

SQFIBROSIS: A ROBUST LIVER FIBROSIS SCORING SYSTEM FOR TELEPATHOLOGY

Yang Yu^{*,1,2}, Jiahao Wang^{*,1}, Hanry Yu^{†,1}

¹ National University of Singapore

² Institute for Infocomm Research (I²R), A*STAR, Singapore

ABSTRACT

Histopathological scoring is the gold standard for assessing liver fibrosis on biopsy images. However, accurate computational assessment of liver fibrosis suffers from data variations of the sampling error caused by staining and imaging heterogeneity in variable clinical lab settings, particularly in the context of computer-assisted diagnostics for telepathology applications. Experienced pathologists address these challenges by leveraging their expertise to reference samples, and selecting relevant spatial features while filtering out noise. Drawing from this domain knowledge, we present sqFibrosis (spatially-sensitive and quantitative fibrosis), a robust and automated scoring index with pathological-relevant features incorporated to provide quantitative analysis for staging liver fibrosis using conventional collagen-stained images and standard light microscopes. Specifically, sqFibrosis consolidates 63 architectural and morphological features from various collagen fibers (portal, fibrillar and septal collagen), resulting in a data-efficient single index for liver fibrosis evaluation. This index also demonstrates the robustness to sampling errors introduced from staining and imaging variations. Furthermore, it is well-suited for liver fibrosis assessment within telepathology settings.

Index Terms— Machine Learning, Histopathology, Liver Fibrosis, Data Efficient, Staining and Imaging Variations.

1. INTRODUCTION

Telepathology, as the practice of pathology at a distance, uses telecommunications technology to facilitate the transfer of pathology data between distant locations for diagnosis, education, and research [1]. It has been successfully used for many applications, especially the rendering of histopathology tissue diagnosis [2]. However, telepathology still grapples with staining and imaging variations since it lacks direct control over the image acquisition process. Therefore, there is a clear need for a robust evaluation system with a high tolerance for such sampling errors to address these challenges.

Liver fibrosis, a common liver disease primarily caused by the excessive accumulation of extracellular matrix (ECM)

during the wound healing process [3], faces significant challenges in accurate diagnosis due to staining and imaging variations within the realm of telepathology. Such variations could introduce visually inconsistent patterns among histopathological images, thereby affecting the diagnostic performance in clinical settings, especially without expert domain knowledge. To overcome these limitations, efforts have been made by using digital pathology to improve the quantification of liver fibrosis in telepathology [4]. The current method of choice is collagen proportionate area (CPA) measurement, which quantifies the extent of collagenous extracellular matrix (ECM) deposition without incorporating architectural information about the damaged tissue landscape [5]. Although CPA appears to be of potential value for classifying liver fibrosis, its non-standardized image analysis methods and lack of analysis on the related pathological features impede its applicability in routine practice for telepathology [6]. While other approaches use multi-photon imaging techniques to circumvent the need for histological staining and, therefore, eliminate any variations that might arise from the staining procedure [7], the clinical applicability of these techniques is still being questioned as the multi-photon microscope is not commonly available in clinics or laboratories, unlike the standard light microscope.

Despite the integration of deep learning-based algorithms for digitalized pathology has significantly contributed to the rapid adoption of telepathology for liver fibrosis [8, 9], the performance of such method still heavily relies on substantial source data for training, which may be difficult to obtain in practice. Deep learning-based algorithms also exhibit overfitting issues on source data and could not work well in the realm of telepathology for target data, due to the limited access to a large-scale dataset from a single clinical centre and/or the presence of various noises in multiple centres' datasets. However, experienced pathologists could still make accurate and robust clinical diagnosis regardless of the variations by extracting the most relevant features for disease progression based on the experience learnt from a limited amount of the previously evaluated data [10]. Therefore, we hypothesize that a robust machine learning-based scoring index could be established with an improved signal-to-noise ratio from semantic features by mimicking the pathological learning process for accurate liver fibrosis staging in telepathology.

* Contributed equally

† Corresponding author: medyuh@nus.edu.sg

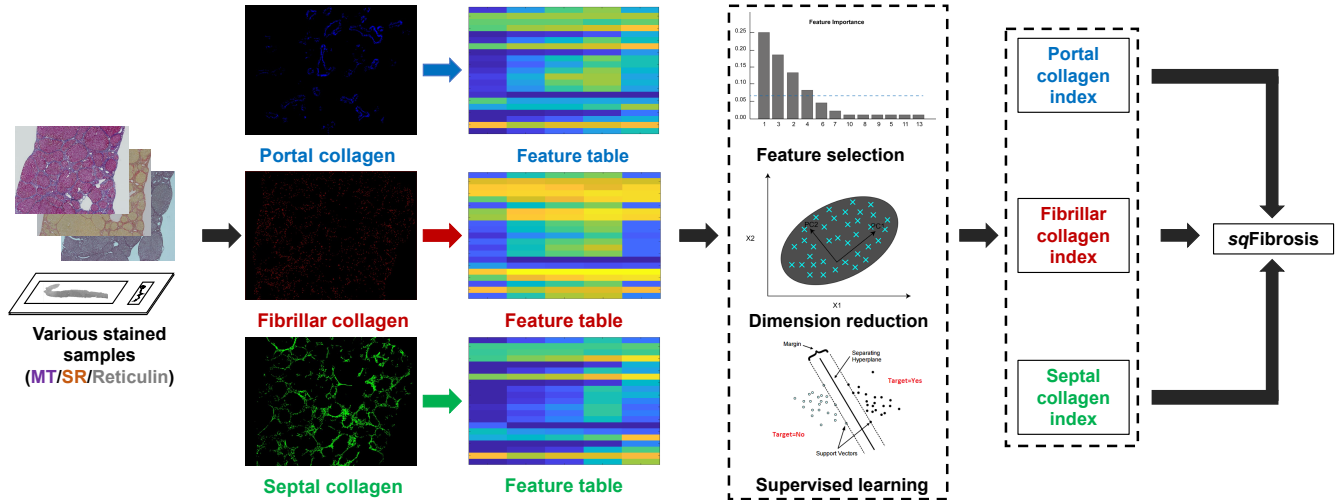


Fig. 1. Schematic illustration of sqFibrosis establishment. It is used in liver biopsy-based fibrosis staging as a fully quantitative and automated method.

To this end, we present sqFibrosis (spatially-sensitive and quantitative fibrosis), a robust and data-efficient liver fibrosis scoring index that enables detailed and accurate quantification of histopathological architectural features using conventional collagen-stained samples and standard light microscopes. By integrating the pathological learning process in telepathology into machine learning-aided imaging analytics, sqFibrosis reduces the sampling error and makes the diagnosis more robust to the staining and imaging variations with detailed spatial information and filtered contributors. It is also resource-efficient to mitigate the challenges of data inaccessibility and scarcity. Furthermore, it could facilitate fully quantitative and automated liver fibrosis assessment in standard clinical or laboratory settings, particularly in the context of telepathology with limited training samples.

2. PROPOSED METHODS

Fig. 1 shows the established workflow of sqFibrosis for telepathology. When provided with a histopathological slide using conventional staining protocols, our objective is to identify distinct collagen distribution patterns, namely, portal collagen (portal tract stroma and portal/periportal fibrosis), fibrillar collagen (perisinusoidal fibrosis), and septal collagen (septal fibrosis caused by bridging necrosis) [11]. These patterns serve as robust representations of disease progression, effectively reducing the data burden for representation learning. We employ unsupervised clustering methods such as k-means [12] to segment various types of collagen in the colour space. Subsequently, we extract these patterns' corresponding architectural and textural features, transforming them into quantitative parameters. In the computational framework for establishing the final index, we extract and categorize 63 architectural and morphological features into three

groups: portal, fibrillar, and septal collagen, respectively. Feature selection is then performed to identify the optimal feature combination. Principal component analysis (PCA) [13] is also exploited for dimension reduction, reducing the dimensionality of the selected collagen features to focus on those relevant to disease progression. Support vector machines (SVM) [14], as the classification model, is utilized to generate individual indexes for three different types of collagen. Lastly, multinomial logistic regression [15] is used to combine these three indexes into a single index for fibrosis evaluation, known as the sqFibrosis.

3. IMPLEMENTATION DETAILS

Sample preparation and image acquisition The Thioacetamide (TAA)-induced animal model was used for studying liver fibrosis in rats. Liver specimens from the left lateral liver of each animal were formalin-fixed, paraffin-embedded, and sectioned into consecutive slides of $50\mu\text{m}$ for Masson Trichrome (MT), Sirius Red (SR) and Reticulin (Retic) staining for histological examination and imaging. Pathological scoring was performed by an experienced pathologist using the Metavir fibrosis staging system (F0-F4) [16]. A total of 100 images (F0: 20 images, F1: 20 images, F2: 20 images, F3: 20 images, F4: 20 images) were captured by the system of a bright field microscope with a 4x objective on stained sections of the stained samples. To cover most of the sample areas, images were acquired for the animal samples with an image size of 4mm by 3mm .

Automated image processing algorithms Methods were developed to identify and differentiate portal, septal and fibrillar collagen. Images were first acquired from stained slides using a bright field microscope and separated into tissue and collagen based on K-mean clustering. The adaptive thresholding

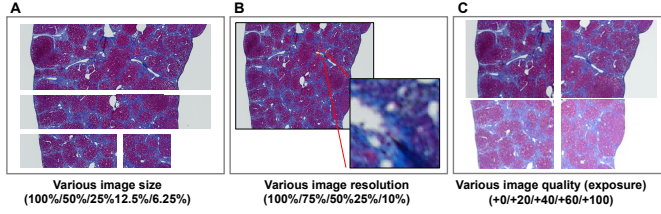


Fig. 2. Exemplary images for showing simulated staining and imaging variations.

was then applied, including morphological opening/closing and hole-filling operations to binarize the tissue images and smooth the spaces. Sinusoidal space was also removed before further processing. The segmented collagen images were further processed using contrast enhancement and adaptive thresholding. The morphological closing and despeckling were also used to connect the fragmented collagen fibers and to remove the noise. The fibrillar collagen was selected first for those fibers with areas less than a certain threshold, representing the fragmented collagen distributed in the tissue (red). For portal collagen, the pixels from the boundary of the portal tract or central vein were first selected and a cut-off distance was chosen to dilate those pixels where all the collagen within the cut-off distance was considered as portal collagen (blue). Lastly, all the remaining collagen was interpreted as septal collagen (green) (Fig 1).

Sampling error simulation To determine the sensitivity of sqFibrosis index to staining and imaging variations, we performed a proof-of-concept study using images with variations in size, resolution and quality (Fig.2). Images with different sizes were obtained from a large section of the liver containing a sufficient number of portal tracts for accurate scoring by an experienced pathologist (Fig.2A); images with varying resolutions were obtained by applying different resizing ratios on a full-size image (Fig.2B); and images of differing quality were prepared by manually adjusting the image contrast by varying the duration of exposure time during image acquisition process (Fig.2C).

Statistical analysis The leave-one-out cross-validation rule was applied for the training and testing datasets with pre-processed images and selected features (architectural and morphological features specific to tubular-shaped objects including length, width, orientation, cross-link spaces and cross-link density, etc. [8]) to test the classification performance. The two-tailed Wilcoxon rank-sum and Kruskal Wallis tests were performed to estimate the statistical differences of sqFibrosis index and CPA between different Metavir fibrosis stages. The DeLong test was used to compare the receiver-operating-characteristics curves (ROCs) and area under ROCs (AUROCs) of sqFibrosis and CPA. The statistical significance level was set as $p < 0.05$.

Table 1. Area under the curve (AUROC) values of sqFibrosis and collagen proportionate area (CPA) using different staining approaches (Masson Trichrome, Sirius Red, Reticulin). T1-T4 represents four binarized classifications for F0 vs. F1-F4; F0-F1 vs. F2-F4; F0-F2 vs. F3-F4; F0-F3 vs. F4. The best results are marked in **bold**.

Metavir Fibrosis Staging	Binarized Classification			
	T1	T2	T3	T4
Masson Trichrome				
CPA	0.966	0.911	0.927	0.926
sqFibrosis (Ours)	0.995	0.990	0.984	0.953
Sirius Red				
CPA	0.856	0.930	0.944	0.961
sqFibrosis (Ours)	0.946	0.939	0.980	0.968
Reticulin				
CPA	0.689	0.721	0.751	0.840
sqFibrosis (Ours)	0.739	0.690	0.830	0.894

4. EXPERIMENTAL RESULTS

We first ask whether the sqFibrosis is capable of scoring liver fibrosis in conventional stained samples on MT as effectively as traditional histological scoring systems, such as the Metavir staging system. Results show that the sqFibrosis was capable of reflecting the continuum of fibrosis stages where its values increased with fibrosis progression and significantly differed between stages ($p < 0.001$) (Fig.3A). We next ask whether sqFibrosis would perform differently if other collagen dyes were to be used such as SR and Retic. MT stains mostly collagen type I while SR stains both collagen type I and III, providing more contrast between tissue and collagen components. Retic is mainly responsible for staining the reticulin network formed by collagen type III. Results indicate that SR-processed images via the sqFibrosis approach, similar to MT, could fully recapitulate the collagen distribution ($p < 0.001$), while signals arising from Retic were relatively lower ($p < 0.01$). This is likely due to the poor tissue penetration ability of the Retic dye (Fig.3A). Notably, while CPA could detect drastic changes in the late stages of fibrosis, it is not capable of differentiating between the normal and early stages (stages F0, F1 and F2) (Fig. 3B).

sqFibrosis also increases with fibrosis progression and showed significant differences among all the stages ($p < 0.001$) for both MT and SR stains with AUROC of 0.95-0.99 and 0.95-0.98, respectively. In contrast, CPA could only show drastic changes in the late stages and could not differentiate between the normal and early stages (stages 0, 1 and 2) for either MT or SR stains with AUROC of 0.91-0.97 and 0.86-0.96, respectively (Table 1). While the sqFibrosis could not detect the subtle difference among early stages of fibrosis for Retic, it could differentiate the difference among late stages of fibrosis with AUROC of 0.83-0.89. On the other hand, CPA fails to do so for all the stages of fibrosis for Retic

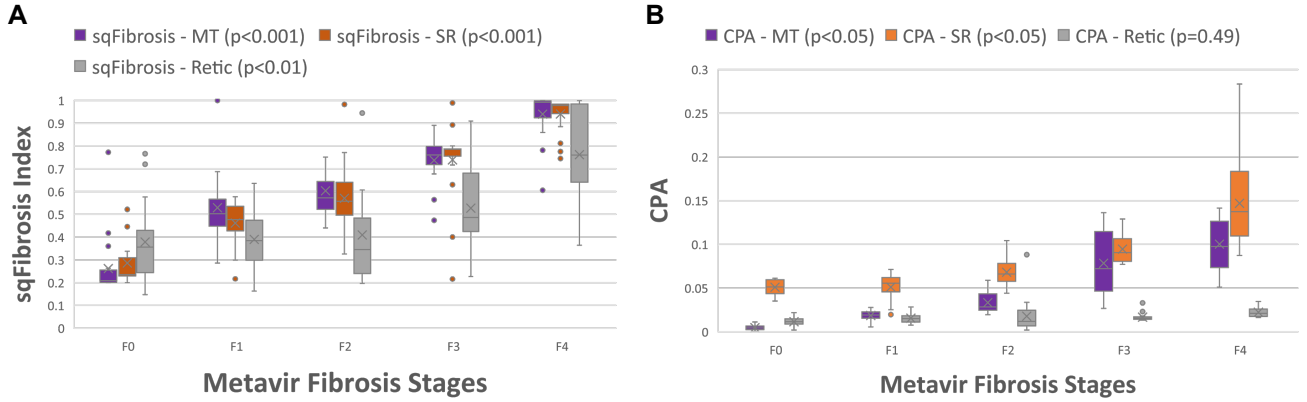


Fig. 3. Changes of sqFibrosis and collagen proportionate area (CPA) with fibrosis progression using Masson Trichrome (MT), Sirius Red (SR) and Reticulin (Retic) staining.

Table 2. Area under the curve (AUROC) values of sqFibrosis and collagen proportionate area (CPA) for liver fibrosis with different simulated imaging variations. T1-T4 represents four binarized classifications for F0 vs. F1-F4; F0-F1 vs. F2-F4; F0-F2 vs. F3-F4; F0-F3 vs. F4. The best results are marked in **bold**.

Binarized Classification	Metavir Fibrosis Staging			
	T1	T2	T3	T4
Image size (100% / 50% / 25% / 12.5% / 6.25%)				
CPA	0.966/0.919/0.865/0.819/0.758	0.911/0.868/0.855/0.810/0.766	0.927/0.913/0.857/0.809/0.752	0.926/0.863/0.839/0.789/0.738
sqFibrosis (Ours)	0.995/0.945/0.916/0.909/0.866	0.990/0.955/0.945/0.901/0.868	0.984/0.968/0.927/0.903/0.858	0.953/0.943/0.905/0.876/0.858
Image resolution (100% / 75% / 50% / 25% / 10%)				
CPA	0.966/0.944/0.913/0.850/0.802	0.911/0.900/0.878/0.844/0.795	0.927/0.914/0.904/0.840/0.786	0.926/0.910/0.863/0.807/0.777
sqFibrosis (Ours)	0.995/0.982/0.952/0.925/0.904	0.990/0.971/0.945/0.919/0.892	0.984/0.965/0.946/0.923/0.899	0.953/0.943/0.917/0.911/0.893
Image quality (+0 / +20 / +40 / +60 / +80)				
CPA	0.966/0.966/0.900/0.900/0.809	0.911/0.910/0.888/0.836/0.806	0.927/0.932/0.903/0.857/0.827	0.926/0.916/0.878/0.832/0.779
sqFibrosis (Ours)	0.995/0.996/0.952/0.924/0.893	0.990/0.980/0.948/0.920/0.904	0.934/0.952/0.938/0.891/0.880	0.953/0.936/0.917/0.880/0.860

staining with AUROC of 0.69-0.84.

Sampling errors caused by variations in image size, resolution and quality, are a major problem when stained samples are used for quantification during routine pathological review, especially in telepathology (Fig.2). The robustness to such variations for the sqFibrosis index as compared to CPA for fibrosis scoring is further investigated. The AUROC values of sqFibrosis decrease slightly along with the reduction in sample size. CPA achieves similar AUROC values as sqFibrosis using a larger sample. However, CPA fails to address the differences when the sample size is reduced to a smaller value. The difference between the AUROC values of sqFibrosis and CPA for differentiating liver fibrosis stages also becomes significant when samples are reduced from one-fourth to one-sixteenth ($p < 0.05$) of the original size (Table 2). We further evaluate sqFibrosis and CPA using varying image resolutions and qualities. Similar results are observed for the AUROC values of sqFibrosis and CPA regarding different image resolution and image qualities where CPA fails to address the different stages of fibrosis with less image resolution and poorer image qualities. The difference between the AUROC values of sqFibrosis and CPA becomes significant from 25% to 10% ($p < 0.05$) of the original resolution and from +60 to +80 of ex-

posure time ($p < 0.05$) for differentiating liver fibrosis stages (Table 2). Such results further prove that the sqFibrosis, with spatial-temporal information from various types of collagen fibers included, could potentially differentiate fibrosis at different stages and is quite robust to the sampling errors caused by the imaging variations.

5. CONCLUSION

We develop a robust and automated scoring index (sqFibrosis) with pathological-relevant features incorporated to provide quantitative scores for staging liver fibrosis using conventional collagen-stained images and standard light microscopes. It could faithfully reproduce the staging results of the histological assessment system and is also robust to sampling error introduced from staining and imaging variations. Thus, such a framework could potentially be a valuable tool to complement the utility of biopsy samples for accurate and robust assessment of liver fibrosis and other liver-related diseases in clinical research through telepathology. Further validations including comparison with deep learning-based algorithms for liver fibrosis assessment and validations on a larger scale clinical biopsies would be explored in the future studies.

6. COMPLIANCE WITH ETHICAL STANDARDS

All the protocols for studying TAA-induced liver fibrosis rat models were reviewed and approved by the Biological Resource Centre (BRC) Institutional Animal Care and Use Committee (IACUC).

7. ACKNOWLEDGMENTS

This work is supported in part by the Institute of Bioengineering & Bioimaging (IBB), Biomedical Research Council, Agency for Science, Technology and Research (A*STAR), A*STAR; A*STAR (A20D3b0073); IAF (H18/01/a0/017); NMRC (CIRG21nov-0032, CIRG22jul-0018)); SMART CAMP; The Institute for Digital Medicine (WisDM); and Mechanobiology Institute of Singapore (A-0003467-22-00) funding to HYU.

8. REFERENCES

- [1] Ronald S Weinstein, Anna R Graham, Lynne C Richter, Gail P Barker, Elizabeth A Krupinski, Ana Maria Lopez, Kristine A Erps, Achyut K Bhattacharyya, Yukako Yagi, and John R Gilbertson, "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future," *Human pathology*, vol. 40, no. 8, pp. 1057–1069, 2009.
- [2] Rashid L Bashshur, Elizabeth A Krupinski, Ronald S Weinstein, Matthew R Dunn, and Noura Bashshur, "The empirical foundations of telepathology: evidence of feasibility and intermediate effects," *Telemedicine and e-Health*, vol. 23, no. 3, pp. 155–191, 2017.
- [3] Scott L Friedman, "Mechanisms of hepatic fibrogenesis," *Gastroenterology*, vol. 134, no. 6, pp. 1655–1669, 2008.
- [4] SS Cross, T Dennis, and RD Start, "Telepathology: current status and future prospects in diagnostic histopathology," *Histopathology*, vol. 41, no. 2, pp. 91–109, 2002.
- [5] Vincenza Calvaruso, Andrew Kenneth Burroughs, Richard Standish, Pinelopi Manousou, Federica Grillo, Gioacchino Leandro, Sergio Maimone, Maria Pleguezuelo, Ilias Xirouchakis, Gian Piero Guerrini, et al., "Computer-assisted image analysis of liver collagen: relationship to ishak scoring and hepatic venous pressure gradient," *Hepatology*, vol. 49, no. 4, pp. 1236–1244, 2009.
- [6] Yi Huang, W Bastiaan De Boer, Leon A Adams, Gerry MacQuillan, Max K Bulsara, and Gary P Jeffrey, "Image analysis of liver biopsy samples measures fibrosis and predicts clinical outcome," *Journal of Hepatology*, vol. 61, no. 1, pp. 22–27, 2014.
- [7] Shuoyu Xu, Yan Wang, Dean CS Tai, Shi Wang, Chee Leong Cheng, Qiwen Peng, Jie Yan, Yongpeng Chen, Jian Sun, Xieer Liang, et al., "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*, vol. 61, no. 2, pp. 260–269, 2014.
- [8] Yang Yu, Jiahao Wang, Chan Way Ng, Yukun Ma, Shupai Mo, Eliza Li Shan Fong, Jiangwa Xing, Ziwei Song, Yufei Xie, Ke Si, et al., "Deep learning enables automated scoring of liver fibrosis stages," *Scientific reports*, vol. 8, no. 1, pp. 16016, 2018.
- [9] Yunchao Yin, Derya Yakar, Rudi AJO Dierckx, Kim B Mouridsen, Thomas C Kwee, and Robbert J de Haas, "Liver fibrosis staging by deep learning: a visual-based explanation of diagnostic decisions of the model," *European Radiology*, vol. 31, no. 12, pp. 9620–9627, 2021.
- [10] Tim Hulsén, Saumya S Jamuar, Alan R Moody, Jason H Karnes, Orsolya Varga, Stine Hedensted, Roberto Spreafico, David A Hafler, and Eoin F McKinney, "From big data to precision medicine," *Frontiers in medicine*, vol. 6, pp. 34, 2019.
- [11] Ramón Bataller, David A Brenner, et al., "Liver fibrosis," *The Journal of clinical investigation*, vol. 115, no. 2, pp. 209–218, 2005.
- [12] Abiodun M Ikotun, Absalom E Ezugwu, Laith Abualigah, Belal Abuhaija, and Jia Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, 2022.
- [13] Ian T Jolliffe and Jorge Cadima, "Principal component analysis: a review and recent developments," *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, pp. 20150202, 2016.
- [14] Jair Cervantes, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua, and Asdrubal Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.
- [15] Chanyeong Kwak and Alan Clayton-Matthews, "Multinomial logistic regression," *Nursing research*, vol. 51, no. 6, pp. 404–410, 2002.
- [16] Pierre Bedossa, Delphine Dargère, and Valerie Paradis, "Sampling variability of liver fibrosis in chronic hepatitis c," *Hepatology*, vol. 38, no. 6, pp. 1449–1457, 2003.