Early Detection of Liver Fibrosis Using Graph Convolutional Networks

Marta Wojciechowska[1,2][0000−0002−9561−8799], Stefano Malacrino[2,3][0000−0003−1237−8572], Natalia Garcia Martin[2][0000−0002−5457−251X], Hamid Fehri[1,2][0000−0003−2017−2670], and Jens Rittscher[1,2,4][0000−0002−8528−8298]

[1] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK
[2] Big Data Institute, University of Oxford, Li Ka Shing Centre for Health Information and Discovery, Oxford, UK
[3] Nuffield Department of Surgical Sciences, University of Oxford
[4] NIHR Oxford Biomedical Research Centre, Oxford, UK
marta.wojciechowska@lincoln.ox.ac.uk

Abstract. Detection of early onset of fibrosis is critical to detecting long term damage to identify potential loss of organ function. While formal grading systems for fibrosis have been established, we argue that a quantitative analysis of fibrosis patterns will improve diagnostic quality and help to standardise clinical reporting. Here we are using deep learning to identify elementary fibrosis patterns. Subsequently, a graphical model is utilised to model the spatial organisation of the fibrosis patterns. Our experimental results demonstrated that this approach correlates well with established clinical grading. The presented method holds the potential to be applied to histology in other organs (e.g. kidney).

Keywords: Digital pathology · Fibrosis · Graph Neural Networks.

1 Introduction

In the context of chronic diseases, fibrosis refers to the deposition of collagen in tissue. Fibrosis can lead to organ dysfunction and even to organ failure. Typically, histochemical stains are being used to visualise collagen which would otherwise be hard to quantify. Assessing the level of fibrosis plays an important role in the diagnosis of liver disease in general and Non-alcoholic Steatohepatitis (NASH) in particular. NASH is a progressive liver disease characterised by inflammation and fibrosis, and has strong links to diabetes, obesity, and cardiovascular diseases. By 2030, the number of NASH patients is expected to increase by 60% and the number of liver deaths related to NASH by 178% [4]. NASH is a histological definition that groups together defects in diverse biochemical processes causing hepatic fat accumulation, inflammation, necrosis and fibrosis.

While the measurement of the Collagen Proportionate Area (CPA) [6] provides a very basic assessment of the collagen content per unit area, it has already been shown that the assessment of morphometrical features allows for a more

accurate histological staging of fibrosis [2]. Here, we propose to utilise deep learning to identify collagen deposition patterns that will form the basis of a more objective and standardised assessment of the fibrosis progression in liver biopsies as defined by the METAVIR standard (see Figure 1). Rather than simply training a deep neural network and utilising the learnt model for a global prediction [18], our proposed method identifies the macro-patterns and characterises the different fibrosis stages through modelling the level of bridging.

2 Previous work

Whole Slide Images (WSIs) are typically split into smaller patches and the tile-level representations are later aggregated to predict the label of the entire slide. Such an approach has been used in diseases where information present at the cell-level is important to form a diagnosis - e.g. in cancer.

The assessment of fibrosis, however, is unique in that the grading system describe changes (bridging) which occur at a macroscopic scale as opposed to cellular scale (see Fig. 1) [5]. In the case of liver disease, a single fibrotic bridge might even span the width of a typical biopsy. For this reason, current approaches use either downsampled WSIs [18] or very large patches of tissue [9] as inputs to a CNN classifier. The limitation of such methods is that they rely on large sections of tissue, the collection of which is not possible in a clinical setting. Other approaches used handcrafted features and leveraged the knowledge of tissue architecture to achieve more interpretable results [17]. The downside of the patch-based methods is that such representations do not take into account the tissue topology at the macroscopic scale.

Graphs provide a unique way of describing spatial relationships between objects [14]. More recently, Graph Convolutional Networks (GCNs) have been applied to histopathology. While CNNs operate on patches directly, GCNs apply the convolutions to predefined graph nodes. Such nodes can represent biological entities as in the case of cell-graphs [20], or larger tissue regions - superpixels [13] or groups of tiles of similar appearance [19]. We propose a novel tissue representation that combines the robustness of deep learning with the prior knowledge of tissue architecture by employing a graph-based model. Such a representation resembles a human reader's understanding of the slide, leading to more interpretable analysis of fibrosis.

3 Data

A set of 271 percutaneous liver biopsies from patients with NAFLD, representing varying stages of fibrosis (Fig. 1) was provided by Perspectum Ltd.. The slides have been stained with Picrosirius Red (PSR), which dyes tissue collagen red. Each biopsy has been assigned a fibrosis grade by an expert pathologist in accordance to the METAVIR standard (F0-F4) [5]. The slides have been scanned with an Aperio scanner at 20x magnification. The dataset has been divided into training and testing folds with a 80:20 ratio.

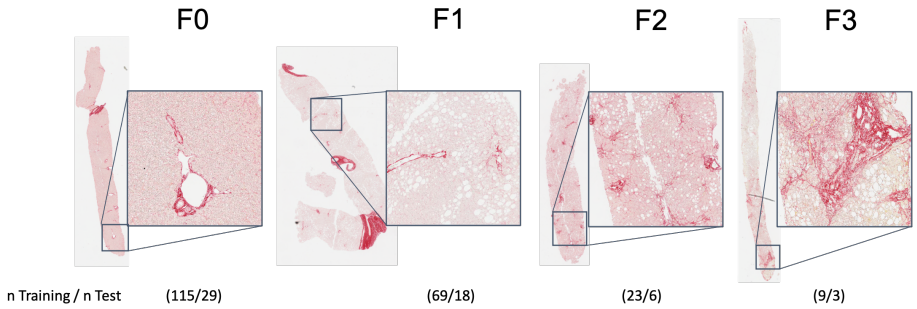


Fig. 1: Liver fibrosis stages (METAVIR standard). Healthy liver (grade F0) shows little collagen (red stain), which is almost only present in portal tract walls. During the initial stages of fibrosis (F1-F2) a network of fine collagen strands appears in the parenchyma. In the higher stages (F3-F4) the pathological fibres form "bridges" connecting portal tracts and central veins, resulting in a cobweb-like macro-pattern. It should be noted that the highest grade (F4 - cirrhosis) is not present in the dataset.

4 Methodology

The task of fibrosis detection can be modelled as a multi-class classification problem. Each slide is decomposed into a set of tiles and one of C classes is assigned to the whole sample. The classes are derived directly from the pathology staging system. In this work, we propose a novel method of tissue graph construction based on the identification of liver tissue landmarks (portal tracts and central veins) and partitioning of the slide into regions centred around said landmarks. Firstly, all slides are pre-processed by segmenting the tissue collagen at full image resolution. The landmarks are identified by a tile clustering pipeline.

A tissue graph is then constructed using the labelled tile representations as nodes. Regions of dense collagen identified by the clustering pipeline are used as input landmarks for a tessellation algorithm. The resulting graph consists of star-like structures, each one containing a biologically relevant region of the tissue (e.g. a single portal tract or region of fibrosis). As a baseline method we used the approach from Yu et al. [18] with a ResNet18 applied to the whole pre-processed and downsampled slides [8]. In order to satisfy the memory limitations, the WSIs were here downsampled by a factor of 32.

4.1 Tile subtyping

A CNN trained on dense annotations created using QuPath [1] is used to identify regions that contain collagen. To address the problem of variation in staining slides are binarised using a specifically trained collagen segmentation CNN. Regions of high collagen content are identified using the tile subtyping pipeline

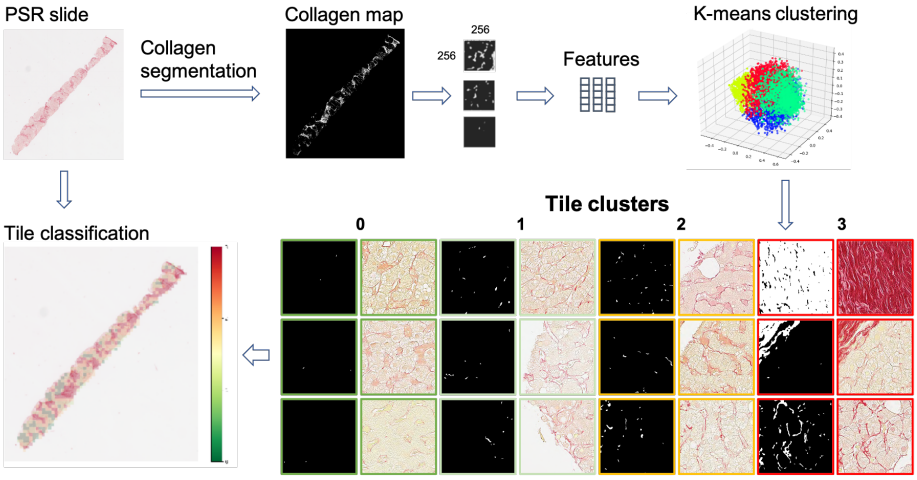


Fig. 2: Tile subtyping pipeline. 4 clusters of collagen tiles have been identified, each representing a different level of collagen content. The classified tiles are coloured according to their predicted cluster. Regions with high collagen content (cluster 3) are later used as Voronoi centres in tissue graph construction. The tile size of 256x256 px was specifically chosen to capture the local pattern of the collagen fibres and at the same time to limit the field of view to only contain one type of collagen pattern at a time.

(Fig. 2). Features from small tiles (256 x 256 px) are extracted from the full resolution slides using an ImageNet pretrained ResNet18 model, and the resulting feature vectors $\in \mathbb{R}^{512}$ are clustered using a k-means clustering algorithm. About 10% of tiles randomly sampled from the whole dataset are used to learn the fibre representation. Experimenting on different numbers of clusters, we found that setting $k = 4$ results in easily interpretable clusters with tiles containing increasing amount of collagen. In particular, tiles from cluster 3 contained regions of dense collagen: liver vasculature and regions of fibrosis.

4.2 Graph construction

The graph construction pipeline is presented in Figure 3. Firstly, regions of high collagen content are identified using the tile labels generated by the tile subtyping pipeline (i.e. tiles with collagen in cluster 3 - see Fig. 2). Next, centroids of each dense collagen region are used as centre points for Voronoi tessellation. The final tissue graph is constructed using features extracted with an ImageNet pretained ResNet18 model from each tile as node features X. The Euclidean distance D from each tile to its corresponding Voronoi centre is encoded as edge weight w. The graph construction was implemented using NetworkX package [7].

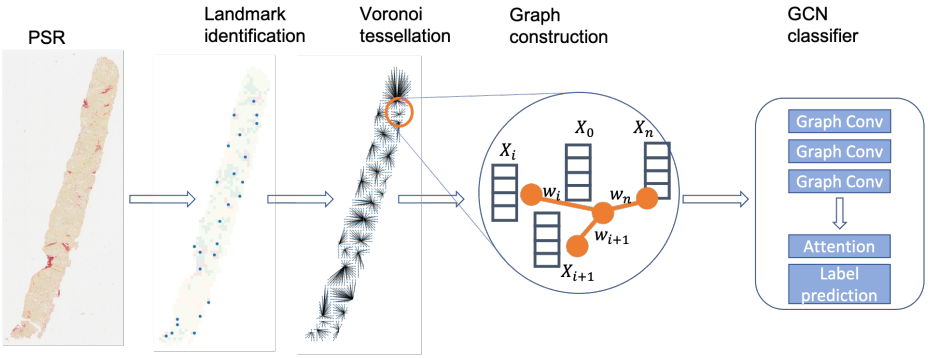


Fig. 3: The slide classification pipeline. First, tissue landmarks are identified and are used as centres for a Voronoi tessellation of image tiles. Next, a tissue graph is constructed using features extracted from individial tiles and the tessellation weights. Finally, the graph representing the whole slide serves as an input to the GCN classifier.

4.3 Graph convolutional layers

Given the constructed graph $G(V, E)$ with N nodes, the associated features matrix $X \in \mathbb{R}^{N,512}$ and adjacency matrix $A \in \mathbb{R}^{N,N}$ are used as an input to a 3-layer GNN. At each layer, a graph convolutional operator from GCN [11] is applied, followed by a non-linear activation function. The final nodes features matrix $H \in \mathbb{R}^{N,64}$ contains nodes representations that encode both the local graph structure and the features of the nodes. Here G^* denotes each of the convolutional layers:

$$H' = ReLu(G^*(X, A)) \;\; H'' = ReLu(G^*(H', A)) \;\; H = tanh(G^*(H'', A)). \quad (1)$$

As alternatives we also tested GAT [16] and GIN [15] and results can be found in Table 1.

4.4 Attention layer

After the last graph layer, an attention layer is used to aggregate the node vectors into a set of 3 vectors, each vector being a representation of the input slide with respect to one of the output classes. For each class, a set of attention scores is computed for the input vectors: each attention score represents the importance of an input vector for that class. The sum of the input vectors weighted by their corresponding attention scores generates the final vector representation of the slide with respect to the class.

Formally, attention for a class c is computed as a function of the input sequence H and a query vector u_c, jointly learned by the network during training. Similarity between each input element H_i and the query vector u_c is computed

through dot product to obtain the corresponding attention weight $\hat{\alpha}_{ci}$. A softmax function is then applied to the vector of weights $\hat{\alpha}_c$ to obtain normalised weights α_c. Finally, the attention weights are used to compute a weighted sum of the input values and obtain the final slide representation v_c.

$$v_c = \sum_i \alpha_{ci} \mathbf{H}_i \qquad \text{with} \qquad \alpha_c = Softmax(\mathbf{H} u_c^\top). \tag{2}$$

Each slide vector v_c is passed to a fully connected layer for the final classification. The classifier has a separate set of weights to compute the raw score for each class $\hat{y}_c$. The class scores $\hat{y}_c$ are then concatenated and Softmax is used to obtain the likelihood vector $\hat{y}$.

$$\hat{y} = Softmax(\hat{y}_0, ..., \hat{y}_C) \qquad \text{with} \qquad \hat{y}_c = \mathbf{W}_c^\top v_c + b_c. \tag{3}$$

The cross-entropy loss is computed from the likelihood vector $\hat{y}$ and the ground truth y. Training was performed to minimise the loss function by backpropagation.

5 Results

The proposed models were implemented using PyTorch [12] and PyTorch Geometric [3]. Training was performed using the Adam optimiser [10] with initial learning rate of 1e−02, $\beta_0 = 0.9$, $\beta_1 = 0.999$, $\epsilon = 1e{-}08$. For the GCN, we use a weight decay factor of $1e-5$. Models were trained for at least 300 epochs and training was stopped when the change of the loss function was negligible. When training the GCN, we applied dropout with probability 0.2 before the final classifiers. We use a batch size of 64 for the GCN and a batch size of 1 for the baseline ResNet18 due to memory limitations. The results reported in Table 1 are averaged across three runs. All models have been trained on an NVIDIA Quadro RTX 6000 GPU. The source code of our analysis is available at https://github.com/mkatw/gnn-fibrosis.

The results of the quantitative analysis of slide fibre content using the identified tile subtypes are presented in Figure 4. The weighted fibrosis score is calculated using the formula from equation (4):

$$WFS = \frac{1}{C-1} \sum_{c=0}^{C-1} cp_c \ , \tag{4}$$

where C is the number of classes and p_c is the fraction of tiles of class c in a given slide. A large variation in the overall fibrosis score can be seen within each of the pathology grades. The variation in the weighted score is expected, as a similar phenomenon is present in CPA measurements across the progressing stages of fibrosis [6]. Measuring the distribution of tile subtypes allows for a more fine-grained quantitative analysis of fibrosis in a slide. Typical examples of graphs

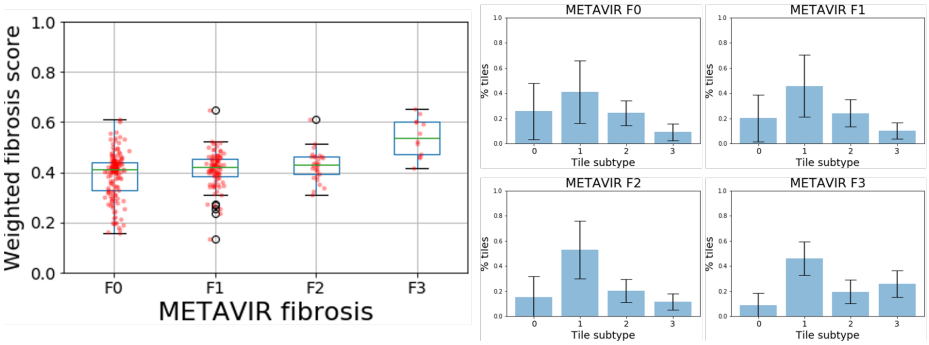


Fig. 4: The weighted fibrosis score across fibrosis stages and the distribution of tile subtypes across each METAVIR stage. The weighted fibrosis score, which is not informed by tissue topology, does not separate between classes F0 and F1 (Wilcoxon signed-rank test coefficients: W=-1.798, p=0.072) and classes F1 and F2 (W=-0.775, p=0.438). However, class F3 can be differentiated from all the other classes: F0 (W=-4.436, p<0.001), F1 (W=-4.181, p<0.001), F2 (W=-3.559, p<0.001), and class F2 can be differentiated from F0 (W=-2.062, p=0.039).

Table 1: F1 scores across fibrosis classes. Results expressed in %.

Model	**F0**	**F1**	**F2**	**F3**
ResNet18	53.81 ± 6.36	21.40 ± 12.73	16.93 ± 15.00	24.44 ± 21.43
GCN	65.09 ± 5.50	19.75 ± 11.63	**28.15** ± 24.48	70.71 ± 12.02
ATN-GCN	51.63 ± 2.83	27.27 ± 15.62	21.99 ± 6.72	**75.56** ± 7.70
GAT	**71.32** ± 8.56	**34.56** ± 21.11	0.00 ± 0.00	44.44 ± 38.49
ATN-GAT	59.15 ± 9.09	33.23 ± 28.91	22.22 ± 38.49	19.05 ± 16.49
GIN	68.31 ± 0.23	13.33 ± 23.09	0.00 ± 0.00	13.33 ± 23.09
ATN-GIN	68.18 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.003 ± 0.00
	F0 - F1	-	**F2**	**F3**
ResNet18	67.98 ± 2.26		26.07 ± 7.74	0.00 ± 0.00
GCN	**93.01** ± 0.53		18.79 ± 18.21	0.80 ± 0.00
ATN-GCN	89.26 ± 2.58		**28.28** ± 11.00	**78.57** ± 6.18
GAT	92.07 ± 1.40		27.78 ± 25.46	61.11 ± 9.62
ATN-GAT	86.22 ± 6.98		0.00 ± 0.00	0.00 ± 0.00
GIN	90.02 ± 0.00		0.00 ± 0.00	0.00 ± 0.00
ATN-GIN	90.02 ± 0.00		0.00 ± 0.00	0.00 ± 0.00
	F0 - F1 - F2	-	-	**F3**
ResNet18	74.19 ± 1.78			6.36 ± 5.53
GCN	99.03 ± 0.00			85.71 ± 0.00
ATN-GCN	**99.36** ± 0.55			**88.57** ± 10.30
GAT	96.85 ± 1.08			38.89 ± 34.69
ATN-GAT	96.90 ± 1.04			0.00 ± 0.00
GIN	96.30 ± 0.00			0.00 ± 0.00
ATN-GIN	96.30 ± 0.00			0.00 ± 0.00

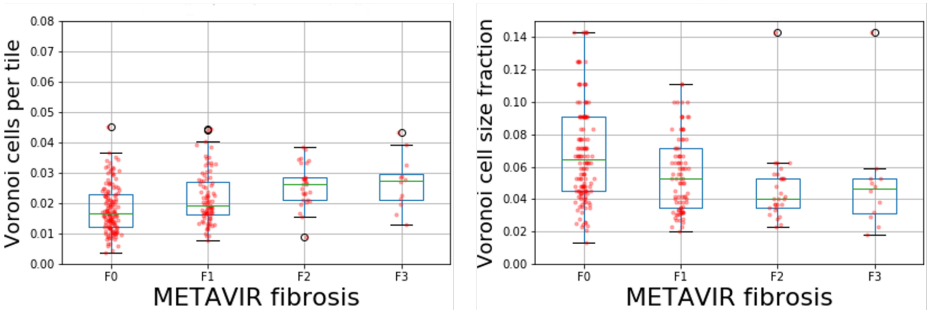
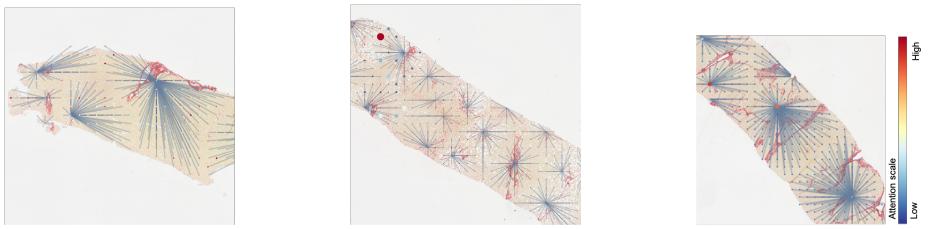


Fig. 5: The number and size of the Voronoi cells across the fibrosis stages. It can be seen that the size of a Voronoi cell relative to the area of the biopsy can change by an order of magnitude, therefore allowing to capture biological objects of varying sizes.

are presented in Figure 6. Selected properties of the constructed tissue graphs are shown in Figure 5.

The F1 scores of the trained models are gathered in Table 1. Here, we have tested three distinct classification setups. In the first, all fibrosis stages (F0-F3) were used directly as prediction classes. In the second and third setup, classes (F0 - F1) and (F0 - F2) were merged, respectively. This was done because stage F0, F1, and F2 cases have a similar macroscopic appearance and are often difficult to distinguish even for a trained pathologist. The experiments demonstrate that our GCN performs better than ResNet18 in each of the classification tasks.



(a) (b) (c)

Fig. 6: Examples of tissue graphs superimposed on the original images. Node size and colour reflect attention activation for the predicted slide class. Notice that the edge length in the individual sub-graphs corresponds to the underlying pattern of fibrosis and that the sub-graphs capture biologically meaningful regions, e.g. (6a) Stage F0 with no signs of fibrosis. (6b) Stage F2 with early fibrotic clusters in the parenchyma. (6c) Stage F3 with prominent bridging.

6 Conclusion

We have proposed a new model that facilitates a quantitative analysis of fibrosis in liver biopsies. The developed pipeline allows for an explicit separation of fibres in the slide into localised fibrosis patterns and the individual regions can be inspected by a pathologist. This representation fully aligns with established fibrosis scoring such as METAVIR. Importantly, this approach can be integrated into an interactive scoring system where pathologists can eliminate specific graphs form the analysis. The method could easily be extended to different fibrosis stainings and to other organs.

Acknowledgements:

MW is funded by the UK Engineering and Physical Sciences Research Council and Medical Research Council (EP/L016052/1) and in part by Perspectum Ltd.. NGM is supported by CRUK Oxford Centre Prize DPhil Studentship (C2195/A27450). HF is funded by NIH 5R01CA193694-02. SM and JR are supported by the PathLAKE Centre of Excellence for digital pathology and artificial intelligence which is funded from the Data to Early Diagnosis and Precision Medicine strand of the HM Government's Industrial Strategy Challenge Fund, managed and delivered by Innovate UK on behalf of UK Research and Innovation (UKRI). Views expressed are those of the authors and not necessarily those of the PathLAKE Consortium members, the NHS, Innovate UK or UKRI. Grant ref: File Ref 104689/application number 18181. JR is in part funded by the National Institute for Health Research Oxford Biomedical Research Centre.

References

1. Bankhead, P., Loughrey, M.B., Fernández, J., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., James, J.A., Salto-Tellez, M., Hamilton, P.W.: Qupath: Open source software for digital pathology image analysis. Scientific Reports **7**(1), 16878 (2017)
2. Calvaruso, V., Burroughs, A.K., Standish, R., Manousou, P., Grillo, F., Leandro, G., Maimone, S., Pleguezuelo, M., Xirouchakis, I., Piero Guerrini, G., et al.: Computer-assisted image analysis of liver collagen: Relationship to ishak scoring and hepatic venous pressure gradient. Hepatology **49**(4), 1236–1244 (2009)
3. Fey, M., Lenssen, J.E.: Fast graph representation learning with pytorch geometric (2019)
4. Friedman, S.L., Neuschwander-Tetri, B.A., Rinella, M., Sanyal, A.J.: Mechanisms of nafld development and therapeutic strategies. Nature Medicine **24**(7), 908–922 (2018)
5. Goodman, Z.D.: Grading and staging systems for inflammation and fibrosis in chronic liver diseases. Journal of Hepatology **47**(4), 598–607 (2007)
6. Goodman, Z.D., Becker, R.L., Pockros, P.J., Afdhal, N.H.: Progression of fibrosis in advanced chronic hepatitis c: evaluation by morphometric image analysis. Hepatology **45**(4), 886–894 (2007)

7. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using networkx. In: Varoquaux, G., Vaught, T., Millman, J. (eds.) Proceedings of the 7th Python in Science Conference. pp. 11 – 15. Pasadena, CA USA (2008)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2016-December**, 770–778 (2016)
9. Heinemann, F., Birk, G., Stierstorfer, B.: Deep learning enables pathologist-like scoring of nash models. Scientific Reports **9**(1), 1–10 (2019)
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. 3rd International Conference for Learning Representations, San Diego, 2015 (2017)
11. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations (ICLR) (2017)
12. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019)
13. Pati, P., Jaume, G., Fernandes, L.A., Foncubierta-Rodríguez, A., Feroce, F., Anniciello, A.M., Scognamiglio, G., Brancati, N., Riccio, D., Di Bonito, M., De Pietro, G., Botti, G., Goksel, O., Thiran, J.P., Frucci, M., Gabrani, M.: HACT-Net: A Hierarchical Cell-to-Tissue Graph Neural Network for Histopathological Image Classification. Lecture Notes in Computer Science **12443 LNCS**, 208–219 (2020)
14. Sharma, H., Zerbe, N., Lohmann, S., Kayser, K., Hellwich, O., Hufnagl, P.: A review of graph-based methods for image analysis in digital histopathology. Diagnostic Pathology **1**(1), 1–51 (2015)
15. Veličković, P., Casanova, A., Liò, P., Cucurull, G., Romero, A., Bengio, Y.: Graph attention networks. 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings pp. 1–12 (2018)
16. Xu, K., Jegelka, S., Hu, W., Leskovec, J.: How powerful are graph neural networks? 7th International Conference on Learning Representations, ICLR 2019 pp. 1–17 (2019)
17. Xu, S., Wang, Y., Tai, D.C., Wang, S., Cheng, C.L., Peng, Q., Yan, J., Chen, Y., Sun, J., Liang, X., Zhu, Y., Rajapakse, J.C., Welsch, R.E., So, P.T., Wee, A., Hou, J., Yu, H.: QFibrosis: A fully-quantitative innovative method incorporating histological features to facilitate accurate fibrosis scoring in animal model and chronic hepatitis B patients. Journal of Hepatology **61**(2), 260–269 (2014)
18. Yu, Y., Wang, J., Ng, C.W., Ma, Y., Mo, S., Fong, E.L.S., Xing, J., Song, Z., Xie, Y., Si, K., et al.: Deep learning enables automated scoring of liver fibrosis stages. Scientific reports **8**(1), 1–10 (2018)
19. Zheng, Y., Jiang, B., Shi, J., Zhang, H., Xie, F.: Encoding histopathological WSIs using GNN for scalable diagnostically relevant regions retrieval. Lecture Notes in Computer Science **11764 LNCS**, 550–558 (2019)
20. Zhou, Y., Graham, S., Alemi Koohbanani, N., Shaban, M., Heng, P.A., Rajpoot, N.: CGC-net: Cell graph convolutional network for grading of colorectal cancer histology images. Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019 pp. 388–398 (2019)