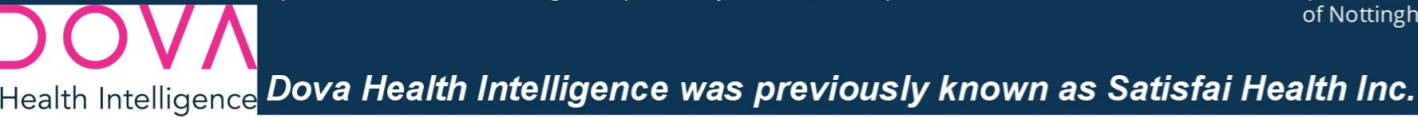


BUILDING A ROBUST ARTIFICIAL INTELLIGENCE SOLUTION FOR USE IN ULCERATIVE COLITIS CLINICAL TRIALS

Byrne, Michael F.^{1,2}; Requa, James²; Panés, Julián³; Bressler, Brian⁴; Panaccione, Remo⁵; Mendel, Robert²; East, James E.⁶; Parsa, Nasim^{7,2}; Banerjee, Rupa⁸; Kalapala, Rakesh⁸; Reddy, Nageshwar⁸; Rughwani, Hardik⁸; Flegg, Daniel⁹; Kobayashi, Taku¹⁰; Asonuma, Kunio¹⁰; Moran, Gordon W.¹¹; Gallinger, Zane¹²; Cheung, Vincent T.⁶; Tan, Malcolm¹³; Ma, Christopher¹³; Travis, Simon P.⁶; Jairath, Vipul¹³

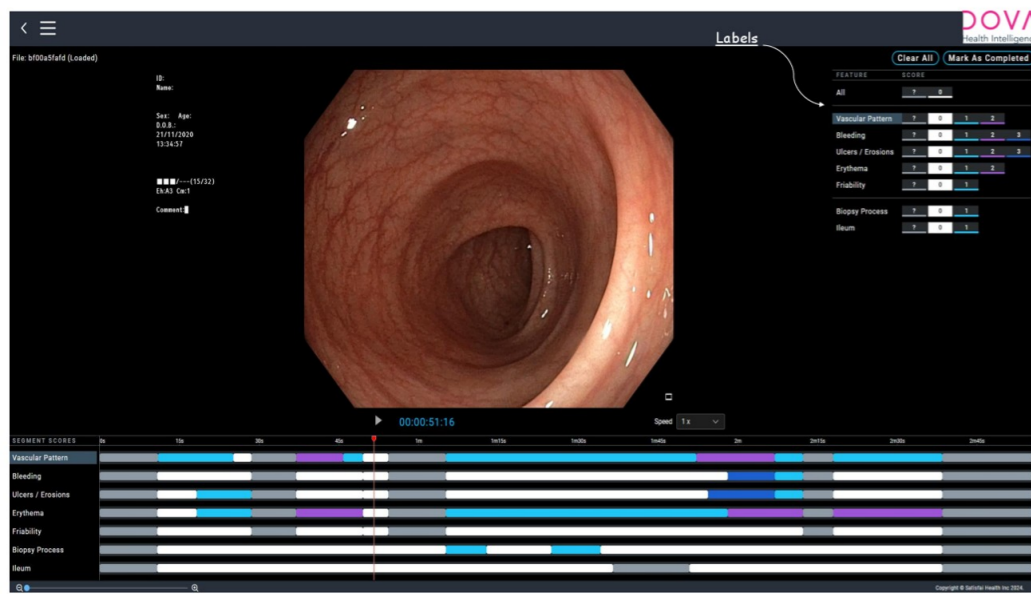
1. Vancouver General Hospital, 2. Dova Health Intelligence (previously Satisfai) 3. Hospital Clinic de Barcelona, 4. St. Paul's Hospital, 5. University of Calgary, 6. University of Oxford, 7. University of Minnesota, 9. University of British Columbia, 10. Center for Advanced IBD Research and Treatment, 11. University of Nottingham, 12. Sinai Health System, 13. Alimentiv Inc.



Background

- AI is increasingly used in UC clinical trials to assess disease activity.
- DovaVision™ is an AI model developed using the Modified Mayo Endoscopic Score (MES) criteria.
- Aim: Improve reproducibility, standardization, and agreement in UC severity assessment.

Methodology



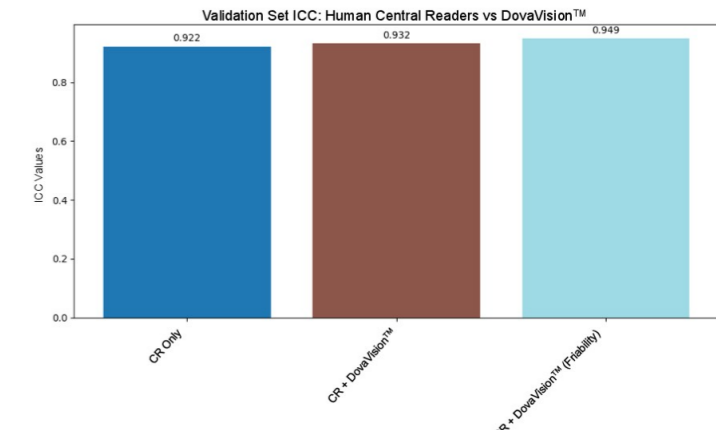
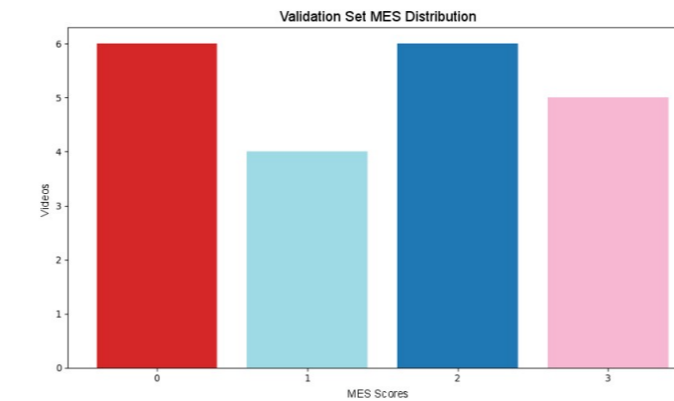
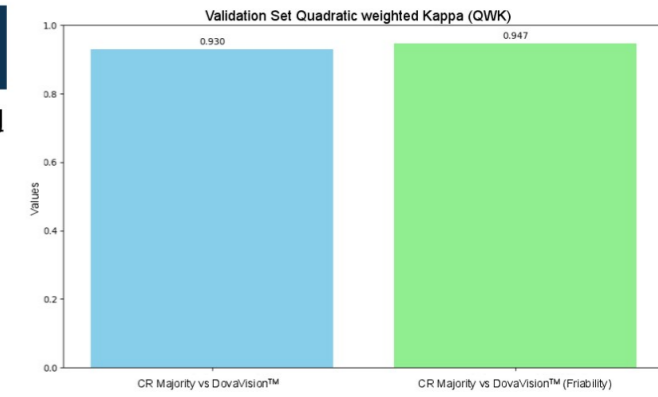
Labeling requirements

- SIA: All videos labeled in Dova's proprietary annotation tool
- Labeler Training: 3 sandbox videos
- Onboarding: Meet ICC/QWK threshold on 10 video set
- Assignments: 20-video batches; 3 labelers/video
- Final label: assigned by majority vote
- Ongoing: labelers maintain ICC/QWK
- 83.8 million individual labels across 7 label categories
- 16.9 million labels after label merging by majority vote

Results

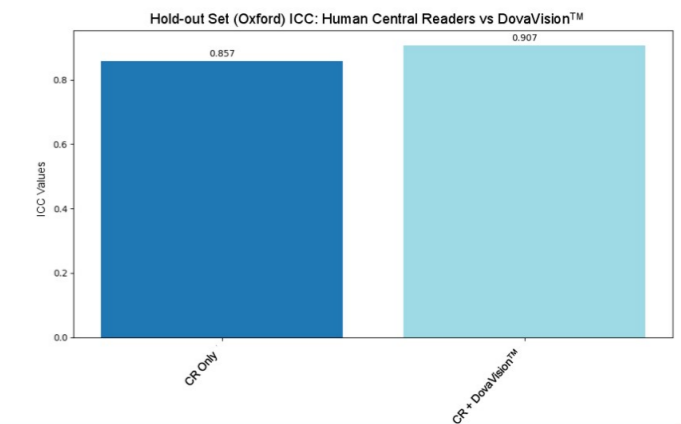
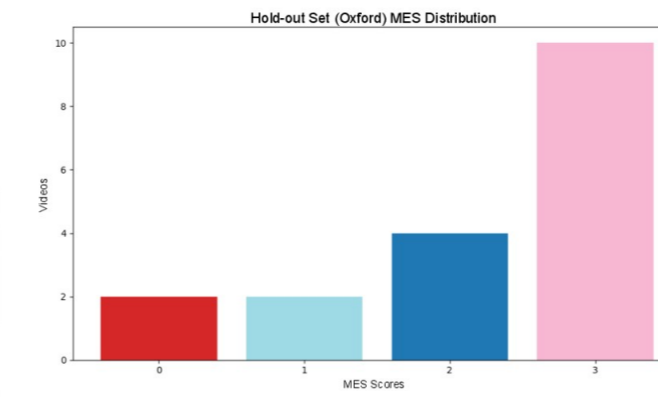
DovaVision™ Validation – Video Level MMES Scoring

- On a validation set of 21 videos, DovaVision™ achieved 0.947 QWK and 0.949 ICC when compared to human central readers
- Central Reader ground truth MMES assigned by 2+1 majority vote (2/3 agreement)
- 1 video in validation set was upgraded from MES1 to MES2 due to Friability which DovaVision™ detected

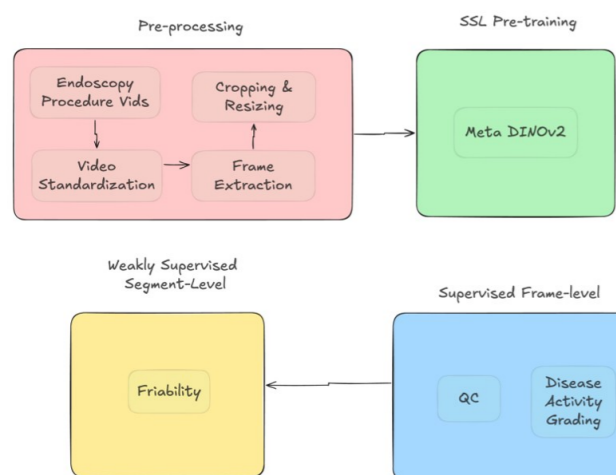


DovaVision™ External Validation

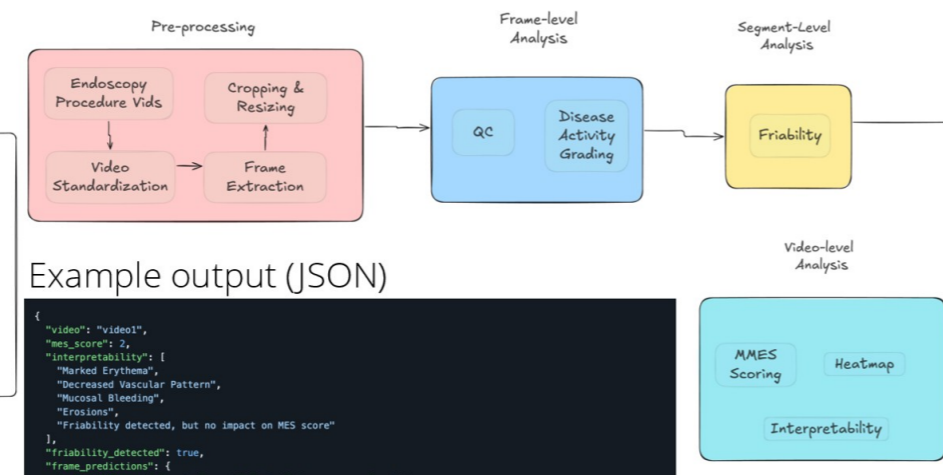
- As a test of our end-to-end inference pipeline, we ran DovaVision™ on a holdout set of 18 videos from a data source not yet included in development
- DovaVision™ achieved 0.88 QWK and 0.907 ICC when compared to human central readers
- In comparison, the pool of 3x central readers assigned / video on this same set showed ICC of 0.857



DovaVision™ Training Pipeline



DovaVision™ Inference Pipeline



Example output (JSON)

```
{
  "video": "video1",
  "mes_score": 2,
  "interpretability": {
    "Marked Erythema",
    "Decreased Vascular Pattern",
    "Mucosal Bleeding",
    "Erosions",
    "Friability detected, but no impact on MES score"
  },
  "friability_detected": true,
  "frame_predictions": {
    "qc": [{"frame_id": 0, "scorable": 0, "biopsy_process": 0}],
    "qc": [{"frame_id": 0, "erythema": 1, "vascular_pattern": 2, "bleeding": 0, "ulcers_erosions": 1}]
  }
}
```

Conclusion

- DovaVision™ enhances reliability, speed, and consistency in UC scoring meeting the modified MES requirements for UC assessment.
- Future applications include AI as a standalone reader with human sign-off and AI integrated into a 2+1 reading model.