
Neural Graph Modelling of Whole Slide Images for Survival Analysis

Anonymous Author(s)

Anonymous Affiliation

Anonymous Email

Abstract

1
2 Evaluation of a cancer patient’s prognostic outlook is an essential step in the clinical
3 decision-making process, involving the assessment of complex tissue structures in
4 multi-gigapixel whole slide images (WSIs). Effective risk stratification of patients
5 from WSIs has proven challenging despite several approaches across the literature
6 due to their large size and inability of existing approaches to effectively model
7 inter-relationships between different tissue components. We propose a graph neural
8 network (GNN) model that performs pairwise ranking of graph representations
9 of WSIs based on survival scores. The proposed approach translates spatially-
10 localised deep features along with their spatial context to a graph neural network
11 to produce survival scores. Analysis over breast cancer patients from The Cancer
12 Genome Atlas (TCGA) shows that the proposed GNN approach is able to rank
13 patients with respect to their disease-specific survival times with a concordance
14 index of 0.672 ± 0.058 . This is a significant improvement over existing state of
15 the art and paves the way for neural graph modelling of WSI data for survival
16 prediction for other cancer types.

1 Introduction

17
18 Breast cancer (BCa) most commonly affects women, with around 55,500 women per year diagnosed
19 in the UK and 1 in 7 women will develop breast cancer in their lifetime [1]. As of 2022 breast cancer
20 is the most commonly diagnosed cancer in the UK and is the second most common form of cancer
21 related deaths in women. According to Public Health England, 85% of women diagnosed with breast
22 cancer in England survive their disease for five years or more [2], with survival rates doubling over
23 the last 40 years driven by more thorough early cancer detection and improved treatment regimes.
24 Mortality rates are projected to fall by a further 26% before 2035.

25 In the process of an individual’s breast cancer diagnosis, a tissue sample is typically taken from
26 the tumour region and analysed by a histopathologist. The sample is stained using Haematoxylin
27 and Eosin (H&E) and then digitised using a high resolution whole slide imaging scanner resulting
28 in a multi-gigapixel ($100,000 \times 100,000$ pixels at 0.25 microns per pixel resolution) whole slide
29 image (WSI). Pathologist assessment of the WSI involves inspection of the morphological features
30 of the tissue cells to draw conclusions relating to the grade, type, and hormone receptor status of
31 the patient’s cancer. All of these features have strong correlations to a given patient’s likelihood of
32 survival and is used to determine treatment options.

33 Currently the analysis of WSIs is performed by human histopathologists and there is a significant
34 possibility for human error and bias in making decisions. Most notable in this regard is the assessment
35 of nuclear pleomorphism, a component step of the Nottingham grading system. This assessment is
36 often the most subjective and so pathologists differ markedly in their nuclear grading. It has been
37 found that breast specialists will assign higher grades than non-specialists [3] based on the assessment
38 of the tissue nuclear pleomorphism. This degree of subjectivity is not conducive to reproducible and
39 reliable categorisations. There is, therefore, significant motivation to create an objective approach
40 method for survival prediction that avoids human subjectivity by stratifying patients based on their
41 risk.

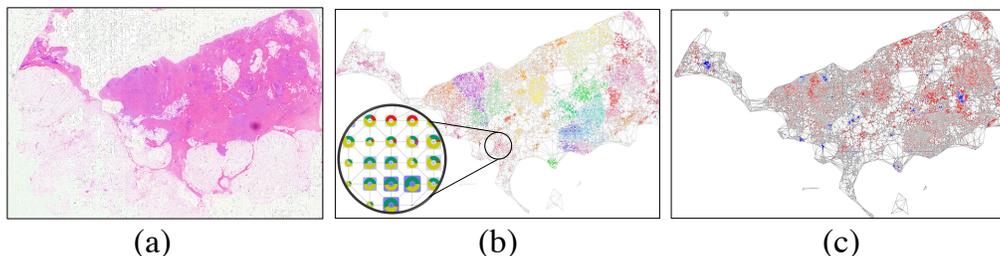


Figure 1: Major steps in the proposed approach: (a) Input Whole Slide Image. (b) Patch level feature extraction and graph construction by spatial clustering and inter-linking neighbouring patches. (c) Using Graph Neural Network based node level survival scoring to generate both node level and WSI level predictions.

42 Weakly supervised learning from whole slide images has seen a number of contributions recently,
 43 particularly aiming to solve classification tasks. However, the regression based problem of survival
 44 prediction is considerably more difficult as disease risk presents its self as a combination of a number
 45 of histological features which correspond to disease status. While there has been some work done
 46 aiming to predict survival from local information extracted from sub-patches of the WSI, there are
 47 only a few frameworks that perform prediction at a global WSI level. Existing methods [4, 5] follow
 48 a structure of taking a large number of random patches from the image and clustering them according
 49 to a rule set, commonly by phenotype. These clusters at patch level will undergo aggregation and
 50 produce a prediction. Alternative methods include MCAT, a co-attention mapping between WSIs
 51 and genomic sequencing formulated in an embedding space [6]. A number of studies have also
 52 investigated model performance when patient data is also fed in however the goal of this work is
 53 to construct a model capable of WSI level survival prediction alone. Furthermore, there have been
 54 successful integrations when using a NTL (Nucleus, Tumour, and Lymphocyte) data along side the
 55 RGB WSI [7]. However, the concordance between predicted and true survival using whole slide
 56 image data remains low.

57 A considerable limitation to the existing systems in the literature is the loss of spatial context between
 58 the random patches. Macro-scale histological structures formed from specific cell types contribute
 59 heavily to a prognostic prediction for a patient. Furthermore, existing implementations have also
 60 relied on external information (such as gene expression patterns) from the WSI to boost scores. The
 61 motivation for this work is to learn survival from the WSI alone.

62 In order to identify significant survival associated features in multi-gigapixel WSIs, we need an
 63 effective way of capturing the inter-relationship between different components in the WSI. For this
 64 purpose, graph based modelling of whole slide images is an attractive solution. Existing work in
 65 this domain (SlideGraph and SlideGraph+) [8, 9] has shown that it is possible to solve classification
 66 problems for breast cancer receptor status prediction using WSI-level graph neural networks. In
 67 this work, we demonstrate the effectiveness of graph based neural models of whole slide images
 68 for survival analysis by using a pairwise ranking loss over predicted survival scores. The proposed
 69 approach results in significant improvement over existing state of the art approaches in this domain
 70 and paves the way for more effective graph based models.

71 2 Methodology

72 2.1 Whole slide image and survival data

73 We collected 1133 whole slide images of Formalin-Fixed paraffin-Embedded (FFPE) Hematoxylin
 74 and Eosin (H&E) stained tissue section of 1084 breast cancer patients from The Cancer Genome
 75 Atlas (TCGA) [10, 11]. The Disease Specific Survival (DSS) data for these patients were collected
 76 from the TCGA Pan-Cancer Clinical Data Resource (TCGA-CDR) [12]. DSS is the length of time
 77 between the data being taken and a disease specific event occurring, in this case death. For some of
 78 the patients the survival data was missing and, consequently, was not used in the study. In line with
 79 clinical practice, patient survival times were censored at 10 years.

80 2.2 Pre-processing

81 Quality of a digitised H&E stained tissue section can be adversely affected by tissue processing
 82 artefacts such as tissue folds and pen-marking originating from histology laboratory. We filter these
 83 artefacts by segmenting the tissue region of whole slide image and ignoring regions with tissue
 84 artefacts (tissue folding, and pen-marking etc.) using our in-house tissue segmentation model. As
 85 an entire WSI at full resolution can be of size $100,000 \times 150,000$ pixels and can be challenging
 86 to fit into a GPU memory, therefore we tile each WSI into patches of size 512×512 pixels at a
 87 spatial resolution of 0.25 microns-per-pixel (MPP). Patches with less than 40% of tissue region (mean
 88 pixel intensity of 60% pixels being higher than 200) are discarded and the rest of patches (tumor and
 89 non-tumor) are used in the study. The total number of 512×512 patches in the data is 8, 487, 768.

90 2.3 Graph Representation of a Whole Slide Image

91 For each WSI, we construct its graph representation, $G_i = G(\mathbf{X}_i)$ where \mathbf{X}_i is the set of features
 92 extracted from each patch within the WSI. In general, a WSI is represented as a set $X = \{\mathbf{p}_m | m =$
 93 $1, \dots, M\}$ where $\mathbf{p}_m \equiv (\mathbf{x}_m, \mathbf{f}_m)$ is composed of the location of the given patch, \mathbf{x}_m , along with its
 94 feature vector, \mathbf{f}_m . The graph construction process can be broken down into three steps which are
 95 explained below (see Figure-1):

96 2.3.1 Patch-Level Feature Extraction

97 The tissue region in the whole slide image is broken down into patches and representative features
 98 of each patch are then extracted. Specifically, we extract Shuffle-Net [13] based deep features. We
 99 encode the patch image of size 512×512 pixels into a 1024-dimensional feature vector by extracting
 100 latent representation of the penultimate fully-connected layer in Shuffle-Net. However, other types of
 101 feature representations can be used in the proposed framework as well such as patch level cellular
 102 composition [14].

103 2.3.2 Spatial Clustering

104 Due to the large size of a whole slide image and the large number of tissue patches, it is necessary to
 105 reduce the size of the graph while maintaining as much of the stored information as possible. This is
 106 done to reduce the computational cost of learning and subsequent analysis with the graphs. In line
 107 with the SlideGraph approach, agglomerative clustering [15] is used to group patches in the original
 108 set P into K clusters represented by the set $C = \{c_k | k = 1, \dots, K\}$ based on spatial neighbourhood
 109 and feature similarity. The number of clusters is different for each WSI depending upon its size and
 110 tissue heterogeneity. For further details, the interested reader is referred to [8, 9].

111 2.3.3 Graph Construction

112 For each WSI, each cluster in its cluster set C is considered to be a vertex in the vertex set V of
 113 its graph representation $G = (V, E)$ [16, 17]. The geometric centre of the node is obtained as:
 114 $\mathbf{g}_c = \frac{1}{|c|} \sum_{\mathbf{p}_j \in c} \mathbf{p}_j$ whereas, the feature representation of each node is taken as the average of patch
 115 features within the cluster and denoted by \mathbf{h}_c . The purpose of using graphs in the proposed approach
 116 is to model the inter-relationship between neighbouring regions in the WSI to capture large-scale
 117 topology. This is done by inter-connecting nodes within a maximum connection distance of 1,500
 118 pixels (between top left corners of patches) to construct the edge set E . This effectively connects all
 119 patches within 1mm in the tissue.

120 2.4 GNN for ranking based on survival times

121 Following the construction of the WSI graph G_i , it can be passed to a graph neural network model
 122 $F(G_i; \theta)$ for prediction of WSI-level survival scores. Here, θ denotes the trainable weight parameters
 123 of the GNN. For training the GNN, we consider the graph representation $G_i, i = 1 \dots N$ of each
 124 WSI in the training set along with the corresponding survival time T_i (in days) and the event indicator
 125 variable $\delta_i \in \{0, 1\}$ representing whether the patient died of breast cancer ($\delta_i = 1$) or not ($\delta_i = 0$).
 126 We use pairwise ranking to train the GNN to predict survival scores based on the constraint that
 127 if the corresponding survival time for patient i is larger than that of patient j and the event for
 128 patient j has taken place, i.e., $T_i > T_j | \delta_j = 1$, then GNN generated survival score for G_i should

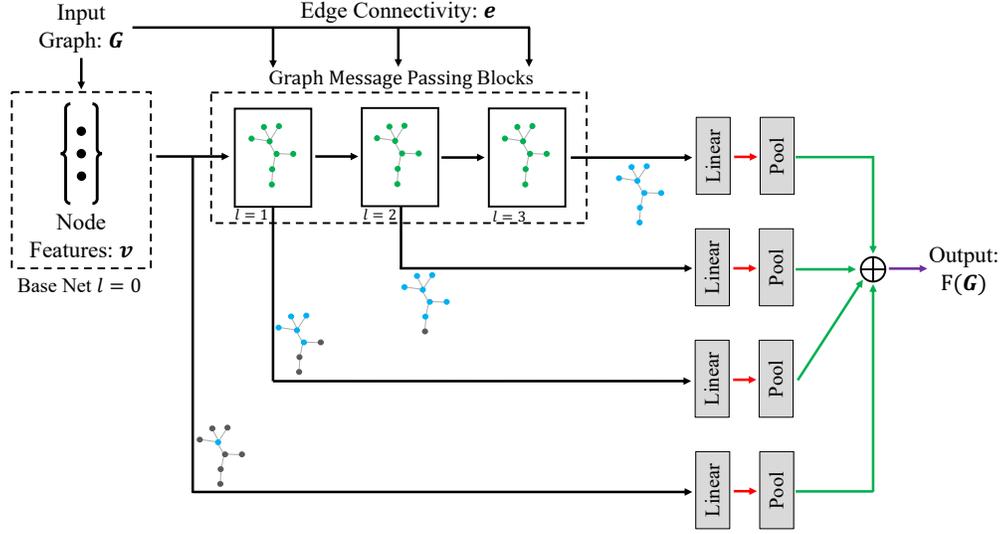


Figure 2: Schematic of graph neural network with message passing block layers using Edge-Conv (equation 2) or GIN-Conv layers (equation 3). The base net is composed of a linear layer, batch normalisation (BN), and gaussian error linear units activation (GELU) layer. Each message passing block includes feature level information from increasing order of neighbourhood nodes, as illustrated by the grey and blue graphs. Black lines indicate latent node representations. Red lines carry node level prediction scores. Green lines are the layer-wise WSI-level outputs which form the final WSI-level output.

129 be larger than that of G_j , i.e., $F(G_i; \theta) > F(G_j; \theta)$ [18]. This results in a pairwise comparison set
 130 $S = \{(i, j) | T_i > T_j, \delta_j = 1, i, j = 1 \dots N\}$ consisting of WSI pairs that can be used for model
 131 training. A pairwise ranking loss is then used for model training with adaptive momentum based
 132 gradient descent as follows:

$$\theta^* = \arg \min_{\theta} \sum_{(i,j) \in S} \max(0, 1 - (F(G_i; \theta) - F(G_j; \theta))). \quad (1)$$

133 The survival prediction framework proposed in this work can utilise different types of graph neural
 134 network architectures. However, we have used the architecture shown in Figure-2. It is constructed
 135 from a base net consisting of a multi-layered perceptron (MLP) that operates on node level feature
 136 representations. The output of the base network is then passed to a series of graph message passing
 137 blocks (GMPBs) which take nodal connectivity in the graph into account. Each time the GMPB
 138 is invoked the architecture accumulates information from increasingly higher-order neighbours.
 139 Additionally, following each GMPB the node features are retained and passed through their own MLPs,
 140 which considers each node's embedding and the difference in embedding with its neighbours. Thus,
 141 for a given neighbourhood, N_k , of the node k the l^{th} GMPB will return the feature embedding, $f_k^{(l)}$,
 142 of a node $\mathbf{p}_k \equiv (\mathbf{x}_k, \mathbf{f}_k) \in X$, which is then passed to a linear layer generating node level predictions
 143 $f_l(\mathbf{p}_k) = \mathbf{w}_l^T \mathbf{f}_k^{(l)}$. The resulting node level prediction scores are then pooled, $\mathbf{F}(G) = \sum_{\mathbf{p} \in P} f_l(\mathbf{p})$
 144 to create layer-wise WSI level scores. These layer wise scores can then be aggregated to produce an
 145 overall WSI level prediction.

146 It is thus the case that graph neural networks (GNN) [19], built using Edge Convolution (Edge-Conv)
 147 or graph isomorphic convolution (GIN-Conv) [20], are suitable for producing a prediction for the
 148 input graph G . The GNN's learning is predicated upon the ability to extract abstract representations
 149 of node level features as a function of their local neighbourhood, by simulating message passing
 150 between neighbouring nodes.

151 Mathematically the Edge-Conv and GIN-Conv layers can be expressed by the following functions,
152 respectively:

$$\mathbf{h}_k^{(l)} = \sum_{u \in \mathcal{N}_k} H_\theta^{(l)}(\mathbf{h}_k^{(l-1)}, \mathbf{h}_u^{(l-1)} - \mathbf{h}_k^{(l-1)}), \quad (2)$$

$$\mathbf{h}_k^{(l)} = H_\theta^{(l)}((1 + \epsilon^{(l)}) \cdot \mathbf{h}_k^{(l-1)} + \sum_{u \in \mathcal{N}_k} \mathbf{h}_u^{(l-1)}). \quad (3)$$

154 Here, $\mathbf{h}_k^{(1)}$ is the feature vector for the k^{th} node on the l^{th} layer (with $\mathbf{h}_k^{(0)} = \mathbf{f}_k$). As shown in
155 equation 2, Edge-Conv works by aggregating node representation of a given node with its difference
156 with the representation of node in its neighborhood $\mathcal{N}_k = \{u \in V | (u, k) \in E\}$. For GIN-Conv, the
157 first term in the right hand side of the equation 3 determines the local contributions from a given
158 node, controlled by ϵ and the second term takes the contributions from the neighbourhood. Finally,
159 $H_\theta^{(l)}$ is the all encapsulating multi-layer perceptron that is capable of learning the non-linear node
160 level transformations that are required. We have experimented with both Edge-Conv and GIN-Conv
161 layers in the paper and found that Edge-Conv offers superior predictive performance.

162 The GCN produces a prediction score from the summation of the node level feature representations.
163 The MLP weights are tuned through the propagation of gradients which result from the ranking loss
164 function discussed earlier.

165 2.5 Code and data availability

166 The proposed approach has been implemented in Python using the PyTorch-Geometric library for
167 graph neural networks. In line with anonymity requirements of the paper, the complete architecture
168 will be released on the organisational Github account for academic use upon acceptance of the paper
169 to allow complete reproducibility. The graph representations, associated survival data and trained
170 predictive models will also be released.

171 3 Experiments and Results

172 In order to evaluate predictive performance of the proposed approach, we have used 5 runs in each
173 of which the dataset was randomly divided into a training and test set with 20% of the overall data
174 for reporting test performance. We ensured that the percentage of cases with events is kept the
175 same across training and test splits. We report the predictive performance of the proposed approach
176 using concordance index (c-index). C-index measures the degree of concordance between relative
177 prediction scores of test patients and their actual survival times. In line with area under the receiver
178 operating characteristic curve (AUROC), the C-index ranges from 0.0 (inverted ranking of survival
179 scores) to 0.5 (no concordance between predicted scores and actual survival times) to 1.0 (perfect
180 concordance between prediction scores and actual survival times). It enables us to compare our
181 predictive performance with previously published results. In addition to this, we have also report the
182 Kaplan-Meier survival curves for high-survival and low-survival group stratifications obtained by
183 thresholding the prediction score generated by the model with a threshold selected using training data
184 examples. The p-value of the log-rank test is also reported.

185 3.1 Quantitative results and Comparison

186 As shown in Table-1 the average concordance index over the test set for the proposed approach is
187 0.672 ± 0.058 which is markedly better in comparison to previous approaches.

188 The Kaplan-Meier curve over a representative test set data split is shown in Figure-3. It shows that the
189 prediction score generated by the proposed approach was able to produce meaningful stratification
190 of patients into two groups whose survival is statistically significantly different ($p = 0.002$ (3 s.f)
191 $\lll 0.05$). The patients in the low group (prediction score below the threshold) have significantly
192 better survival probability over time in comparison to patients in the high risk group. Results of this
193 stratification were consistent over multiple evaluation runs.

194 We believe that the proposed method is able to perform better in comparison to other existing ap-
195 proaches due its effective inclusion of spatial context between the patch level features. As discussed
196 earlier many alternative methods rely on generic grouping of random patches from the WSI. Exami-
197 nation of the global spatial context of micro-features in histology images is a common practice for

198 pathologists and so exclusion of this information limits a given model's ability to perform survival
199 prediction. We have also investigated the impact of various architectural choices which are discussed
200 below.

201 **Choice of Model Layers**

202 We have experimentally compared the predictive performance of GIN vs Edge-Conv which has
203 resulted in a significant improvement in concordance index from 0.602 ± 0.093 (With GIN-Conv) to
204 0.672 ± 0.058 (with Edge-Conv).

205 **Role of Patch Clustering**

206 Our analysis shows that the predictive performance in terms of test concordance index does not
207 change much with (0.672 ± 0.058) and without clustering (0.685 ± 0.072). However, patch-level
208 clustering reduces the size of the graphs leading to faster computation times. Without clustering, the
209 average per train-test split run time is approximately 1800 whereas with clustering it stands at 770
210 seconds. This is, as expected, dependent upon the amount of memory available on the GPU.

211 **Role of Knowledge Jump Connections**

212 We have investigated the role of "knowledge jumping" or connections from each layer to the final
213 prediction experimentally and found that the difference between average test concordance with
214 (0.672 ± 0.058) and without (0.687 ± 0.074) these connections is marginal. The addition of these
215 connections seems to reduce the variance in prediction results.

216 **Choice of node level features**

217 In addition to deep features obtained from Shuffle-Net, we have also used estimates of patch-level
218 cellular composition i.e., counts of neoplastic, inflammatory, connective and epithelial cells in
219 each patch, as node level features. These estimates are obtained from a machine learning method
220 called ALBRT (see [14] for details). Using these features, the best average concordance index was
221 0.63 ± 0.09 which is lower in comparison to deep features from Shuffle-Net. We conjecture that
222 despite lower concordances such clinical features may offer novel insights into the role of different
223 cells in conjunction with other histologically important features such as blood vessels and mitotic
224 figures along with genomic or transcriptomic features.

225 **Effect of edge connectivity threshold**

226 We have also analyzed the impact of edge connectivity threshold used for defining the edges in the
227 graph on test concordance figures. We have found that an optimal test concordance (0.689 ± 0.08) is
228 obtained with a connectivity threshold of 2000 pixels whereas 1500 pixels offers lower variation. The
229 predictive performance started to drop beyond the 1mm tissue region bounds on edge connectivity
230 between nodes. However, we believe that this range is dependent upon the choice of features and can
231 be chosen as a design parameter for different problems.

232 **Impact of censoring**

233 We have used 10 years as the censoring threshold. However, if no such artificial censoring is used the
234 average test concordance index improves to 0.702 ± 0.076 . This is expected due to the increase in
235 the number of pairs of examples that contribute to the loss function in training.

236 **Processing times**

237 The approach has been developed in Python using the PyTorch Geometric library for graph neural
238 networks. On an NVIDIA RTX 3080 GPU, the per train-test split execution time is around 12 minutes
239 (on average). The creation of graphs from node level features for the whole dataset requires (on
240 average) 35 minutes whereas the extraction of deep features and pre-processing can be expected to
241 take 8 to 10 minutes per whole slide image (WSI) for pre-processing, patch extraction and node level
242 feature computation depending upon the size of tissue within the WSI.

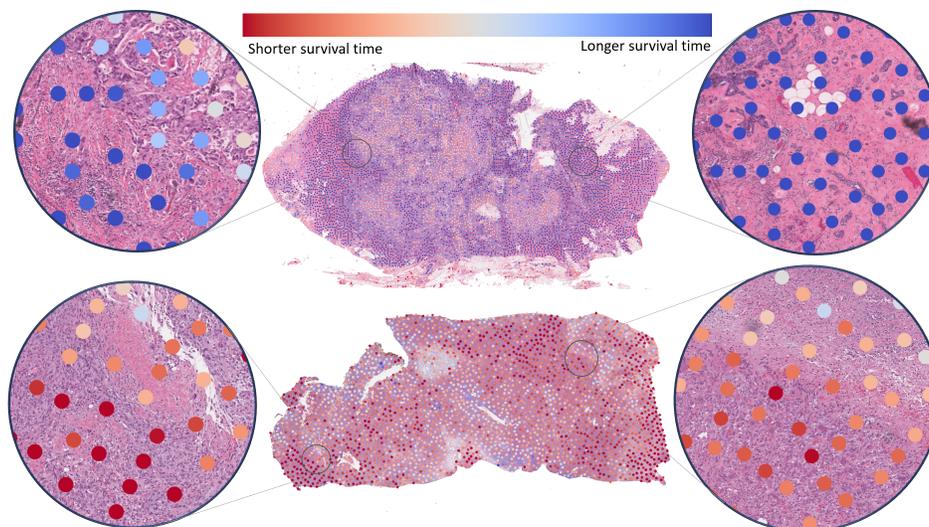


Figure 4: Graph visualisation of TCGA-B6-A0RS ($\delta = 1$, $T = 3063$ days, WSI model score = 4.515) (Top) and TCGA-AC-A2QJ ($\delta = 1$, $T = 446$ days, WSI model score = 1.548) (Bottom) using node level prediction scores with red to blue false colour mapping. Our initial investigation into regions associated with low survival indicates their association to larger number of pleomorphic tumour cells along with increased cellularity.

264 centres for performance assessment. Linking histopathology based features with other genomic and
 265 transcriptomic features can further improve predictive performance.

266 Author Contribution and Acknowledgements

267 These will be added after acceptance due to anonymity requirements for review.

268 References

- 269 [1] Cancer Research UK. Breast cancer statistics. [https://www.cancerresearchuk.org/health-](https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer)
 270 [professional/cancer-statistics/statistics-by-cancer-type/breast-cancer](https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer) Accessed: 24/08/2022.
 271 1
- 272 [2] Public Health England. Cancer survival in england for patients diagnosed between 2014 and
 273 2018, and followed up to 2019. [https://www.gov.uk/government/statistics/cancer-survival-in-](https://www.gov.uk/government/statistics/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-until-2019/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-to-2019)
 274 [england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-until-2019/cancer-](https://www.gov.uk/government/statistics/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-until-2019/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-to-2019)
 275 [survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-to-](https://www.gov.uk/government/statistics/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-until-2019/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-to-2019)
 276 [2019](https://www.gov.uk/government/statistics/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-until-2019/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-to-2019) Accessed: 24/08/2022. 1
- 277 [3] B Dunne and J J Going. Scoring nuclear pleomorphism in breast cancer. *Histopathology*, 39(3):
 278 259–265, September 2001. 1
- 279 [4] Jiawen Yao, Xinliang Zhu, Jitendra Jonnagaddala, Nicholas Hawkins, and Junzhou Huang.
 280 Whole slide images based cancer survival prediction using attention guided deep multiple
 281 instance learning networks. *Medical Image Analysis*, 65:101789, oct 2020. doi: 10.1016/j.
 282 *media*.2020.101789. 2, 7
- 283 [5] Xinliang Zhu, Jiawen Yao, Feiyun Zhu, and Junzhou Huang. Wsisa: Making survival prediction
 284 from whole slide histopathological images. In *Proceedings of the IEEE Conference on Computer*
 285 *Vision and Pattern Recognition (CVPR)*, July 2017. 2
- 286 [6] Richard J. Chen, Ming Y. Lu, Wei-Hung Weng, Tiffany Y. Chen, Drew F.K. Williamson, Trevor
 287 Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival
 288 prediction in gigapixel whole slide images. In *Proceedings of the IEEE/CVF International*
 289 *Conference on Computer Vision (ICCV)*, pages 4015–4025, October 2021. 2, 7

- 290 [7] Huidong Liu and Tahsin Kurc. Deep learning for survival analysis in breast cancer with whole
291 slide image data. *Bioinformatics*, 38(14):3629–3637, 06 2022. ISSN 1367-4803. doi: 10.1093/
292 bioinformatics/btac381. URL <https://doi.org/10.1093/bioinformatics/btac381>. 2
- 293 [8] Wenqi Lu et al. Capturing cellular topology in multi-gigapixel pathology images. In *2020*
294 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*,
295 pages 1049–1058, 2020. doi: 10.1109/CVPRW50498.2020.00138. 2, 3
- 296 [9] Wenqi Lu, Michael Toss, Muhammad Dawood, Emad Rakha, Nasir Rajpoot, and Fayyaz
297 Minhas. Slidegraph+: Whole slide image level graphs to predict her2 status in breast cancer.
298 *Medical Image Analysis*, 80:102486, 2022. ISSN 1361-8415. doi: [https://doi.org/10.1016/](https://doi.org/10.1016/j.media.2022.102486)
299 [j.media.2022.102486](https://doi.org/10.1016/j.media.2022.102486). URL [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S1361841522001335)
300 [S1361841522001335](https://www.sciencedirect.com/science/article/pii/S1361841522001335). 2, 3
- 301 [10] The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast
302 tumours. *Nature*, 490(7418):61–70, Oct 2012. ISSN 1476-4687. doi: 10.1038/nature11412.
303 URL <https://doi.org/10.1038/nature11412>. 2
- 304 [11] Katherine A Hoadley, Christina Yau, Toshinori Hinoue, Denise M Wolf, Alexander J Lazar,
305 Esther Drill, Ronglai Shen, Alison M Taylor, Andrew D Cherniack, Vésteinn Thorsson, et al.
306 Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of
307 cancer. *Cell*, 173(2):291–304, 2018. 2
- 308 [12] Jianfang Liu, Tara Lichtenberg, Katherine A Hoadley, Laila M Poisson, Alexander J Lazar,
309 Andrew D Cherniack, Albert J Kovatich, Christopher C Benz, Douglas A Levine, Adrian V
310 Lee, et al. An integrated tcga pan-cancer clinical data resource to drive high-quality survival
311 outcome analytics. *Cell*, 173(2):400–416, 2018. 2
- 312 [13] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient
313 convolutional neural network for mobile devices. *CoRR*, abs/1707.01083, 2017. URL [http://](http://arxiv.org/abs/1707.01083)
314 arxiv.org/abs/1707.01083. 3
- 315 [14] Muhammad Dawood, Kim Branson, Nasir M. Rajpoot, and Fayyaz ul Amir Afsar Minhas.
316 Albrt: Cellular composition prediction in routine histology images, 2021. URL [https://](https://arxiv.org/abs/2108.08306)
317 arxiv.org/abs/2108.08306. 3, 6
- 318 [15] Daniel Müllner. Modern hierarchical, agglomerative clustering algorithms, 2011. URL [https://](https://arxiv.org/abs/1109.2378)
319 arxiv.org/abs/1109.2378. 3
- 320 [16] David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, and Lars
321 Petersson. A survey on graph-based deep learning for computational histopathology, 2021.
322 URL <https://arxiv.org/abs/2107.00272>. 3
- 323 [17] Harshita Sharma, Norman Zerbe, Sebastian Lohmann, Klaus Kayser, Olaf Hellwich, and Peter
324 Hufnagl. A review of graph-based methods for image analysis in digital histopathology. *The*
325 *Diagnostic Pathology Journal*, 1, 08 2015. doi: 10.17629/www.diagnosticpathology.eu-2015-1:
326 61. 3
- 327 [18] Fayyaz Minhas et al. L1-regularized neural ranking for risk stratification and its application
328 to prediction of time to distant metastasis in luminal node negative chemotherapy naïve breast
329 cancer patients. In *Machine Learning and Principles and Practice of Knowledge Discovery in*
330 *Databases*, pages 390–400, Cham, 2021. Springer International Publishing. ISBN 978-3-030-
331 93733-1. 4
- 332 [19] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional
333 networks, 2016. URL <https://arxiv.org/abs/1609.02907>. 4
- 334 [20] Keyulu Xu et al. How powerful are graph neural networks? *CoRR*, abs/1810.00826, 2018. URL
335 <http://arxiv.org/abs/1810.00826>. 4
- 336 [21] Maximilian Ilse, Jakob Tomczak, and Max Welling. Attention-based deep multiple instance
337 learning. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International*
338 *Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*,
339 pages 2127–2136. PMLR, 10–15 Jul 2018. URL [https://proceedings.mlr.press/v80/](https://proceedings.mlr.press/v80/ilse18a.html)
340 [ilse18a.html](https://proceedings.mlr.press/v80/ilse18a.html). 7