

Diagnostic Accuracy of Convolutional Neural Network Based Machine Learning Algorithms in Endoscopic Severity Prediction of Ulcerative Colitis: A Systematic Review and Meta-analysis

Vinay Jahagirdar, MD¹; Saurabh Chandan, MD²; Babu P. Mohan, MD, MS³; Gursimran S. Kochhar, MD⁴

¹University of Missouri Kansas City School of Medicine, ²CHI Health Creighton School of Medicine, ³University of Utah School of Medicine, ⁴Allegheny Health Network

1. BACKGROUND

- Endoscopic severity of ulcerative colitis (UC) predicts clinical outcomes and is essential to guide treatment and evaluate therapeutic response.
- The Mayo endoscopic score (MES) is commonly used to objectively classify mucosal damage. It ranges from 0 to 3, with a higher score reflecting increased severity. With advances in machine learning, artificial intelligence is being employed for automating image analysis.
- Convolutional neural network (CNN) is a powerful deep learning method for image recognition, and in this study, we aim to look at diagnostic accuracy parameters of CNN based machine learning algorithms to predict UC severity.

2. METHODS

- Databases were searched from inception to May 2022 using specific terms for studies evaluating the diagnostic accuracy parameters of machine learning algorithms in assessing UC severity.
- Inclusion was restricted to studies that employed CNN based algorithms.
- Outcomes of interest were the pooled accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).
- Multiple 4X4 contingency tables assessing the diagnostic accuracy of the algorithms were considered independent of each other as the goal was to study the overall direction of pooled rates and not calculate precise point estimates.
- Standard meta-analysis methods were employed using the random-effects model, and heterogeneity was assessed using the I² statistics.

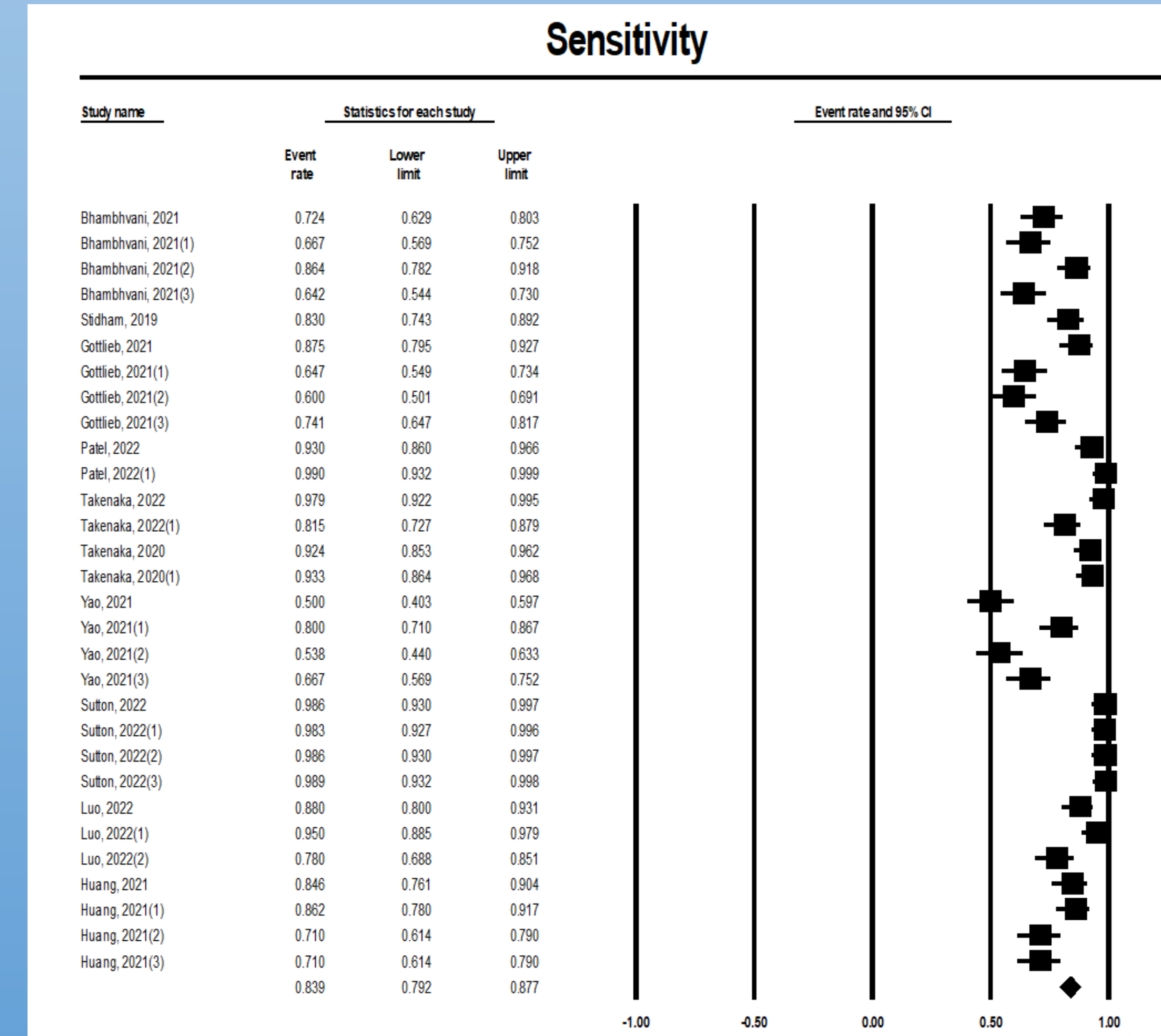


Figure 1: Forest Plot of Sensitivity

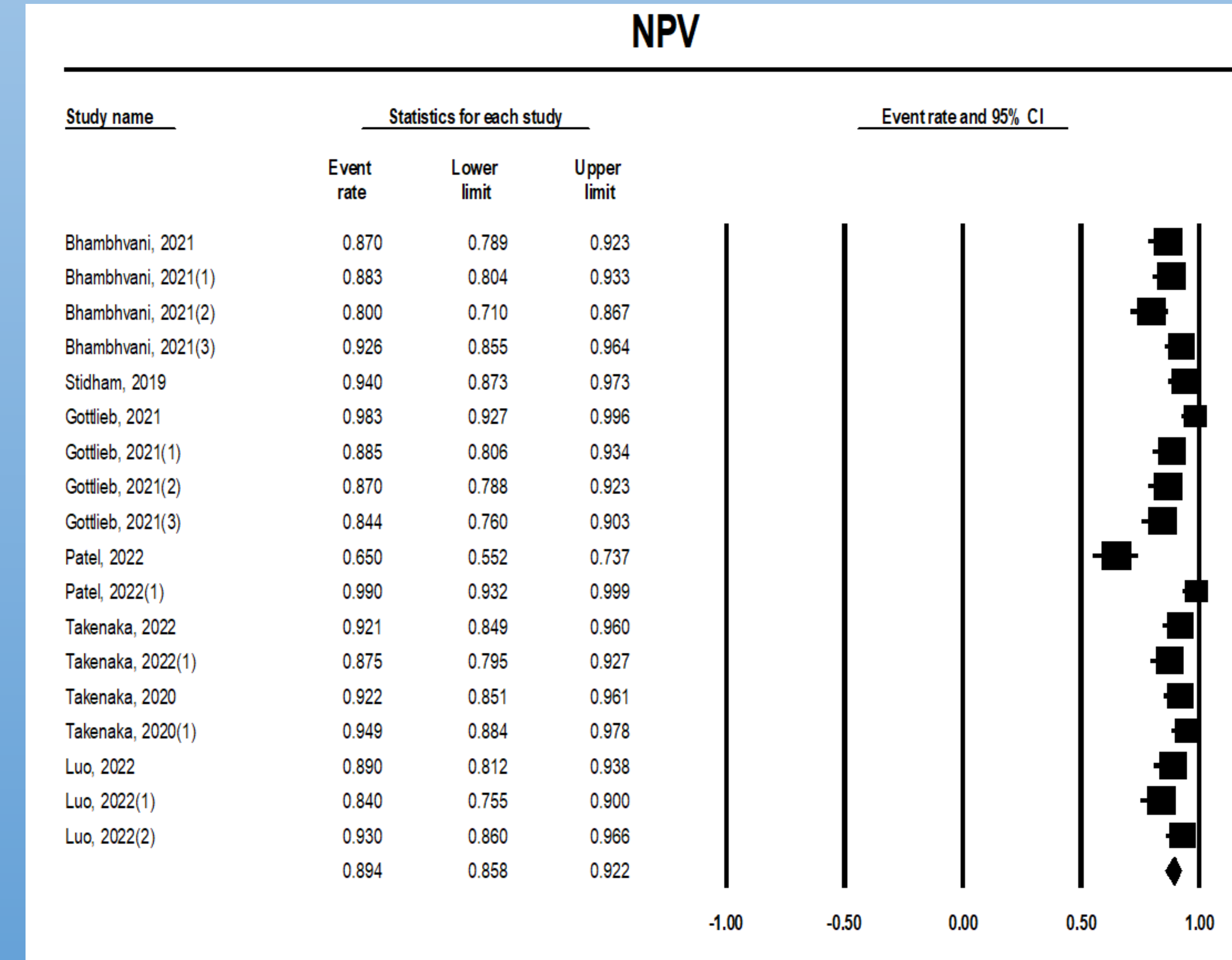


Figure 2: Forest Plot of Negative Predictive Value

| Outcome | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|--|--|--|--|--|
| Pooled rates (95% CI, I ² % heterogeneity) | 91.2% (87.4-93.9, 84%) 20 datasets | 83.9% (79.2-87.7, 89%) 30 datasets | 92.3% (89.5-94.4, 84%) 30 datasets | 86.5% (80.7-90.8, 89%) 18 datasets | 89.4% (85.8-92.2, 78%) 18 datasets |

Table 1: Summary of pooled results

Based on our meta-analysis of 12 studies, CNN-based machine learning algorithms demonstrated excellent pooled accuracy parameters.

3. RESULTS

- 12 studies were included that exclusively used CNN algorithm. Studies that used support vector machines or a combination were excluded.
- The CNN algorithm was trained and tested to predict Mayo score severity 0, 1, 2 & 3, individually in majority of the studies.
- In few studies the CNN algorithm was used to differentiate between Mayo 0 vs 1, and Mayo 0-1 vs 2-3.
- Although, 'ground-truth' differed, individual 4X4 tables were considered as independent of each other for the purpose of this study.
- Interestingly, on sub-group analysis, we found that UCEIS had higher sensitivity and positive predictive value in predicting severity of UC, than MES.

4. CONCLUSION

- Based on our meta-analysis of 12 studies, CNN-based machine learning algorithms demonstrated excellent pooled diagnostic accuracy parameters.
- Further work seems to be needed to get the NPV >90.
- Real-life clinical studies are needed to establish the role of AI in endoscopic severity assessment of UC.