

Reviewer #1

Specific comments to authors

This is an useful study that proves that patients with acute biliary pancreatitis are hidden among patients with idiopathic acute pancreatitis. Please clarify whether the biliary sludge and microlithiasis visualized by this EUS are in the gallbladder or common bile duct, or only in the common bile duct. If only the gallbladder is visualized, it is difficult to prove whether it is true acute pancreatitis due to cholelithiasis. Therefore, only those visualized in the common bile duct should be examined.

RESPONSE TO REVIEWER 1:

The authors thank the reviewer for this comment that is highly relevant to clinical practise. Biliary sludge and microlithiasis were only evaluated as pancreatitis aetiologies if concrement evidence was detected in the gallbladder and common bile duct or only in the common bile duct by endosonography. Two cases of the validation cohort with only detection of sludge in the gallbladder were evaluated only as sludge-induced pancreatitis and included in the study due to concomitant elevated liver values above three times the norm. Other possible causes of pancreatitis were additionally excluded and thusso biliary aetiology was suggested in this case presenting with gallbladder sludge and correspondingly high liver values [1,2]. We would also like to thank the reviewer for the generally positive feedback on our study.

The corresponding text in the manuscript section "Participants" was supplemented as follows (page 5, line 23/24): "218 patient cases with acute pancreatitis and endosonography were then further stratified into a cohort (47 patients) with no other cause of pancreatitis than endosonographically detected biliary microconcrements (biliary sludge/microlithiasis; **detection of concrements in the common bile duct or gallbladder**



and common bile duct) and 171 patients with other causes of AP (Figure 1)."

[1] Zarnescu N, Costea R, Zarnescu (Vasiliu) E, et al. Clinico-biochemical factors to early predict biliary etiology of acute pancreatitis: age, female gender, and ALT. J Med Life 2015; 8: 523–526 [PMID: 26664483 PMCID: PMC4656965]

[2] Phillip V, Huber W, Hagemes F, et al. Incidence of acute pancreatitis does not increase during Oktoberfest, but is higher than previously described in Germany. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc 2011; 9: 995-1000.e3 [PMID: 21723238 DOI: 10.1016/j.cgh.2011.06.016]



Reviewer#2

Specific comments to authors

Manuscript ID: 86626 Title: A machine-learning based decision tool selecting patients with idiopathic acute pancreatitis for endosonography to exclude a biliary etiology

RESPONSE TO REVIEWER 2:

First of all, we thank the reviewer for the detailed review of our study and would like to provide a point-to-point response below.

Introduction

 What is the significance of accurately identifying the etiology of acute pancreatitis, particularly in cases classified as idiopathic? How does the lack of a clear definition for biliary sludge and microlithiasis pose challenges in assessing their role as causes of acute pancreatitis?

A diagnosis of idiopathic pancreatitis has direct diagnostic and prognostic consequences for patients compared to a diagnosis of biliary pancreatitis (due to sludge and microlithiasis). Patients whose aetiology remains labelled idiopathic have a higher risk of pancreatitis recurrence compared to the cohort of biliary pancreatitis (12% vs. 25%) [1]. Therefore, the identification of a treatable aetiology, such as biliary pancreatitis, is of high clinical relevance for patients. Until 2023, there was no uniform consensus definition of the terms sludge and microlithiasis. In our systematic review published in GUT this year, we were able to show that in around 20% of cases, the terms sludge and microlithiasis were even used synonymously in the literature. Therefore, these inconsistently used definitions made a qualitatively accurate risk attribution in the context of pancreatitis impossible until now [2]. We have incorporated this information into the manuscript on page 4, line 7 to 11.



[1] Hajibandeh S, Jurdon R, Heaton E, Hajibandeh S, O'Reilly D. The risk of recurrent pancreatitis after first episode of acute pancreatitis in relation to etiology and severity of disease: A systematic review, meta-analysis and meta-regression analysis. J Gastroenterol Hepatol. 2023 Jun 27. doi: 10.1111/jgh.16264. Epub ahead of print. PMID: 37366550.

[2] Żorniak M, Sirtl S, Beyer G, Mahajan UM, Bretthauer K, Schirra J, Schulz C, Kohlmann T, Lerch MM, Mayerle J; LMU Microlithiasis Expert Survey Team. Consensus definition of sludge and microlithiasis as a possible cause of pancreatitis. Gut. 2023 Apr 18:gutjnl-2022-327955.

2. What are the current guidelines and recommendations for the management of idiopathic acute pancreatitis, particularly in cases suspected to be caused by biliary sludge and microlithiasis? How does the development of a predictive tool using a machine learning-based approach contribute to the decision-making process and potential interventions?

In principle, endosonography is recommended in cases of pancreatitis labelled as idiopathic to detect/exclude biliary origin, pancreas divisum or other anatomical alterations. Alternatively, or in the case of inconclusive findings MRI/MRCP examination should be performed according to the German pancreatitis as well as international guidelines. Robust data has been published showing that the use of endosonography can detect a cause of pancreatitis in 60% of idiopathic pancreatitis, half of which are biliary (gallstones, microlithiasis, sludge) induced. The currently ongoing PICUS trial in the Netherlands will provide information on the exact timing of endosonography to increase the detection rate [3-5].



[3] Beyer G, Hoffmeister A, Michl P, et al. S3-Leitlinie Pankreatitis – Leitlinie der Deutschen Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS) – September 2021 – AVVMF Registernummer 021-003. Z Für Gastroenterol 2022; 60: 419–521 [PMID: 35263785 DOI: 10.1055/a-1735-3864].

[4] Umans DS, Rangkuti CK, Sperna Weiland CJ, et al. Endoscopic ultrasonography can detect a cause in the majority of patients with idiopathic acute pancreatitis: a systematic review and metaanalysis. Endoscopy 2020; 52: 955–964 [PMID: 32557477 DOI: 10.1055/a-1183-3370].

[5] Umans DS, Timmerhuis HC, Hallensleben ND, et al; Dutch Pancreatitis Study Group. Role of endoscopic ultrasonography in the diagnostic work-up of idiopathic acute pancreatitis (PICUS): study protocol for a nationwide prospective cohort study. BMJ Open. 2020 Aug 20;10(8):e035504.

3. How does the proposed machine learning tool, based on routine laboratory values, assist clinicians in estimating the probability of the presence of biliary sludge and/or microlithiasis in patients with acute pancreatitis? What are the potential benefits of this tool in reducing the need for unnecessary endosonography and guiding the selection of appropriate interventions, such as cholecystectomy or biliary sphincterotomy?

Using the Microlithiasis Prediction Score, clinicians can predict a biliary etiology caused by microlithiasis or sludge in the CBD at the time of presentation to the emergency department or in the first days of hospitalisation related to pancreatitis with a positive prediction of 75.7 and a negative prediction of 78.5 %. Through this score-based selection of patients, patients formally subsumed in the cohort of idiopathic pancreatitis patients



can be directed to a more targeted use of endosonography to establish the diagnosis of biliary pancreatitiy. Once the diagnosis of biliary pancreatitis has been made, recurrence can be prevented by cholecystectomy.

Methods

a. Study design

1.What were the specific inclusion criteria used to select the patient cohorts for this retrospective study of acute pancreatitis?

The selection criteria used for study inclusion are listed in the "participants" section (p. 5, line 10 continued). Patients with acute pancreatitis and endosonography during the pancreatitis-associated stay were included. The clustering was then performed in the group with evidence of microlithiasis/sludge and a group without evidence of sludge/microlithiasis (see Fig.1 and Supplement Fig.1).

Corresponding section on patient selection:

"Only patients who met the diagnostic criteria for acute pancreatitis as defined in the APA/IAP guidelines and adapted into the German S3 guideline were enrolled in the analysis [9], [10]. The first classifier used was whether patients received an EUS during their initial hospital stay, reducing the number of patients for further analysis to 360. The endosonographies were each performed by an experienced endoscopist. In the majority (79%) of pancreatitis stays, EUS was performed on days 1-3. Of the 360 patients with endosonography, a total of 142 cases were excluded from further analysis due to incomplete records or missing coding. The 218 patient cases with acute pancreatitis and endosonography were then further stratified into a cohort (47 patients) with no other



cause of pancreatitis than endosonographically detected biliary microconcrements (biliary sludge/microlithiasis) and 171 patients with other causes of AP (Figure 1). In the two study groups (acute pancreatitis + EUS: 47 x microlithiasis versus 171 x nonmicrolithiasis (other cause; Supplement Fig. 1), history, alcohol consumption, sonography, ERCP or EUS findings, start or change of existing medication, known hereditary pancreatitis (available genetic tests of the most prevalent susceptibility genes), and laboratory findings (lipase levels, IgG subclasses, liver enzymes, triglyceride and calcium level (corrected for blood serum albumin level) were retrospectively evaluated. In the context of laboratory value analyses, the values from the first blood analysis after admission of the respective patient stay were used in each case. The aim was to select patients in whom microlithiasis / sludge was likely to subject them to EUS to reduce the number of EUS as an invasive, expensive, and burdened with complications procedure. To independently validate our machine-based algorithm we obtained identical clinical data and inclusion criteria from two high-volume German pancreas centres (University hospital of the Technical University Munich: 22 x Microlithiasis AP, 51 x Other AP; University Medical Centre Göttingen: 14 x Microlithiasis AP, 30 x Other-AP; Supplement Fig. 1)."

b. Participants

 What were the diagnostic criteria used to identify patients with acute pancreatitis for inclusion in the study? Were these criteria based on the APA/IAP guidelines and the German S3-Guideline?

Exactly, we used the diagnostic algorithm listed in the German S3 guideline on pancreatitis. Acute upper abdominal pain (often but not necessarily radiating to the back), an increase in serum lipase to at least three times the norm, and an image-morphological



pancreatitis correlation are considered diagnostic criteria for acute pancreatitis, whereby 2 of the 3 criteria must be met for the diagnosis to be made.

The corresponding passage in the manuscript section "Participants" reads as follows (p. 5, line 13-15): "Only patients meeting the diagnostic criteria of acute pancreatitis as set in the APA/IAP guidelines and adapted in the German S3-Guideline were enrolled in the analysis"

[3] Beyer G, Hoffmeister A, Michl P, et al. S3-Leitlinie Pankreatitis – Leitlinie der Deutschen Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS) – September 2021 – AWMF Registernummer 021-003. Z Für Gastroenterol 2022; 60: 419–521 [PMID: 35263785 DOI: 10.1055/a-1735-3864].

2. How were the patients stratified into the two study groups (microlithiasis and nonmicrolithiasis)? What were the specific criteria used to classify a patient as having microlithiasis or another cause of acute pancreatitis?

The diagnosis of microlithiasis versus no microlithiasis (Other AP) was based on the result of endosonography. Only if endosonographic evidence of microlithiasis or sludge could be obtained the patient was assigned to the microlithiasis-pancreatitis group. Since endosonography is the gold standard for detecting sludge and microlithiasis, we relied on this endoscopic procedure.

[2] Żorniak M, Sirtl S, Beyer G, Mahajan UM, Bretthauer K, Schirra J, Schulz C, Kohlmann T, Lerch MM, Mayerle J; LMU Microlithiasis Expert Survey Team. Consensus definition of sludge and microlithiasis as a possible cause of pancreatitis. Gut. 2023 Apr 18:gutjnl-2022-327955.



3.Can you provide more information about the retrospective evaluation of patient data? What specific variables were assessed, and how were they evaluated in relation to microlithiasis or other causes of acute pancreatitis?

As listed in point 1, the patients were classified into Microlithiasis AP and Other AP groups based on the endosonography findings. For all patients, all laboratory values and images obtained during the pancreatitis stay were screened to avoid misclassification. All relevant laboratory values obtained are listed in Table 1. Specifically, for the cohort of patients labelled idiopathic, IgG4 levels were also screened and evaluated along with existing images to avoid overlooking occult aetiologies such as autoimmune pancreatitis. In no case we detected elevated IgG4 values and/or imaging signs of autoimmune pancreatitis. Based on this very detailed analysis of patient documents, we achieved a robust classification of the two study groups.

c. Test methods

1. How were the baseline variables filtered and selected for inclusion in the machine learning-based predictor model? Were any specific criteria applied to determine the variables' relevance and impact on the prediction of microlithiasis?

We thank the reviewer for raising this question. We would like to clarify that, while creating a machine learning predictor model, we employed multi-step procedures for selecting baseline variables, such as:

(1) Data Pre-processing: From all the collected variables, we removed the variables



which have zero variance and which have near-zero variance.

- (2) Feature Engineering: Feature engineering involves creating new variables or transforming existing ones to better represent the underlying patterns in the data. To achieve this, we classified all the numeric variables into three categories, such as within physiological limit (WL), above upper physiological limit (ULN), and below lower physiological limit (LLN), based on clinical reference limits. We retained the categorical variable as is.
- (3) Feature Selection: aims to identify the subset of features that contribute the most to the predictive power of the model while reducing the risk of overfitting, we employed h2o automl algorithm, which explore multiple machine learning algorithms (in our case we used: generalised linear regression (GLM), random forest (RF), decision tree, Gradient boosting machines (GBM), Extreme gradient boosting (XGBoost), StackedEnsembles, and Deep Neural Network (Deep Learning)). We selected 20 models per algorithm. Of all the selected models, the model with the minimum logloss and a higher area under the curve has been used for feature selection. Top-ranking features with more than 0.05 variable importance were retained for further iterative modelling.
- (4) Model Building and Validation: The top-ranking features were subject to iterative h2o.automl model (with the same analogy used in step 3) using the training cohort. Models are trained in 80% of the training cohort and tested in 20% of the training cohort to assess their predictive performance. The performance of the model is evaluated using appropriate metrics (such as accuracy, precision, recall, logloss, AUC, PPV and NPV).

2.Can you provide more details on the machine learning methods used in the study? What specific machine learning algorithms were employed, and how were their parameters optimized during the training process?



H2O.ai's AutoML utilises a variety of algorithms under the hood to perform its tasks. We use the following algorithms:

- 1. Generalized Linear Models (GLM)
- 2. Random Forest
- 3. Gradient Boosting Machines (GBM)
- 4. XGBoost: Extreme Gradient Boosting (XGBoost)
- 5. Stacked Ensembles
- 6. Deep Learning
- 2. How was the performance of the predictor model assessed and evaluated? What measures or metrics were used to determine the model's accuracy and predictive capabilities? Additionally, could you provide information on the external validation dataset and how it was utilized to validate the final predictive model?

We trained the model on 80% of the training cohort and assessed the performance of the final model on the remaining 20% of the test cohort using the following metrices:

- 1. Accuracy
- 2. Precision (positive predictive value)
- 3. Recall (sensitivity)
- 4. F1 score
- 5. Logloss
- 6. Area under ROC curve (AUC) and
- 7. Confusion Matrix

To reduce overfitting, we used a 10-fold cross-validation. Regarding the external validation dataset, it was a separate data set that was not used during the training or



hyperparameter tuning process. This data set served as an independent test of the performance and generalisation of the model. It ensured that the models were not biased due to overfitting to the training data. The external validation data set was evaluated using the same metrics mentioned above, and the results were compared to the performance on the test set used during model development.

Results

- a. Microlithiasis predictive score results of the identification cohort
 - 1. How were the patients in the identification cohort categorized into the microlithiasis/sludge cohort versus the Other-AP cohort? Were specific diagnostic criteria or imaging techniques used to determine the presence of microlithiasis/sludge in the biliary system?

Categorisation was based on the individual endosonography findings of each patient. Due to the lack of a harmonised and internationally agreed definition for the terms sludge and microlithiasis at the time of the study, the respective endoscopy findings were adopted and not reevaluated. At all centres, endosonography was performed exclusively by advanced endoscopy experts. Endosonography is the gold standard for the diagnosis of sludge and microlithiasis.

- 2. Could you provide more information on the variables used in the machine learningbased microlithiasis prediction algorithm? How were these variables measured and what significance did they have in predicting the presence of microlithiasis/sludge in acute pancreatitis patients?
- 3. What were the performance metrics used to evaluate the accuracy and predictive



capabilities of the ML-based microlithiasis prediction algorithm? Can you provide a more detailed explanation of how sensitivity, positive predictive value (PPV), negative predictive value (NPV), and specificity were calculated and interpreted in the context of the study results?

All performance metrics were obtained from the H2o automl command h2o.performance command

(https://docs.h2o.ai/h2o/latest-stable/h2o-docs/performance-and-prediction.html).

In the context of the study, true positives (TP) are cases where the algorithm correctly predicted the presence of microlithiasis, true negatives (TN) are cases where the algorithm correctly predicted the absence of microlithiasis, false positives (FP) are cases where the algorithm incorrectly predicted microlithiasis, and false negatives (FN) are cases where the algorithm failed to predict microlithiasis when it was actually present.

Sensitivity was calculated using:

'Sensitivity = True Positives / (True Positives + False Negatives)"

In the study, a high sensitivity ensured that microlithiasis cases were not missed.

The Positive Predictive Value (PPV) (precision) was calculated using the formula: 'PPV = True Positives / (True Positives + False Positives)"

In the context of the study, a high PPV would mean that the algorithm's positive predictions are reliable and not likely to be false alarms.

The negative predictive value (NPV) was calculated using: 'NPV = true negatives / (True Negatives + False Negatives)"

A high NPV would indicate that the algorithm is good at ruling out microlithiasis when it is truly absent.

Specificity was calculated using: 'Specificity = true negatives / (True Negatives + False



Positives)"

High specificity would mean that the algorithm is good at minimising false positives, ensuring that cases without microlithiasis are not incorrectly identified as positive.

- b. Microlithiasis predictive score validation cohort
- How were the patients in the validation cohort selected and categorized into the microlithiasis AP and Other-AP groups? Were the inclusion criteria and diagnostic methods consistent with those used in the identification cohort?

The patient categorisation was carried out according to the same criteria as in the identification cohort in order to methodically ensure high quality.

2. Can you provide more information on the automated machine learning (autoML) process used for the iterative reduction of variables and model optimization in the validation cohort? What specific algorithms or techniques were employed in the autoML approach?

We thank reviewer for raising this query. We explained the complete flow of h2o automl modeling in answer to Question 1 in section c methods. Since external validation was employed to establisg the performance of model, we did not perform model optimisation using the validation cohort.

3. The sensitivity and specificity values for the microlithiasis predictive score in the validation cohort are reported as 0.96 and 0.31, respectively. Can you discuss the implications of these values in terms of accurately predicting the presence of



microlithiasis in acute pancreatitis patients? How do these performance metrics contribute to the overall utility and reliability of the microlithiasis predictive score?

Our score helps to select patients for EUS with a high sensitivity and a very high negative predictive value and thus will reduce costs and complications of unnecessary EUS exams as well as allows selecting patients for further treatment to prevent recurrence of biliary pancreatitis at the time of presentation in the emergency department. The robustness of the model is shown in the alluvial plot in **Figure 3** with only 3 out of 81 patients being misclassified as microlithiasis and not as other-AP, corresponding to the discretely higher NPV (compared to the PPV) in the validation cohort (**Table 3**). The applicability for everyday clinical practice in the context of our current study is therefore based primarily on the very high negative predictive power. Within the framework of the currently planned prospective score validation, the aim will be to reduce the number of score variables without losing negative predictive power.



Reviewer#3

Specific comments to authors

This is a carefully done study and the findings are of considerable interest. The authors "present a robust and validated machine learning-based predictor model consisting of routinely recorded parameters at admission that can predict biliary sludge and microlithiasis as cause of acute pancreatitis". This article provides a research method for establishing a good machine and network based etiology prediction model, and obtains good experimental results through the validated pattern. Furthermore, an explanation of following questions should be pointed. 1. What is the role of this model in clinical treatment? 2. In addition to the Cohort study, has the accuracy of this model been verified in a randomized controlled study? 3. Has Ig4 been tested on every patient as an excluded diagnosis? I would be very glad to re-review the paper in greater depth once it has been edited because the subject is interesting.

RESPONSE TO REVIEWER 3:

First of all, we would like to thank the reviewer for the very positive feedback on our study. With regard to the three questions listed, we would like to add the following comments:

1. The clinical relevance of our machine-learning based algorithm for patient selection biliary (sludge/microlithiasis) versus idiopathic pancreatitis is based on the fact, that by using our prediction model, patients are assigned a treatable aetiology by using endosonography and confirming the diagnosis of biliary pancreatitis not requiring additional follow-up. Idiopathic pancreatitis patients are also more likely to suffer from recurrent pancreatitis (20 % risk) compared to biliary acute pancreatitis, making



the diagnostic fork in the road (biliary versus idiopathic) a very relevant one for pancreatitis patients [1-4].

[1] Hajibandeh S, Jurdon R, Heaton E, Hajibandeh S, O'Reilly D. The risk of recurrent pancreatitis after first episode of acute pancreatitis in relation to etiology and severity of disease:
A systematic review, meta-analysis and meta-regression analysis. J Gastroenterol Hepatol.
2023 Jun 27. doi: 10.1111/jgh.16264. Epub ahead of print. PMID: 37366550.

[2] Balakrishnan V, Unnikrishnan AG, Thomas V et al. Chronic pancreatitis. A prospective nationwide study of 1,086 subjects from India. JOP 2008; 9: 593–600.

[3] Conwell DL, Banks PA, Sandhu BS et al. Validation of Demographics, Etiology, and Risk Factors for Chronic Pancreatitis in the USA: A Report of the North American Pancreas Study (NAPS) Group. Dig Dis Sci 2017; 62: 2133–2140 [354]

[4] Levy P, Barthet M, Mollard BR et al. Estimation of the prevalence and incidence of chronic pancreatitis and its complications. Gastroenterol Clin Biol 2006; 30: 838–844

- **2.** A score validation within the framework of a prospective randomised trial has not yet taken place, but is currently in preparation.
- **3.** Autoimmune pancreatitis as a rare pancreatitis aetiology entity in European countries occurred as a confirmed diagnosis in only one case in our pancreatitis cohort (1/171; 0.6 %; see Supplement Figure 1). Of the 117 acute pancreatitis patients classified as idiopathic at the three centres, 52% had IgG subclasses measured (all without evidence of IgG4 level elevation). In terms of image morphology, there were no indications of the presence of autoimmune pancreatitis type 1 or 2 either in the patients with IgG4 determination or in the idiopathic acute pancreatitis patients without IgG determination. Due to the retrospective nature of the study, we were unable to



increase the rate of specific IgG4 levels, but were able to exclude the presence of a relevant rate of occult autoimmune pancreatitis due to the multiparameter diagnosis of autoimmune pancreatitis (ICDC criteria; [5]).

The Supplement Figure 1 Legend was supplemented accordingly:

Supplement Figure 1: Distribution of non-microlithiasis (Other-AP) patients according to underlying etiologies by respective centre. Listed are the respective non-biliary microconcrement-triggered acute pancreatitis of the patients who received an endosonography in the course of the diagnostic work-up, located under the encroachment "Other-AP". IgG4 levels were determined in 52% of idiopathically classified AP patients, each without evidence of elevation suggestive of underlying autoimmune pancreatitis. Imaging evidence of autoimmune pancreatitis was accordingly not found in the idiopathic-AP cohort.

[5] Shimosegawa **T**, Chari ST, Frulloni L et al. International consensus diagnostic criteria for autoimmune pancreatitis: guidelines of the International Association of Pancreatology. Pancreas 2011; 40: 352–