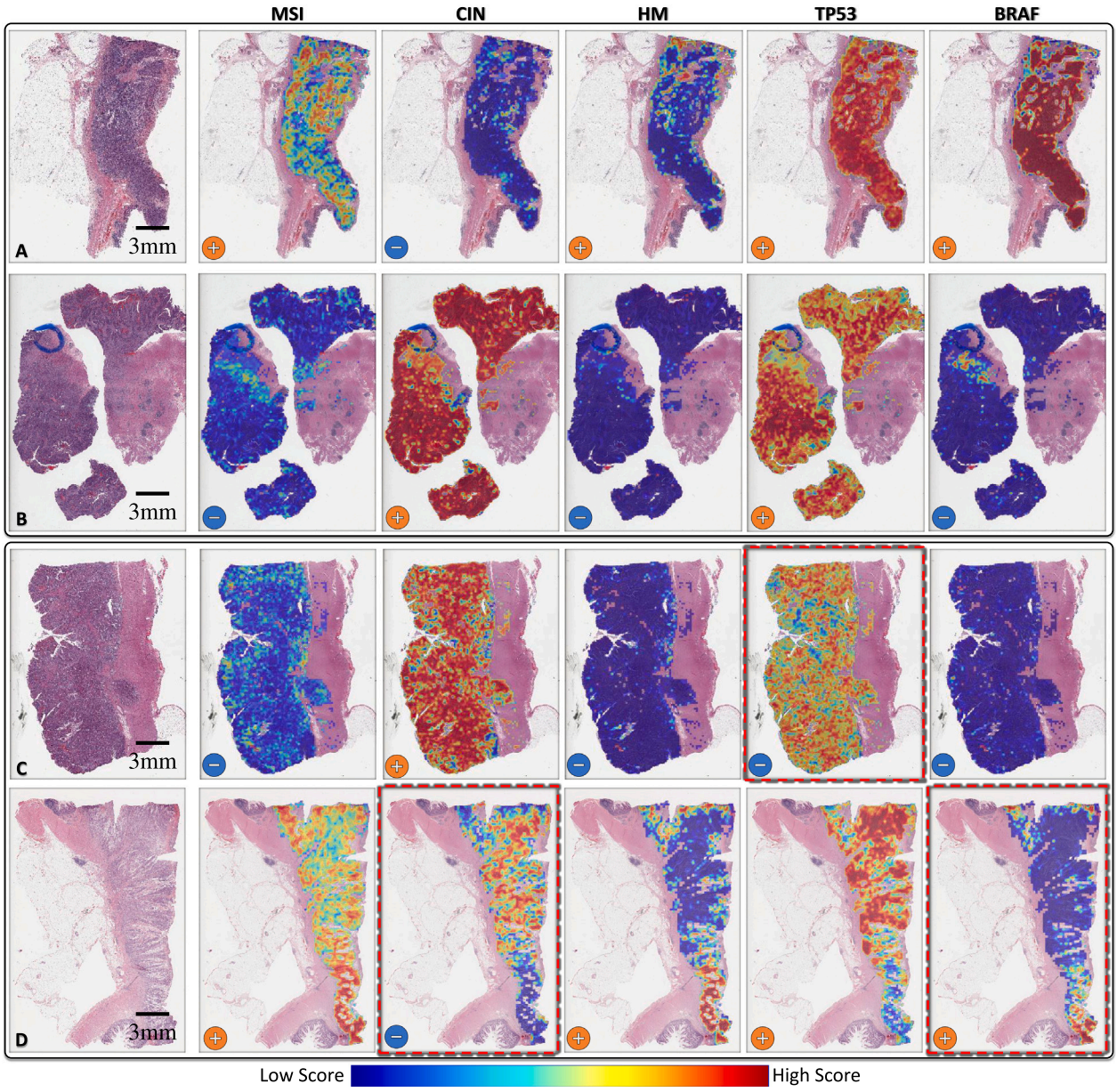

Fig. 6. Prediction maps of CIN, HM, *TP53*, *BRAF*, and MSI for four cases from the TCGA cohort. The top two rows show successful prediction of all mutations while the bottom two rows showcase at least one wrong prediction which is highlighted by a red dashed border. Positive and negative signs indicate the ground truth labels for each mutation. For a more detailed view of these results, please visit our demo viewer at https://tiademos.dcs.warwick.ac.uk/bokeh_app?demo=sna.

and then test them on the PAIP dataset. The results for this experiment are reported in Table 6 where the proposed Social-IDaRS achieves the AUROC (AUPRC) of 0.99 (0.98) outperforming IDaRS by 1% (3%), CLAM by 15% (33%), and ML-SNA (Zamanitajeddin et al., 2021) by 40% (47%). Notably, the performance of Social-CLAM is also increased by 4% and 14% for AUROC and AUPRC, which implies that utilizing SNA features in DL models can improve generalizability and increase the precision of the model considerably.

Table 6
Results of cross-validation experiments on PAIP2020 dataset for MSI status prediction.

Method	AUROC	AUPRC
Zamanitajeddin et al. (2021)	0.59	0.51
CLAM (Lu et al., 2021b)	0.84	0.65
IDaRS (Bilal et al., 2021)	0.98	0.95
Social-CLAM (Proposed)	0.88	0.79
Social-IDaRS (Proposed)	**0.99**	**0.98**

4.5. Qualitative results

To visualize the performance of the best performing method (Social-IDaRS) for various predictions on WSIs, we illustrate the prediction heatmaps overlay on four example WSIs in Fig. 6. For each problem, the prediction heatmaps are generated by placing the prediction scores returned by the problem-specific Social-IDaRS model for extracted tiles on the relative positions. Then, heatmaps are smoothed using a Gaussian filter for better visualization. In Fig. 6, prediction maps for CIN, HM, *TP53*, *BRAF*, and MSI prediction problems are depicted where the ground truth labels are shown with + and − signs inside each case indicating the binary label for each category based on Fig. 4 and Section 4.1 notation.

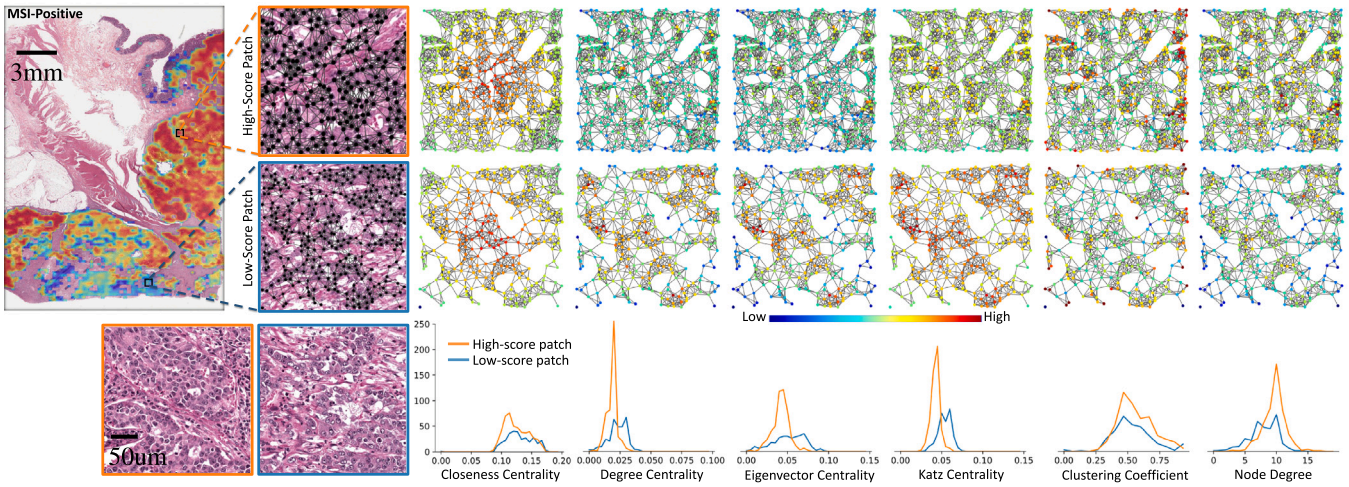

Fig. 7. Cell networks with nodes coloured based on SNA measures for two patches of an MSI-High sample. The plots in the bottom row show the histogram of SNA measure for the high-score and low-score patches.

The idea is that if the GT label for a category is positive more high-score (reddish) regions and if the label is negative more low-score (bluish) regions should be visible in the heatmap. The top two rows in Fig. 6 (samples A and B) show accurate predictions for all categories where heatmap colours for negative cases are mostly bluish whereas positive cases present more reddish regions. Please note that the thresholds used for different experiments (CIN, HM, etc.) are different, therefore, for a CIN case a high threshold can be utilized to only select cases with lots of high-score regions as positives. On the other hand, for HM prediction, even if small areas of the region are positive it can be considered positive. The proposed Social-IDaRS is capable of capturing positive cases even with a small number of positive tiles due to its special score aggregation strategy. The heatmap high-score region distribution for different categories can also have clinical implications where for CIN, *TP53*, and *BRAF* positive cases almost the whole area of the tumour show a high-score signal (the whole tumour positive and even some small positive regions can be seen for negative cases), whereas for HM, positive signals are only presented in local regions of the positive tumour and negative cases show almost no high-score regions indicating that hypermutated tumour can be found only locally.

Samples C and D that are shown in the bottom two rows of Fig. 6 contain 1 and 2 wrong predictions where the GT label for the category is negative but many high-score regions are predicted or vice versa. These wrong predictions are highlighted using red dashed borders. In particular, for sample C, *TP53* prediction shows many high-score regions whereas the GT label for this category is negative. On the other hand, for *BRAF* prediction in sample D, the GT label is positive but only a small region of the tumour is predicted as high-score *BRAF* mutated while for this category it is required to see large high-score regions for a positive case.

To better highlight the relevance of proposed statistical SNA features, we apply the Social-IDaRS model for MSI status prediction on an MSI-High sample and overlay the prediction heatmap on the WSI overview in Fig. 7. Then, we select the patches with the highest (orange border) and lowest (blue border) prediction scores to plot the cell networks for them. The cell networks for the selected patches are repeated in 6 columns, each column dedicated to a different SNA measure, where node colour highlights the value of specific SNA measure at the node (blue for low values and red for high values). Furthermore, we extract the distribution (histogram) of all SNA measures for both high-score and low-score patches and plot them separately in the last row of Fig. 7 where orange and blue distribution curves belong to high-score and low-score patches, respectively.

In Fig. 7, it is obvious that depending on the arrangement and connections of the nuclei in high-score and low-score patches, the

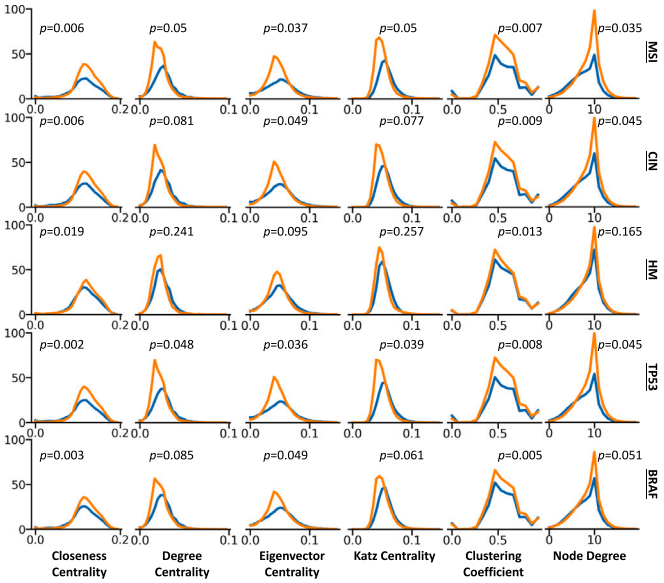

Fig. 8. Statistical comparison of average distributions of SNA measures over the top 20 most attentive regions of Social-IDaRS model for positive (orange curves) and negative (blue curves) samples of various prediction tasks.

SNA measure values for each node differ. Take the degree centrality (DC) measure for example, where the colour of most of the nodes in the high-score patch is green-blue whereas, in the low-score patch, all colours can be seen, especially lots of orange nodes. This means that the high-score patch has more nuclei with lower degree centrality (shown as a large peak at lower values of degree centrality distribution in Fig. 7) while nodes in the low-score patch's cell network show degree centrality values over a wider range with no distinguished peak anywhere in the range (blue degree centrality histogram in Fig. 7). Although it may be hard to distinguish differences in SNA measures of cell networks, the histograms of values of different SNA measures show distinct differences between the high-score and low-score patches.

4.6. Statistical analysis of cell distributions

Previously, we illustrated values of the SNA measures and their related distributions only for a couple of low-score and high-score patches in Fig. 7. Although that visualization can give qualitative insights into

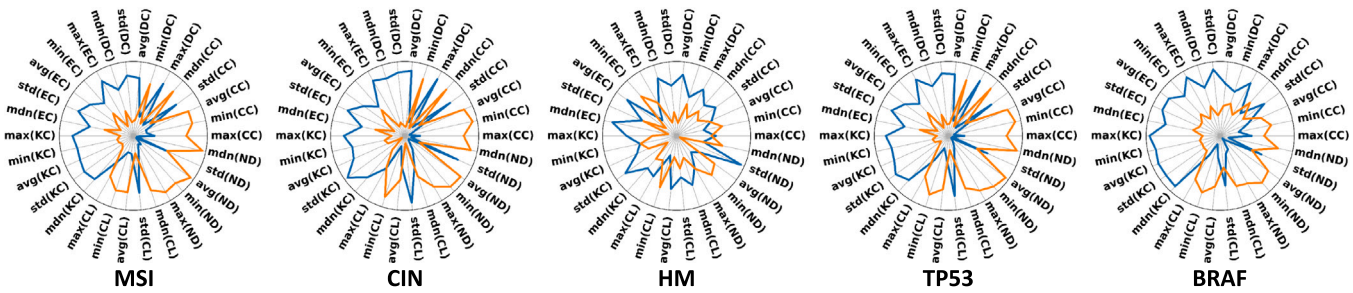

Fig. 9. Radar plot of five statistics (max: maximum, min: minimum, avg: average, std: standard deviation, mdn: median) of different SNA measures (CC: Closeness Centrality, DC: Degree Centrality, EC: Eigenvector Centrality, KC: Katz Centrality, CL: Clustering Coefficient, ND: Node Degree) over the top 20 most attentive regions of Social-IDaRS model for positive (orange colour) and negative (blue colour) samples of various prediction tasks.

the role of SNA in measuring cell relationship, it may not be clear how much SNA measures can differentiate between positive and negative samples of different cell alterations over the whole population. To do this, we select the top 20 most attentive tiles from positive and negative samples of each problem based on Social-IDaRS scores, then extract their distribution and statistical features, calculate the average for positive and negative classes, and finally visualize them in Figs. 8 and 9.

In Fig. 8, the average distribution for positive (orange) and negative (blue) samples of different problems are depicted in different rows. Also, in each row, the distribution of different SNA measures is plotted separately so we can fairly compare them. At first glance, it is evident that there is a clear differentiation between the average distribution of positive and negative patches for all SNA measures through all the tasks. Take MSI prediction for example (first row of Fig. 8), the average distribution curves between positive (MSI-High) and negative (MSI-Low and MSS) samples are clearly different, either showing different peaks (position and value) or having different standard deviation. When focusing on the bottom left part of Fig. 8, one can see that for positive MSI samples, the population of nodes (cells) with various closeness centralities is always higher than that of negative samples. Referring to the closeness centrality definition in Eq. (3), this implies that in positive MSI-High samples cells are generally (on average) closer to each other in comparison to negative MSS samples. This can be seen for all other tasks as well (very much highlighted in CIN and *TP53*). Similar behaviour can be described for the clustering coefficient, where for almost every CCo mid-range value, positive samples have a higher number of nuclei than negative samples, suggesting that in positive samples cells tend to form clusters more than in negative samples (however this is not very evident in HM prediction task).

Similarly, the peak property of distributions of other SNA measures can help interpret the overall nuclei relationships between different populations. For example, distributions for both Eigenvector and Katz centralities show different peaks over positive and negative samples. In the third column from the left, it is obvious that in positive samples, there are considerably higher numbers of cells with low Eigenvector centrality values where a dominant peak arises around the Eigenvector value of 0.04. Whereas, for negative samples, the distribution is flatter with not much dominant peak (see the third column of the first row in Fig. 8). As Eigenvector centrality tries to measure node influence (nodes that have a high degree but are connected to other nodes with high degrees) in the network, the curves here suggest that positive cases tend to have considerably more nuclei with mid-range influence and fewer nodes with very high or very low influence in comparison to negative patches where all nodes maintain the similar level of influence in the network (flatter distribution curve).

In order to compare the distributions quantitatively and highlight their differences, we perform the student's t-test on average positive and negative distributions for each SNA measure and each prediction task. Each t-test was conducted independently for a specific SNA measure and prediction task combination, ensuring that each statistical evaluation was unique and not influenced by others. The p-values of all the experiments are also reported in Fig. 8. As it can be seen, for most of the distributions p-value is significant (<0.05), suggesting that the differences in the average positive and negative distributions of SNA measures are statistically significant. However, this does not hold for some SNA measures in some prediction tasks, for example, Degree and Katz centralities in HM prediction (while closeness centrality and clustering coefficient are significantly different).

Another interesting take from Fig. 8 is that although the general shape of distributions of positive and negative samples for each SNA measure does not change much over different prediction tasks, there are still some differences that vary from task to task. Take Katz centrality or Node Degree for example, in HM prediction the distributions for positive and negative samples are very close whereas for MSI prediction one can see considerable differences. This implies that not all cell relationship properties that are extracted using SNA measures are predictive for all the tasks. In other words, different SNA measures quantify different aspects of cell relationships and each can be beneficial to a certain task by providing the relevant features, while not being very beneficial to other prediction tasks. Therefore, we suggest extracting all the SNA measures in the cell relationship feature extraction step to avoid missing valuable information for a specific prediction task.

Moreover, we extract the basic statistics of different SNA measures of high-scored patches for each prediction task and visualize them in the form of radar plots in Fig. 9. For each SNA measure, the minimum (min), maximum (max), average (avg), standard deviation (std), and median (mdn) of the values are extracted, normalized, and averaged over positive and negative samples to draw orange and blue plots in Fig. 9. It is easy to note that for all prediction tasks and SNA measures, there are evident differences between statistical SNA features of positive and negative samples, suggesting the relevance and differentiability of the proposed features. Furthermore, these features are more interpretable as singular values. For example in CIN, comparing the statistics of closeness centrality (max(CC), min(CC), avg(CC), mdn(CC)) values for positive samples with those for negative samples shows a significant difference between the two groups. This means that the CIN-positive samples can take higher values of closeness centralities than negative samples which implies that in CIN-positive samples cells tend to live in closer communities (in line with what we have observed with SNA distributions in Fig. 8). On the other hand, the same features for HM-positive and negative tumours are not much different and all around the mid-level, proposing that cells in hypermutated tumours (HM-positive) and wild-type tumours (HM-negative) are distributed similarly close. An in-depth list of the main SNA features and their interpretation in cell networks can be found in Table S3.

These kinds of observations and interpretations can be extracted from all of the plots illustrated in Fig. 9. In general, it seems that for CIN, *TP53*, *BRAF*, and MSI the average normalized values for most of the statistics of the negative class are higher than those of the positive class. Also. the trend of differences between positive and negative classes in statistical features is similar (with minor variations) for all prediction experiments except for HM prediction.

5. Discussion

Predicting tumour mutation or molecular alteration status in most cancers requires either IHC staining or genome sequencing, both of which are expensive to operate on and might not be possible if there is no access to cancer tissue. Various studies in recent years tried to predict these entities from imaging information (WSIs) and artificial intelligence. Mostly, MIL-based methods are utilized to leverage the accessible weak labels of the slides which are usually acquired using the aforementioned tools. As it is often not clear what is contributing to the molecular alteration status of the tumour or its type in the H&E stained sample, it is desired that MIL-based methods provide some explainability on the prediction maps. Furthermore, raw pixel-based inputs to DL models may not provide them with enough abstract information to extract relevant features from WSIs for predicting such hard entities.

To mitigate these challenges, we propose to enhance the SOTA deep learning-based methods for molecular pathways and key mutation prediction by incorporating cell relationship information. In particular, we used the iterative draw-and-ranking sampling strategy to train a deep neural network classifier that accepts in the input both image-based and cell-relationship-based features (Social-IDaRS) as well as an attention-based MIL method that accepts a similar combination of features (Social-CLAM). The cell relationship information is collected using a series of measures from the SNA domain which are then abstracted into 90 statistical features (for each image) that can represent cell-to-cell relationships within the image.

By extensive cross-validation experiments on the TCGA-CRC dataset, we show that Social-IDaRS can improve the AUROC and AUPRC of CIN, HM, *TP53*, *BRAF*, and MSI prediction tasks by 2.4% and 8.8% (on average) over the best-performing model in the literature. By performing similar experiments using Social-CLAM, we show that SNA features carrying cell relationship information can improve the average AUROC and AUPRC by 4% and 7% over all tasks in comparison to CLAM (Lu et al., 2021b) baseline. The outstanding performance of the proposed Social-IDaRS method on the external PAIP dataset achieving AUROC and AUPRC of 0.99 and 0.98, respectively, is evidence of the generalizability of Social-IDaRS.

While deep learning architectures like ResNet50, employed in the original IDaRS and CLAM methods, are proficient at autonomously extracting features from images, their reliance on traditional convolutional layers often restricts them to localized pixel patterns, potentially missing crucial long-range cell relationships pivotal for understanding molecular pathways in CRC. Addressing this limitation, we integrated SNA measures, offering a global and abstract depiction of these cell relationships, enhancing both model accuracy and interpretability. This methodological adjustment, rooted in cellular social relationships, underlines the discriminative power of the proposed SNA features, significantly boosting the performance of deep learning models. What sets these SNA features apart is their direct extractability from an image, contrasting other methods that resort to auxiliary inputs from different modalities (Chen et al., 2020, 2022). Hence, not only do the SNA-based features offer a more holistic view of cellular dynamics and bridge the gap between raw pixels and molecular intricacies, but they also present an almost cost-free and consistently available advantage over more expensive modalities that often lack comprehensive data.

We show that using SNA-based features can add interpretability to some of the advanced predictions that DL models generate. By plotting the distribution of SNA measures (Fig. 8) and their basic statistical features (Fig. 9), we visualize how different social cell relationships can be associated with different prediction tasks. Based on our SNA-based feature averaged on most attentive regions of the Social-IDaRS model, we realized that in CIN+, *TP53*+, *BRAF*+, and MSI+ samples there are more connections (relationships) between nuclei and they are closer to each other (in terms of network connections, see Section 3.4) in comparison to negative samples. On the other hand, all the statistics in eigenvector and Katz centralities were higher for CIN-, *TP53*-, *BRAF*-, and MSI- in comparison to their positive counterparts, suggesting that in the negative samples, there are more influential nuclei although the cell connections are of higher degrees in positive samples (based on node degree, degree centrality, and clustering coefficient statistics). Another interesting observation is the remarkable similarity of radar plots for MSI and *TP53*. This similarity implies that, based on extracted SNA features, *general trends* of cell relationships between negative and positive cases are almost identical for MSI and *TP53* mutation. However, we emphasize the "general trends of cell relationships" because radar plots in Fig. 9 are based on normalized values and real values of related features are different for MSI and *TP53*. We believe that these kinds of insights can be helpful in designing explainable models or proposing a clinical algorithm for such prediction tasks and therefore require more attention in the future.

To help with the explainability of the Social-IDaRS model, we have also provided a visualization tool, called TIAviz, to overlay prediction heatmaps for multiple example WSIs. The CIN, HM, *TP53*, *BRAF*, and MSI prediction heatmaps in TIAviz are generated similarly to Fig. 6 to highlight the most and least attentive regions of the model. The interactive viewer and the option of switching between different heatmap overlays can help identify the regions contributing to the prediction of different molecular pathways or mutations.[2]

In the proposed method, the extracted SNA features to describe the tumour micro-environment are based solely on nuclei relationships. Although they have proven to be beneficial in Social-IDaRS and Social-CLAM models for various tasks (see results in Section 4.3), they still do not include any information regarding the type of nuclei. We think that the integration of nuclei types in the proposed social cell network pipeline and the introduction of new SNA features that allow the incorporation of that information can improve the performance of prediction models on all tasks. Furthermore, including nuclei types can add another layer of explainability and interpretability to the method, for example, we can investigate the composition of influential cells in the cell network to gain more insights into the cancer microenvironment. However, realizing this idea is challenging as it requires accurate prediction of cell types and finding the optimal way of embedding that information within cell networks. Moreover, having nuclei types, it may be possible to generate plausible WSI-level cell networks and investigate the effect of their properties on sophisticated prediction tasks such as mutation or molecular alteration prediction.

Another limitation of the current study, and other similar research studies at the moment, is the lack of publicly available external data for generalization tests. It has been shown that DL models have the capacity to generalize better to unseen data if they are trained with larger and more versatile data. However, curating such data is time-consuming and expensive. While enough training and validation data is not available, models for predicting such sophisticated tasks should be as explainable as possible to eliminate the chance of overfitting.

6. Conclusions

We proposed a novel approach to predict molecular pathways and key mutations in colorectal cancer by incorporating cell-to-cell relationship information as auxiliary inputs in a weakly-supervised deep learning framework. By leveraging SNA measures, we were able to extract useful information from cell-to-cell connections and improve the performance of two SOTA models, introducing Social-CLAM and Social-IDaRS, for multiple prediction tasks including CIN, HM, *TP53*, *BRAF*, and MSI. Our approach outperformed existing SOTA methods and demonstrated the importance of cell relationships and cellular communities in the tumour microenvironment. Additionally, we provided

[2] Our demo is available at https://tiademos.dcs.warwick.ac.uk/bokeh_app?demo=sna.

interpretability to our predictions by visualizing the distribution of SNA measures and their basic statistical features. The proposed method is interpretable and generalizable, making it a valuable tool for precision medicine in colorectal cancer. This study also highlights the potential of incorporating SNA in deep learning models for histology image analysis and the need for further research in this area.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: MOSTAFA JAHANIFAR and NASIR RAJPOOT report financial support provided by the BigPicture project, which is funded by the European Commission, outside of the submitted work. NASIR RAJPOOT reports financial support provided by UK Research and Innovation (UKRI), outside of the submitted work. NASIR RAJPOOT reports financial support from GlaxoSmithKline, UK, outside of the submitted work. NASIR RAJPOOT reports a relationship with Histofy Ltd that includes: board membership. MOSTAFA JAHANIFAR is consultant for Histofy Ltd.

Data availability

The data is publicly available but the code is under university IP and cannot be shared.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2023.103071.

References

Ahmedt-Aristizabal, D., Armin, M.A., Denman, S., Fookes, C., Petersson, L., 2021. A survey on graph-based deep learning for computational histopathology. Comput. Med. Imaging Graph. 102027.

Al-Sohaily, S., Biankin, A., Leong, R., Kohonen-Corish, M., Warusavitarne, J., 2012. Molecular pathways in colorectal cancer. J. Gastroenterol. Hepatol. 27 (9), 1423–1431.

Alemi Koohbanani, N., Jahanifar, M., Gooya, A., Rajpoot, N., 2019. Nuclear instance segmentation using a proposal-free spatially aware deep learning framework. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 622–630.

Alemi Koohbanani, N., Jahanifar, M., Tajadin, N.Z., Rajpoot, N., 2020. NuClick: a deep learning framework for interactive segmentation of microscopic images. Med. Image Anal. 65, 101771.

Bilal, M., Nimir, M., Snead, D., Taylor, G.S., Rajpoot, N., 2022a. Role of AI and digital pathology for colorectal immuno-oncology. Br. J. Cancer 1–9.

Bilal, M., Raza, S.E.A., Azam, A., Graham, S., Ilyas, M., Cree, I.A., Snead, D., Minhas, F., Rajpoot, N.M., 2021. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. Lancet Digit. Health 3 (12), e763–e772.

Bilal, M., Tsang, Y.W., Ali, M., Graham, S., Hero, E., Wahab, N., Dodd, K., Sahota, H., Lu, W., Jahanifar, M., et al., 2022b. AI based pre-screening of large bowel cancer via weakly supervised learning of colorectal biopsy histology images. medRxiv.

Boehm, K.M., Khosravi, P., Vanguri, R., Gao, J., Shah, S.P., 2022. Harnessing multimodal data integration to advance precision oncology. Nat. Rev. Cancer 22 (2), 114–126.

Boland, C.R., Goel, A., 2010. Microsatellite instability in colorectal cancer. Gastroenterology 138 (6), 2073–2087.

Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Med. 25 (8), 1301–1309.

Cao, R., Yang, F., Ma, S.-C., Liu, L., Zhao, Y., Li, Y., Wu, D.-H., Wang, T., Lu, W.-J., Cai, W.-J., 2020. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer. Theranostics 10, 11080.

Carbonneau, M.-A., Cheplygina, V., Granger, E., Gagnon, G., 2018. Multiple instance learning: A survey of problem characteristics and applications. Pattern Recognit. 77, 329–353.

Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., Mahmood, F., 2020. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Trans. Med. Imaging.

Chen, R.J., Lu, M.Y., Williamson, D.F., Chen, T.Y., Lipkova, J., Noor, Z., Shaban, M., Shady, M., Williams, M., Joo, B., et al., 2022. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell 40 (8), 865–878.

Chikontwe, P., Nam, S.J., Go, H., Kim, M., Sung, H.J., Park, S.H., 2022. Feature recalibration based multiple instance learning for whole slide image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 420–430.

Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., Tsirigos, A., 2018. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature Med. 24 (10), 1559–1567.

Couture, H.D., 2022. Deep learning-based prediction of molecular tumor biomarkers from H&E: A practical review. J. Pers. Med. 12 (12), 2022.

Echle, A., Grabsch, H.I., Quirke, P., van den Brandt, P.A., West, N.P., Hutchins, G.G., Heij, L.R., Tan, X., Richman, S.D., Krause, J., et al., 2020. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. Gastroenterology 159 (4), 1406–1416.

Echle, A., Laleh, N.G., Schrammen, P.L., West, N.P., Trautwein, C., Brinker, T.J., Gruber, S.B., Buelow, R.D., Boor, P., Grabsch, H.I., 2021. Deep learning for the detection of microsatellite instability from histology images in colorectal cancer: a systematic literature review. ImmunoInformatics 100008.

El-Deiry, W.S., Goldberg, R.M., Lenz, H.-J., Shields, A.F., Gibney, G.T., Tan, A.R., Brown, J., Eisenberg, B., Heath, E.I., Phuphanich, S., et al., 2019. The current state of molecular testing in the treatment of patients with solid tumors, 2019. CA: Cancer J. Clin. 69 (4), 305–343.

FDA, et al., 2020. FDA Approves First-Line Immunotherapy for Patients with MSI-H/dMMR Metastatic Colorectal Cancer. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-first-line-immunotherapy-patients-msi-hdmmr-metastatic-colorectal-cancer.

Ghaffari Laleh, N., Muti, H.S., Loeffler, C.M.L., Echle, A., Saldanha, O.L., Mahmood, F., Lu, M.Y., Trautwein, C., Langer, R., Dislich, B., Buelow, R.D., Grabsch, H.I., Brenner, H., Chang-Claude, J., Alwers, E., Brinker, T.J., Khader, F., Truhn, D., Gaisa, N.T., Boor, P., Hoffmeister, M., Schulz, V., Kather, J.N., 2022. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. Med. Image Anal. 79, 102474. http://dx.doi.org/10.1016/j.media.2022.102474, URL: https://www.sciencedirect.com/science/article/pii/S1361841522001219.

Graham, S., Jahanifar, M., Azam, A., Nimir, M., Tsang, Y.-W., Dodd, K., Hero, E., Sahota, H., Tank, A., Benes, K., et al., 2021a. Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 684–693.

Graham, S., Jahanifar, M., Vu, Q.D., Hadjigeorghiou, G., Leech, T., Snead, D., Raza, S.E.A., Minhas, F., Rajpoot, N., 2021b. Conic: Colon nuclei identification and counting challenge 2022. arXiv preprint arXiv:2111.14485.

Graham, S., Minhas, F., Bilal, M., Ali, M., Tsang, Y.W., Eastwood, M., Wahab, N., Jahanifar, M., Hero, E., Dodd, K., et al., 2023. Screening of normal endoscopic large bowel biopsies with artificial intelligence: a retrospective study. Gut 72 (9), 1709–1721. http://dx.doi.org/10.1136/gutjnl-2023-329512, URL: https://gut.bmj.com/content/72/9/1709.

Graham, S., Vu, Q.D., Jahanifar, M., Raza, S.E.A., Minhas, F., Snead, D., Rajpoot, N., 2022b. One model is all you need: multi-task learning enables simultaneous histology image segmentation and classification. Med. Image Anal. 102685.

Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. 58, 101563. http://dx.doi.org/10.1016/j.media.2019.101563, URL: https://www.sciencedirect.com/science/article/pii/S1361841519301045.

Hagen, L., Keller, T., Neely, S., DePaula, N., Robert-Cooperman, C., 2018. Crisis communications in the age of social media: A network analysis of Zika-related tweets. Soc. Sci. Comput. Rev. 36 (5), 523–541.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hong, R., Fenyö, D., 2022. Deep learning and its applications in computational pathology. BioMedInformatics 2 (1), 159–168.

Hu, Y., Sirinukunwattana, K., Gaitskell, K., Wood, R., Verrill, C., Rittscher, J., 2022. Predicting molecular traits from tissue morphology through self-interactive multi-instance learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 130–139.

Jahanifar, M., Tajeddin, N.Z., Koohbanani, N.A., Rajpoot, N.M., 2021. Robust interactive semantic segmentation of pathology images with minimal user input. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 674–683.

Janowczyk, A., Madabhushi, A., 2016. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J. Pathol. Inform. 7 (1), 29.

Jaume, G., Pati, P., Bozorgtabar, B., Foncubierta, A., Anniciello, A.M., Feroce, F., Rau, T., Thiran, J.-P., Gabrani, M., Goksel, O., 2021. Quantifying explainers of graph neural networks in computational pathology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8106–8116.

Jiménez-Sánchez, D., Ariz, M., Chang, H., Matias-Guiu, X., De Andrea, C.E., Ortiz-de Solórzano, C., 2022. NaroNet: Discovery of tumor microenvironment elements from highly multiplexed images. Med. Image Anal. 78, 102384.

Kather, J.N., Heij, L.R., Grabsch, H.I., Loeffler, C., Echle, A., Muti, H.S., Krause, J., Niehues, J.M., Sommer, K.A., Bankhead, P., et al., 2020. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat. Cancer 1 (8), 789–799.

Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.-A., Gaiser, T., Marx, A., Valous, N.A., Ferber, D., et al., 2019a. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS Med. 16 (1), e1002730.

Kather, J.N., Pearson, A.T., Halama, N., Jäger, D., Krause, J., Loosen, S.H., Marx, A., Boor, P., Tacke, F., Neumann, U.P., et al., 2019b. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nature Med. 25 (7), 1054–1056.

Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrika 18 (1), 39–43.

Krause, J., Croft, D.P., James, R., 2007. Social network theory in the behavioural sciences: potential applications. Behav. Ecol. Sociobiol. 62 (1), 15–27.

Laleh, N.G., Muti, H.S., Loeffler, C.M.L., Echle, A., Saldanha, O.L., Mahmood, F., Lu, M.Y., Trautwein, C., Langer, R., Dislich, B., et al., 2021. Benchmarking artificial intelligence methods for end-to-end computational pathology. bioRxiv.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature 521 (7553), 436–444.

Li, J., Zheng, Y., Wu, K., Shi, J., Xie, F., Jiang, Z., 2022. Lesion-aware contrastive representation learning for histopathology whole slide images analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 273–282.

Liu, Y., Sethi, N.S., Hinoue, T., Schneider, B.G., Cherniack, A.D., Sanchez-Vega, F., Seoane, J.A., Farshidfar, F., Bowlby, R., Islam, M., et al., 2018. Comparative molecular analysis of gastrointestinal adenocarcinomas. Cancer Cell 33 (4), 721–735.

Lu, M.Y., Chen, T.Y., Williamson, D.F., Zhao, M., Shady, M., Lipkova, J., Mahmood, F., 2021a. AI-based pathology predicts origins for cancers of unknown primary. Nature 594 (7861), 106–110.

Lu, W., Toss, M., Dawood, M., Rakha, E., Rajpoot, N., Minhas, F., 2022. SlideGraph+: Whole slide image level graphs to predict HER2 status in breast cancer. Med. Image Anal. 102486.

Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F., 2021b. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat. Biomed. Eng. 5 (6), 555–570.

Luchini, C., Bibeau, F., Ligtenberg, M., Singh, N., Nottegar, A., Bosse, T., Miller, R., Riaz, N., Douillard, J.-Y., Andre, F., et al., 2019. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. Ann. Oncol. 30 (8), 1232–1243.

Lv, Z., Yan, R., Lin, Y., Wang, Y., Zhang, F., 2022. Joint region-attention and multi-scale transformer for microsatellite instability detection from whole slide images in gastrointestinal cancer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 293–302.

Madabhushi, A., Lee, G., 2016. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med. Image Anal. 33, 170–175.

Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Velázquez Vega, J.E., Brat, D.J., Cooper, L.A., 2018. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc. Natl. Acad. Sci. 115 (13), E2970–E2979.

Newman, M., 2018. Networks. Oxford University Press.

Oliveira, S.P., Neto, P.C., Fraga, J., Montezuma, D., Monteiro, A., Monteiro, J., Ribeiro, L., Gonçalves, S., Pinto, I.M., Cardoso, J.S., 2021. CAD systems for colorectal cancer from WSI are still not ready for clinical acceptance. Sci. Rep. 11 (1), 1–15.

Pocock, J., Graham, S., Vu, Q.D., Jahanifar, M., Deshpande, S., Hadjigeorghiou, G., Shephard, A., Bashir, R.M.S., Bilal, M., Lu, W., et al., 2022. TIAToolbox as an end-to-end library for advanced tissue image analytics. Commun. Med. 2 (1), 1–14.

Pritchard, C.C., Grady, W.M., 2011. Colorectal cancer molecular biology moves into clinical practice. Gut 60 (1), 116–129.

Reisenbüchler, D., Wagner, S.J., Boxberg, M., Peng, T., 2022. Local attention graph-based transformer for multi-target genetic alteration prediction. arXiv preprint arXiv:2205.06672.

Schirris, Y., Gavves, E., Nederlof, I., Horlings, H.M., Teuwen, J., 2022. DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. Med. Image Anal. 79, 102464.

Scott, J., 1988. Social network analysis. Sociology 22 (1), 109–127.

Shaban, M., Awan, R., Fraz, M.M., Azam, A., Tsang, Y.-W., Snead, D., Rajpoot, N.M., 2020. Context-aware convolutional neural network for grading of colorectal cancer histology images. IEEE Trans. Med. Imaging 39 (7), 2395–2405.

Shephard, A.J., Graham, S., Bashir, S., Jahanifar, M., Mahmood, H., Khurram, A., Rajpoot, N.M., 2021. Simultaneous nuclear instance and layer segmentation in oral epithelial dysplasia. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 552–561.

Singh, M.P., Rai, S., Pandey, A., Singh, N.K., Srivastava, S., 2021. Molecular subtypes of colorectal cancer: An emerging therapeutic opportunity for personalized medicine. Genes Dis. 8 (2), 133–145.

Srinidhi, C.L., Ciga, O., Martel, A.L., 2021. Deep neural network models for computational histopathology: A survey. Med. Image Anal. 67, 101813.

Teichmann, M., Aichert, A., Bohnenberger, H., Ströbel, P., Heimann, T., 2022. End-to-end learning for image-based detection of molecular alterations in digital pathology. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II. pp. 88–98.

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N., 2016. Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans. Med. Imaging 35 (8), 1962–1971.

Vuong, T.T.L., Vu, Q.D., Jahanifar, M., Graham, S., Kwak, J.T., Rajpoot, N., 2022. IMPaSh: A novel domain-shift resistant representation for colorectal cancer tissue classification. arXiv preprint arXiv:2208.11052.

Wagner, S.J., Reisenbüchler, D., West, N.P., Niehues, J.M., Zhu, J., Foersch, S., Veldhuizen, G.P., Quirke, P., Grabsch, H.I., van den Brandt, P.A., et al., 2023. Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study. Cancer Cell 41 (9), 1650–1661.

Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., 2013. The cancer genome atlas pan-cancer analysis project. Nature Genet. 45 (10), 1113–1120.

Xi, Y., Xu, P., 2021. Global colorectal cancer burden in 2020 and projections to 2040. Transl. Oncol. 14 (10), 101174.

Yamashita, R., Long, J., Longacre, T., Peng, L., Berry, G., Martin, B., Higgins, J., Rubin, D.L., Shen, J., 2021. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. Lancet Oncol. 22, 132–141.

Zamanitajeddin, N., Jahanifar, M., Rajpoot, N., 2021. Cells are actors: Social network analysis with classical ml for sota histology image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 288–298.

Zhao, Y., Lin, Z., Sun, K., Zhang, Y., Huang, J., Wang, L., Yao, J., 2022. SETMIL: spatial encoding transformer-based multiple instance learning for pathological image analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 66–76.

Zhou, Y., Graham, S., Alemi Koohbanani, N., Shaban, M., Heng, P.-A., Rajpoot, N., 2019. Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.