**Supplementary File 2. Obuchowski index.1**

The Obuchowski index calculates the accuracy of a diagnostic test when the gold standard is measured on a continuous, ordinal or nominal scale. The Obuchowski index is a weighted average of the area under the receiver operating characteristic (AUROC) values obtained for all possible pairs of fibrosis stages (i.e. 10 pairs for the five [F0–F4] fibrosis stages) to be differentiated. It estimates the probability that a test will correctly rank two randomly chosen patients with different stages of fibrosis.2 The Obuchowski Index is a rank based measure that can be calculated without constructing a receiver operating characteristic curve (ROC), although they can be interpreted similarly, they are not associated with ROC curves.3

1Obuchowski NA. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. Acad Radiol 2005;12(9):1198-1204. doi: 10.1016/j.acra.2005.05.013. PMID: 16099683.

2 Choi KJ, Jang JK, Lee SS, Sung YS, Shim WH, Kim HS, *et al*. Development and Validation of a Deep Learning System for Staging Liver Fibrosis by Using Contrast Agent-enhanced CT Images in the Liver. Radiology 2018;289(3):688-697. doi: 10.1148/radiol.2018180763. PMID: 30179104.

3Nguyen P. nonbinROC: Software for Evaluating Diagnostic Accuracies with Nonbinary Gold Standards. Journal of Statistical Software 2007;21(10) doi: 10.18637/jss.v021.i10.