

A point-based objective scoring tool using surgical videos for grading difficulty levels of laparoscopic cholecystectomy

Yuchen Luo

frank.luo@student.unimelb.edu.au

The University of Melbourne

Yuen Chiu Kang

Northern Health

Ankur Sidhu

Northern Health

Mani Suleiman

Northern Health

Nezor Houli

Northern Health

Siobhan C McKay

Northern Health

Wanda Stelmach

Northern Health

Russell Hodgson

The University of Melbourne

Research Article

Keywords: Video assessment, Laparoscopic cholecystectomy, Surgical education, Difficulty grading

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10647709/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Purpose

This study aims to develop a standardised grading scale for case difficulty based on objective, visually observable criteria, thereby enhancing its utility for broader surgical education and research purposes.

Methods

Eighty de-identified LC videos were reviewed by three expert hepatopancreatobiliary surgeons. Spearman correlation analysis was used to eliminate redundant variables, and multivariate regression identified independent predictors of "severe" difficulty. Points were assigned based on odds ratios to create a weighted scoring instrument, which was validated using Receiver Operating Characteristic curves.

Results

Interrater agreement demonstrated five criteria achieving 'excellent' agreement and four showing 'good' agreement, with no criteria classified as 'poor'. Significant independent predictors for 'severe' difficulty included 100% adhesion burying gallbladder (OR 29.3, $p = 0.016$), inability to grasp the gallbladder (OR 21.3, $p = 0.029$), and necessity for decompression (OR 59.7, $p < 0.001$). Cases classified as more difficult correlated with higher incidence of wall thickening (86.7% vs 5.6%, $p < 0.001$), exudate/gangrene (60% vs 0, $p < 0.001$), impacted stones (26.7% vs 0, $p < 0.001$), and inability to grasp (46.7% vs 0, $p < 0.001$). The scoring tool achieved an ROC AUC of 0.83 at a cutoff point of 3 for predicting 'severe' cases and an ROC AUC of 0.76 at a cutoff point of 1 for identifying 'moderate' cases.

Conclusions

This study established a statistically weighted, point-based scoring instrument for grading LC difficulty. By providing an objective and reproducible framework based on intraoperative video features, the tool facilitates standardised performance auditing, and the potential for future integration with automated artificial intelligence grading algorithms.

Introduction

Laparoscopic cholecystectomy (1) represents a frequently performed surgical procedure; however, the level of difficulty associated can vary significantly. The operation may be routinely performed employing standard techniques by a trainee surgeon with a relatively less complex case. Conversely, certain cases may present greater challenges, necessitating the expertise of experienced consultant surgeons. (2) Given the variability in difficulty, it becomes increasingly complex to evaluate operative performance. Implementing a stratified operative difficulty framework via a criterion-based scoring rubric offers significant advantages for fairer assessment, standardised technical auditing, and research. Furthermore, this methodology may catalyze the integration of advanced technologies including computer vision and artificial intelligence, thereby enabling the development of sophisticated algorithms

and real-time surgical guidance systems. Such advancements offer potential to improve precision, offering profound advantages for both clinical practice and surgical research. (3)

The current available scoring scales incorporate an extensive array of data encompassing preoperative demographics, clinical parameters, as well as operative characteristics. (4–6) The four established difficulty grading systems articulated in major publications, namely the Nassar, Sugrue, Parkland, and Cuschieri grading scales, are predominantly centred on intraoperative findings with relevance of case difficulty assessment to the gallbladder's appearance, ease of mobilization, dissection, and excision, and they also cover a comprehensive range of aspects pertinent to the management of the surgical presentation. (2, 7–9) Although such instruments provide rigor and standardization for assessment and research, there remains a need for objective evaluation tools based exclusively on intraoperative visual data, which could potentially yield more standardised and reproducible results for larger scale use in assessment and research purposes.

The criteria that delineate grading outcomes as either easy or difficult remain a contentious area of research, and existing studies have employed a diverse range of outcome measures, including operative time, conversion to open surgery, length of hospital stay, as well as morbidity and mortality rates. (10–13) However, many of these variables are influenced by numerous confounding factors, such as the operator's work experience, skill set, their actual performances, and issues related to equipment failure and perioperative ward-based management. (9)

A key strength of this study lies in its focus on objective, video-based features allowing for assessment by non-surgical specialists with no knowledge of data outside of the video. This study aims to identify these factors with high interrater agreement and achieve greater predictive values for a rater-defined case difficulty grade outcome, which paves the way for standardised implementation across surgical education and research, while significantly advancing the potential for future AI-driven automated applications. (14, 15)

Material and methods

This study constitutes a cohort study conducted at Northern Health, located in Melbourne, Australia. Ethics was approved by the Austin Health high-risk research and ethics committee (LNR/18/Austin/142A).

In total, eighty de-identified video recordings were procured with informed consent. Laparoscopic cholecystectomy in the American approach (Four ports with patient in supine position) was included, and French approach (Four port with patient in lithotomy position) was excluded. The videos were subsequently reviewed and edited to double speed by a single consultant surgeon. Three consultant HPB surgeons were blindly tasked with the assessment of these videos. Prior to rating, each rater received training through two demonstrative videos that were excluded from the study. Each video was appraised by two out of the three raters.

The research team, through an extensive literature review, also informed by its pilot phase publication, established a definitive set of variables for the assessment of case difficulty levels derived from four established grading systems reported in the literature: the Nassar, Cuschieri, Parkland, and Sugrue scales, and instead combining all four major scales and focusing primarily on real-time video features. (2, 7–9, 14) Following this, the cases were subjectively classified as ‘mild’, ‘moderate’, and ‘severe’ in relation to their complexity levels. (Supplementary Fig. 1)

Statistical analysis

Percent agreement was utilized for these binary assessment items to effectively evaluate concordance. Statistical analyses were executed using R version 4.3.1. An internationally recognized guideline for studying and reporting interrater reliability was applied in this investigation. (16) For the comparison of variables, χ^2 or Fisher tests were employed for categorical data. The students’ t-test and Mann-Whitney U test were used for continuous variables, contingent upon their distribution. Univariate and multivariate analyses were performed to elucidate factors influencing positive outcomes. The area under curve (17) the Receiver Operating Characteristic Curve (18) was analysed to ascertain the optimal cutoff point model within the scoring tool developed in this study.

Results

Based on the 80 videos rated, five criteria were identified as ‘excellent’ agreement, four as ‘good’ agreement, while two were categorized under ‘moderate’ reliability. No criteria were classified as exhibiting ‘poor’ reliability among the raters. (Table 1)

Table 1
– Interrater reliability of difficulty rating assessment

Assessment items	Value	Strength
Study phase	Validation phase (80 videos)	
Unrelated adhesion	0.89	Good
Limited view of Hartmann's pouch	0.81	Good
Adhesions burying gallbladder	0.69	Moderate
Thickened wall	0.72	Moderate
Empyema	0.91	Excellent
Gangrene	1.00	Excellent
Intrahepatic gallbladder	0.96	Excellent
Impacted stone	0.88	Good
Unable to grasp	0.98	Excellent
Decompression required	0.99	Excellent
Overall difficulty	0.76	Good
NOTE: Strength of agreement reference: Poor < 0.5, Moderate 0.5–0.75, Good 0.75–0.9, Excellent > 0.9.		

Three consultant HPB surgeons subjectively categorised the difficulty grades as 'mild', 'moderate', and 'severe' for the purposes of correlating against the objective factors. A comparative analysis of these three categories revealed that cases classified as more difficult were correlated with a higher incidence of 'unrelated adhesions' (6.7% in the 'severe' group versus 4.5% in the 'mild' group, $p = 0.004$), 'Adhesions burying 100% of GB' (13.3% in the vs 1.1%, $p < 0.001$), 'wall thickening' (86.7% vs 5.6%, $p < 0.001$), 'exudate/gangrene' (60% vs 0, $p < 0.001$), 'impacted stones' (26.7% vs 0, $p < 0.001$), 'unable to grasp' (46.7% vs 0, $p < 0.001$), and 'decompression required' (66.7% vs 0, $p < 0.001$). (Table 2)

Table 2
– Difficulty level assessment items by comparison

Items	Mild	Moderate	Severe	P-value
Number	89	56	15	
Seniority				< 0.001
Registrar	32 (36.0%)	31 (55.4%)	11 (73.3%)	
Fellow	20 (22.5%)	16 (28.6%)	4 (26.7%)	
Consultant	37 (41.6%)	9 (16.1%)	0 (0%)	
Unrelated adhesion	4 (4.5%)	8 (14.3%)	1 (6.7%)	0.256
Limited view of Hartmann's pouch	3 (3.4%)	11 (19.6%)	2 (13.3%)	0.004
Adhesions burying GB				< 0.001
None	58 (65.2%)	16 (28.6%)	6 (40%)	
1–50%	26 (29.2%)	33 (58.9%)	6 (40%)	
51–99%	4 (4.5%)	6 (10.7%)	1 (6.7%)	
100%	1 (1.1%)	1 (1.8%)	2 (13.3%)	
Wall thickening	5 (5.6%)	24 (42.9%)	13 (86.7%)	< 0.001
Exudate/Gangrene	0 (0%)	9 (16.1%)	9 (60%)	< 0.001
Intrahepatic GB	1 (1.1%)	1 (1.8%)	1 (6.7%)	0.30
Impacted stone	0 (0%)	14 (25.0%)	4 (26.7%)	< 0.001
Unable to grasp	0 (0%)	5 (8.9%)	7 (46.7%)	< 0.001
Decompression required	0 (0%)	10 (17.9%)	10 (66.7%)	< 0.001
GB - Gallbladder				

A Spearman correlation analysis was conducted to elucidate the relationships amongst variables, considering the potential overlap within the assessment domain. The outcomes indicated that two variables, specifically 'wall thickening' and 'exudate/gangrene', were significantly correlated with numerous other assessment items, exhibiting Spearman Correlation coefficients exceeding 0.40. Furthermore, the results illustrated that 'decompression required' was associated with 'unable to grasp' (0.47), and showing a lesser correlation with 'impacted stone' (0.22). The remaining variables demonstrated a relatively independent performance. Consequently, the investigators excluded 'wall thickening' and 'exudate/gangrene' from further linear regression model analysis due to these aforementioned confounding correlations identified. Given that 'impacted stone', 'unable to grasp', and 'decompression required' displayed relatively graded correlations and clinical relevance, e.g., an

impacted stone leads to an inability to grasp and subsequently necessitates decompression, these three items were amalgamated as 'mobilization difficulty' for subsequent analysis. (Fig. 1)

In the context of univariate analysis, the assessment items indicative of a 'severe' difficulty level included 'adhesion to gallbladder' (OR 12.3, $p = 0.021$), 'impacted stone' (OR 5.1, $p = 0.059$), 'unable to grasp' (OR 24.5, $p < 0.001$), and 'decompression required' (OR 27.0, $p < 0.001$). The grouped 'mobilization difficulty' variable also exhibited statistical significance, evidenced by progressively graded odds ratios in predicting 'severe' cases. In compiling all variables for a multivariate analysis aimed at predicting a 'severe' difficulty level, the most effective model identified variables such as 'adhesions to gallbladder' with 100% adhesions burying GB demonstrating an odds ratio (OR) of 29.3 (95% CI 1.56–504.9, $p = 0.016$), alongside grouped mobilization difficulty items such as 'impacted stone' (OR 8.6, $p = 0.098$), 'unable to grasp' (OR 21.3, $p = 0.029$), and 'decompression required' (OR 59.7, $p < 0.001$). (Table 3)

Table 3

– Univariate and multivariate analysis of selected difficulty level rating items in predicting “severe” difficulty level cases

Assessment items	Odds ratio	95% CI lower	95% CI upper	p-value
Univariate analysis				
Unrelated adhesion	0.79	0.04	4.5	0.83
Limited view of Hartmann’s pouch	1.4	0.2	6.0	0.65
Adhesions burying GB (none)	1.0			
1–50%	1.3	0.4	4.2	0.7
51–99%	1.2	0.06	8.3	0.85
100%	12.3	1.3	119.6	0.021
Wall thickening	26	6.7	172.3	< 0.001
Exudate/Gangrene	22.7	6.8	82.7	< 0.001
Intrahepatic GB	5.1	0.23	56.7	0.19
Impacted stone	3.4	0.86	11.5	0.059
Unable to grasp	24.5	6.5	101.4	< 0.001
Decompression required	27	8.1	102.4	< 0.001
Mobilisation difficulty (Grouped)				
1. Impacting stone	5.2	0.24	46.04	0.18
2. Unable to grasp	13.8	0.59	155.7	0.043
3. Decompression required	41.3	10.8	208.6	< 0.001
Multivariate analysis (Best performing model)				
Adhesions burying GB				
1–50%	0.73	0.16	3.09	0.66
51–99%	1.09	0.04	14.5	0.95
100%	29.3	1.56	504.9	0.016
Mobilisation difficulty				
Impacting stone	8.6	0.37	104	0.098
Unable to grasp	21.3	0.81	324.4	0.029
CI – Confidence Interval; GB – Gallbladder				

Assessment items	Odds ratio	95% CI lower	95% CI upper	p-value
Decompression required	59.7	13.1	438.0	< 0.001
CI – Confidence Interval; GB – Gallbladder				

Subsequent to the identification of the variables that demonstrated optimal performance in the multivariate regression model, points were assigned to these variables predicated on the power and comparison based on their odds ratios derived from both univariate and multivariate analyses. For instance, 'Adhesions burying 100% of GB' was associated with an OR of 29.3, whereas the mobilization difficulty subcategory 'decompression required' exhibited an odds ratio of 59.7; consequently, the investigators designated 3 points for the former and 6 points for the latter. (Table 4) This novel scoring system demonstrated an ROC AUC of 0.83 at a cutoff point of 3 in predicting 'severe' cases, as well as a ROC AUC of 0.76 at a cutoff point of 1 in predicting "moderate" cases. (Fig. 2)

Table 4
– Scoring system proposed for assessing difficulty level of laparoscopic cholecystectomy

Assessment items	Ratio based on odds ratio	Max points
Adhesions burying gallbladder		
None	0	0 point
1–50%	1	1 point
51–99%	2	2 points
100%	3	3 points
Mobilisation difficulty		
None	0	0 point
Impacting stone	2	2 points
Unable to grasp	4	4 points
Decompression required	6	6 points
Total points		0–9 points

Discussion

This study highlighted the evidence of using intraoperative findings based on videos features accessible to anyone, correlating to a rater-defined difficulty grading outcome. This study successfully developed an objective scoring system of laparoscopic cholecystectomy complexity offering potential for many future purposes in fields including education, research and adoption of AI technology.

Although this study builds its assessment variables from the four established grading systems reported in the literature, it focused primarily on real-time video features. (2, 7–9) These four major frameworks incorporated several non-operative and potentially confounding factors. For instance, the Sugrue scale included body mass index (BMI) and previous abdominal surgery, positing that these elements determine the ‘ease of access’ to the gallbladder. (2) Likewise, the Cuschieri scale accounted for liver cirrhosis with portal hypertension, reflecting increased technical challenge and procedural risks. (3) The Surgical Cholecystectomy Score by Kruthika S et.al., employed time as a direct proxy for difficulty in identifying the cystic artery and duct, rather than treating time solely as an outcome of complexity. (19) In contrast to these existing instruments, our study introduced a more objective evaluation tool grounded exclusively in intraoperative visual findings on video, and it vigorously investigated the interrater reliability of each video-based feature and excluded correlating assessment items utilising statistical methods. The high interrater agreement also underscores the instrument’s robust reliability across expert assessors, effectively justifying the reproducibility of our study outcomes and its transition into an automated AI framework. These findings not only enable broader scalability but also secure the reliability and reproducibility critical for rigorous educational and research outcomes.

Identifying the assessment items represents a fundamental step; however, elucidating the determinants of the final ‘mild’, ‘moderate’, and ‘severe’ difficulty levels is equally critical for defining the relationships between variables and outcomes. Although conversion to open surgery remains a frequently employed surrogate for difficulty grading, its incidence has become exceedingly low in contemporary practice. (3, 9, 19) Moreover, institutional cultures and individual surgeons’ risk tolerance vary considerably and directly influence conversion decisions. (9) Postoperative parameters, including length of hospital stay, 30-day complication rates, and mortality, have likewise been utilised as proxies for difficulty assessment. (3) Nevertheless, multiple confounding factors may intervene, among which variability in operative performance constitutes the predominant source of bias, as an accomplished surgeon may render a technically demanding case appear straightforward. (20–23) Additional influences, such as equipment availability, assistant inexperience, and intraoperative teaching or communication process, may further prolong operative time. (2) These conventional proxies, however entrenched in surgical audit, are structurally incapable of isolating technical difficulty from the confounding influence of operator skill, institutional culture and perioperative circumstance. By contrast, direct video-based features, if identified, are accessible to everyone and, if identified accurately, less prone to assessment bias.

Allocation of points to a scoring scale forms the most fundamental process of this study. Predominantly, established scoring frameworks have been constructed via expert consensus, wherein point allocations are determined according to clinical insights. (2–4, 6, 9, 24–26) In contrast to conventional approaches, this study employs statistical methodology involving linear regression analysis for weighting, Spearman correlation to eliminate redundant factors, and interrater reliability testing for agreement and reproducibility. By standardizing the operative context, this evidence-based difficulty grading scale enables the independent evaluation of technical skills, thereby serving as a foundational bridge to performance assessment. (14) However, the process of quantifying metrics and identifying intraoperative landmarks, despite its reliability and objectivity, imposes considerable time demands on

assessors and generates substantial administrative burden. The automation of these assessments and difficulty classifications could substantially alleviate such issues, offering any surgical doctor a uniform, data-driven, and evidence-based insights into their technical performance or development. With ongoing progress in computer vision and artificial intelligence, it is plausible that deep learning algorithms could be trained to identify the factors drawn by human assessment in this study, thereby automating the grading process with enhanced precision and conferring notable benefits to surgeons and researchers in the near future. (23, 27)

Limitations

Since the assessment focused on video recordings, difficulty assessments excluded procedures outside the peritoneal cavity, omitting the body positioning, laparoscopic port insertion, and intraoperative cholangiogram. Additionally, the sample size of 80 videos is modest compared to certain multi-centre cohort studies, and the difficulty grading outcomes were drawn from expert surgeons' subjective conclusion instead of comparison against using other alternative grading systems, surgeon self-analysis or patient outcomes. (28, 29)

Conclusion

This study introduced a reliable, points-based scoring instrument that classified the difficulty grading of laparoscopic cholecystectomy. It has delivered a consistent, reproducible technique for evaluating surgical difficulty through distinct video-based features. The findings in this study will enable objective evaluation of different difficulty strata, yielding substantial advantages for performance assessment, standardised education and research, and future AI technological application for larger scale studies.

Declarations

Author Contribution

RH conceived and designed the study. YK, AS collected data. NH, SM, RH performed video assessment. MS performed statistical analysis. YL drafted the manuscript. WS, RH provided senior supervision and contributed to the direction of the project. All authors edited the manuscript and approved the final version for submission.

Acknowledgement:

This research was supported by the Commonwealth through an Australian Government Research Training Program Scholarship [DOI: <https://doi.org/10.82133/C42F-K220>] The corresponding author is also the recipient of the Northern Health Foundation and Division of Surgery stipend scholarship.

Word count: 2050 (body), 244 (abstract)

References

1. Ege B, Dinc T, Yildiz BD, Balci Z, Bozkaya H (2015) Utility of endoscopy for diagnosis of barrett in a non-Western society: endoscopic and histopathologic correlation. *Int Surg* 100(4):720–725
2. Sugrue M, Sahebally SM, Ansaloni L, Zielinski MD (2015) Grading operative findings at laparoscopic cholecystectomy- a new scoring system. *World J Emerg Surg* 10:14
3. Griffiths EA, Hodson J, Vohra RS, Marriott P, Katbeh T, Zino S et al (2019) Utilisation of an operative difficulty grading scale for laparoscopic cholecystectomy. *Surg Endosc* 33(1):110–121
4. Vivek MA, Augustine AJ, Rao R (2014) A comprehensive predictive scoring method for difficult laparoscopic cholecystectomy. *J Minim Access Surg* 10(2):62–67
5. Randhawa JS, Pujahari AK (2009) Preoperative prediction of difficult lap chole: a scoring method. *Indian J Surg* 71(4):198–201
6. Gupta N, Ranjan G, Arora MP, Goswami B, Chaudhary P, Kapur A et al (2013) Validation of a scoring system to predict difficult laparoscopic cholecystectomy. *Int J Surg* 11(9):1002–1006
7. Hanna GB, Shimi SM, Cuschieri A (1998) Randomised study of influence of two-dimensional versus three-dimensional imaging on performance of laparoscopic cholecystectomy. *Lancet* 351(9098):248–251
8. Madni TD, Leshikar DE, Minshall CT, Nakonezny PA, Cornelius CC, Imran JB et al (2018) The Parkland grading scale for cholecystitis. *Am J Surg* 215(4):625–630
9. Nassar AHM, Hodson J, Ng HJ, Vohra RS, Katbeh T, Zino S et al (2020) Predicting the difficult laparoscopic cholecystectomy: development and validation of a pre-operative risk score using an objective operative difficulty grading system. *Surg Endosc* 34(10):4549–4561
10. Bourgouin S, Mancini J, Monchal T, Calvary R, Bordes J, Balandraud P (2016) How to predict difficult laparoscopic cholecystectomy? Proposal for a simple preoperative scoring system. *Am J Surg* 212(5):873–881
11. Siddiqui MA, Rizvi SAA, Sartaj S, Ahmad I, Rizvi SWA (2017) A Standardized Ultrasound Scoring System for Preoperative Prediction of Difficult Laparoscopic Cholecystectomy. *J Med Ultrasound* 25(4):227–231
12. Ashfaq A, Ahmadieh K, Shah AA, Chapital AB, Harold KL, Johnson DJ (2016) The difficult gall bladder: Outcomes following laparoscopic cholecystectomy and the need for open conversion. *Am J Surg* 212(6):1261–1264
13. Dutch Pancreatitis Study G, da Costa DW, Schepers NJ, Bouwense SA, Hollemans RA, van Santvoort HC, et al. Predicting a 'difficult cholecystectomy' after mild gallstone pancreatitis. *HPB*. 2019;21(7):827 – 33
14. Luo Y, Kang YC, Sidhu A, Suleiman M, Houli N, Kumta S et al (2026) Video-Based Assessment of Operative Competency in Laparoscopic Cholecystectomy: A Prospective Observational Study to Inform Rubric Design for AI-Assisted Evaluation. *ANZ J Surg* 96(1–2):102–108

15. Luo Y, Kumta S, Stelmach W, Hodgson R (2026) An Australian perspective of using video for the assessment of laparoscopic surgery and support for artificial intelligence in performance evaluation. *Langenbeck's Archives of Surgery*
16. Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 15(2):155–163
17. Cao Y, Manias E, Omonaiye O, Beauchamp A, Jessup R, Oldenburg B et al (2025) Associations between health literacy, well-being and general health in adults with type 2 diabetes, considering culturally and linguistically diverse background: Results from the first diabetes MILES Australia study. *Diabet Med* 42(10):e70090
18. M. K. Population Growth in Australia: The Australia Institute; (2015) [Available from: <https://australiainstitute.org.au/wp-content/uploads/2020/12/Population-Paper-FINAL.pdf>]
19. Asuri SK, Singh K, Prakash D, Rai O, Ramachandran S (2024) Surgical Cholecystectomy Score (SCS) for grading the difficulty of laparoscopic cholecystectomy. *Langenbecks Arch Surg* 409(1):203
20. Bouarfa L, Schneider A, Feussner H, Navab N, Lemke HU, Jonker PP et al (2011) Prediction of intraoperative complexity from preoperative patient data for laparoscopic cholecystectomy. *Artif Intell Med* 52(3):169–176
21. Tongyoo A, Liwattanakun A, Sriussadaporn E, Limpavitayaporn P, Mingmalairak C (2024) New Proposed Classification of Difficulty in Laparoscopic Cholecystectomy. *J Laparoendosc Adv Surg Tech A* 34(5):407–414
22. Lipman JM, Claridge JA, Haridas M, Martin MD, Yao DC, Grimes KL et al (2007) Preoperative findings predict conversion from laparoscopic to open cholecystectomy. *Surgery* 142(4):556–563 discussion 63 – 5
23. Ward TM, Hashimoto DA, Ban Y, Rosman G, Meireles OR (2022) Artificial intelligence prediction of cholecystectomy operative course from automated identification of gallbladder inflammation. *Surg Endosc* 36(9):6832–6840
24. Asai K, Iwashita Y, Ohyama T, Endo I, Hibi T, Umezawa A et al (2022) Application of a novel surgical difficulty grading system during laparoscopic cholecystectomy. *J Hepatobiliary Pancreat Sci* 29(7):758–767
25. Asai K, Ohyama T, Watanabe M, Moriyama H, Kujiraoka M, Watanabe R et al (2024) Validation of a surgical difficulty grading system in laparoscopic cholecystectomy for acute cholecystitis. *J Hepatobiliary Pancreat Sci* 31(2):80–88
26. Schrenk P, Woisetschläger R, Rieger R, Wayand WU (1998) A diagnostic score to predict the difficulty of a laparoscopic cholecystectomy from preoperative variables. *Surg Endosc* 12(2):148–150
27. Orimoto H, Hirashita T, Ikeda S, Amano S, Kawamura M, Kawano Y et al (2025) Development of an artificial intelligence system to indicate intraoperative findings of scarring in laparoscopic cholecystectomy for cholecystitis. *Surg Endosc* 39(2):1379–1387
28. Kanakala V, Borowski DW, Pellen MG, Dronamraju SS, Woodcock SA, Seymour K et al (2011) Risk factors in laparoscopic cholecystectomy: a multivariate analysis. *Int J Surg* 9(4):318–323

29. Kama NA, Kologlu M, Doganay M, Reis E, Atli M, Dolapci M (2001) A risk score for conversion from laparoscopic to open cholecystectomy. Am J Surg 181(6):520–525

Figures

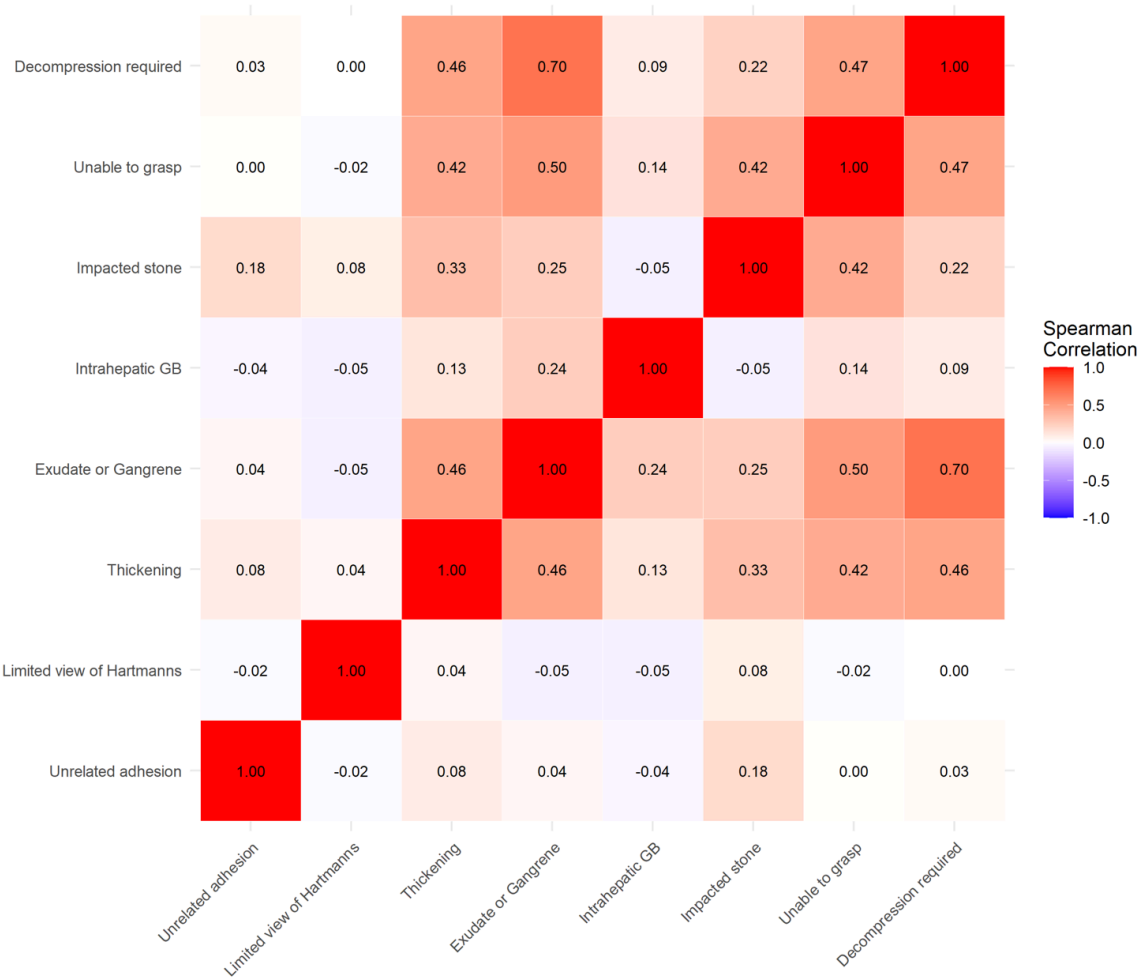


Figure 1

Spearman correlation heatmap of difficulty level rating items.

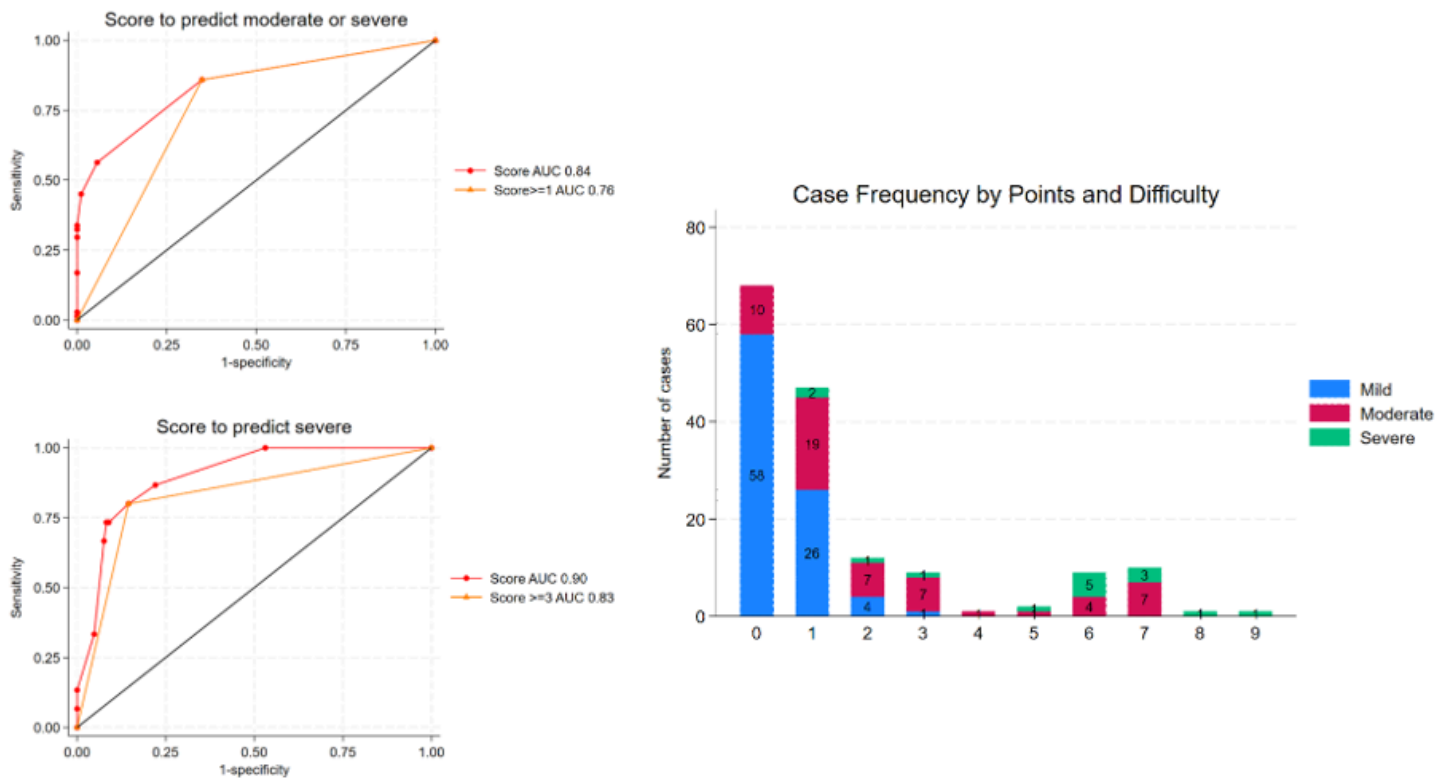


Figure 2

Receiver Operating Characteristic (ROC) curve of the developed scoring system with Area Under Curve (AUC) 0.76 at cutoff point of 1 in predicting moderate cases, and ROC AUC 0.83 at cutoff point of 3 in predicting severe cases.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.docx](#)
- [Supplementarymaterial.docx](#)