Retrospective Study
Machine-learning based decision tool selecting patients with idiopathic acute pancreatitis for endosonography to exclude a biliary etiology

Sirtl S et al. Microlithiasis prediction score for AP
Abstract

BACKGROUND

Biliary microlithiasis/sludge is detected in approximately 30% of patients with idiopathic acute pancreatitis (IAP). As recurrent biliary pancreatitis can be prevented, the underlying etiology of IAP should be established.

AIM

To develop a machine-learning based decision tool for the use of endosonography (EUS) in pancreatitis patients to detect sludge and microlithiasis.

METHODS

We retrospectively used routinely recorded clinical and laboratory parameters of 218 consecutive patients with confirmed AP admitted to our tertiary care hospital between 2015 and 2020. Patients who did not receive EUS as part of the diagnostic work-up and whose pancreatitis episode could be adequately explained by other causes than biliary sludge and microlithiasis were excluded. We trained supervised machine learning (ML) classifiers using H₂O.ai automatically selecting the best suitable predictor model to predict microlithiasis/sludge. The predictor model was further validated in two independent retrospective cohorts from two tertiary care centers (117 patients).

RESULTS

Of 28 categorized patients’ variables recorded at admission were identified to compute the predictor model with an accuracy of 0.84 [95% confidence interval (CI): 0.791-0.9185], positive predictive value of 0.84 and negative predictive value 0.80 in the identification cohort (218 patients). In the validation cohort, the robustness of the prediction model was confirmed with an accuracy of 0.76 (95%CI: 0.673-0.8347), positive predictive value of 0.76 and negative predictive value of 0.78 (117 patients).

CONCLUSION
We present a robust and validated ML-based predictor model consisting of routinely recorded parameters at admission that can predict biliary sludge and microlithiasis as cause of AP.

**Key Words**: Acute pancreatitis; Idiopathic acute pancreatitis; Biliary pancreatitis Microlithiasis; Sludge; Endosonography


**Core Tip**: Occult biliary lithiasis represents the largest moncausally treatable aetiology group within idiopathic acute pancreatitis cases. The identification of this subgroup protects patients from pancreatitis recurrences and over- or underdiagnosis. Based on 28 easy to collect and widely available patient variables, a machine-learning based prediction score can be used to predict the presence or absence of biliary sludge or microlithiasis in the context of pancreatitis hospitalisation. We provide a web-based prediction tool to select patients for endosonography to investigate microlithiasis or sludge as cause of pancreatitis and treat them accordingly.

**INTRODUCTION**

Pancreatitis is a high incidence disease and the underlying cause for the highest number of patients admitted to hospital admission of all benign gastrointestinal-disorders[1]. In approximately 25% of patients with acute pancreatitis (AP), etiology cannot be established during the first episode of pancreatitis[2,3]. If etiology of AP cannot be identified by history, laboratory chemistry and imaging pancreatitis is classified as “idiopathic” [idiopathic AP (IAP)]. Unclassified or idiopathic pancreatitis represents the third largest group of pancreatitis and is therefore of great importance from both a
medical and a socioeconomic point of view requiring thorough work up\textsuperscript{[3,4]}. All efforts should be made to elucidate a treatable etiology to prevent further episodes of AP. A recent meta-analysis has shown that biliary etiology is the most common cause of idiopathic pancreatitis with a prevalence of 30\%\textsuperscript{[5]}. Specifically, in light of morbidity and mortality of AP it is crucial to differentiate the potentially treatable etiology of AP triggered by biliary sludge and microlithiasis from idiopathic or other causes of AP. Unfortunately, due to a lack of unifying definition of biliary sludge and microlithiasis it is currently impossible to assess the risk of sludge and/or microlithiasis as cause of AP. In the absence of clear evidence guideline suggest to treat those patients with cholecystectomy and maybe biliary sphincterotomy. In line, the diagnostic IAP workup requires excluding biliary microconcrements as it is believed that detection and concrement removal and/or cholecystectomy can prevent further episodes of pancreatitis in over 85\% of cases\textsuperscript{[6,7]}. To facilitate decision making whether the patient should be referred to endosonography (EUS) followed by endoscopic retrograde cholangiopancreatography (ERCP) or cholecystectomy, we developed a predictive tool using a machine learning (ML)-based approach to estimate the probability of the presence of biliary sludge and/or microlithiasis at the time of presentation to the emergency department. The ML tool, which is based on routine laboratory values, will help clinicians to enrich the likelihood to detect microlithiasis or sludge at admission on EUS and hereby reduce the number of EUS exams in presumed acute idiopathic pancreatitis.

**MATERIALS AND METHODS**

**Study design**

We retrospectively studied 1340 confirmed and hospitalized patient cases of AP treated at LMU University Hospital Munich (tertiary care hospital) between 01/01/2015 and 01/10/2020 (ICD-10 codes used: K85.00-K85.91). Patient cohorts with identical inclusion criteria from the University hospital of the Technical University Munich and the
University Medical Center Goettingen served as validation cohort. The study was conducted in accordance with the updated STARD guideline of 2015[^8].

**Participants**

Only patients meeting the diagnostic criteria of AP as set in the APA/IAP guidelines and adapted in 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 EUS, a total of 142 cases were excluded from further analysis due to incomplete records or missing coding. 218 patient cases with AP and EUS 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). In the two study groups (AP + EUS: 47 × microlithiasis versus 171 × non-microlithiasis (other cause; Supplementary Table 1), history, alcohol consumption, sonography, ERCP or EUS findings, start or change of existing medication, known hereditary pancreatitis (available genetic testing of most prevalent susceptibility genes), and laboratory findings (lipase levels, immunoglobulin G subclasses, liver enzymes, triglyceride and calcium level (corrected for blood serum albumin level) were retrospectively evaluated. In the context of the laboratory value analyses, the values from the first blood analysis after admission of the respective patient stay was used in each case. The aim was to select patients in which 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 centers (University Hospital of the Technical University Munich: 22 × microlithiasis AP, 51 × other AP; University Medical
Center Goettingen: 14 × microlithiasis AP, 30 × other-AP; Supplementary Table 1). The definitions of the entities “biliary sludge” and “biliary microlithiasis” were taken from the endoscopic reports during the retrospective data evaluation and were not re-evaluated due to the current lack of an accepted unifying definition. Due to the differences between the participating centers in the use and partial equation of the two terms biliary sludge and microlithiasis, sludge-triggered pancreatitis was subsumed as biliary AP caused by microlithiasis.

**Test methods**

All aspects of data reporting, predictive modeling, and validation reporting were performed in accordance with the TRIPOD guidelines\(^{11}\). A diagnostic reference standard for laboratory or imaging-based prediction of biliary sludge or microlithiasis in the context of AP has not yet been published. To derive the ML-based predictor model (index test), the following steps were performed (Figure 2): (1) Baseline variables (n = 192) were filtered leaving out variables with zero and near zero variance; (2) All numeric variables were classified into within limit, above upper limit, and below lower limit, based on clinical reference limits. All categorised variables were retained; (3) The training cohort was divided into a training (80%) and a test set (20%). Endpoint balancing was achieved by stratifying the classes by inducing the sampling rate of patients with microlithiasis and reducing the sampling rate of patients with other-AP. ML was performed based on all filtered baseline variables and data from the training set, resulting in a predictor based on all variables (base predictor model); and (4) To improve robustness and interpretability, low-impact variables were iteratively removed. An iterative predictive model with a reduced number of variables (n = 26) was obtained based on the performance in the test set.

All predictor models were constructed using the H2O.ai platform (https://www.h2o.ai) selecting (with h2o.automl) the best suitable ML method in the training set. The parameters of each method were optimized by employing an internal ten-fold cross-validation on the training set. The optimal method was then applied to
the test set to assess the final performance. In each loop, the best performing predictor model was identified from all predicted outcomes obtained using the performance measure logloss. Variables with a higher proportion of missing data (> 25% missing data) were also not excluded per se in order to base the final model on the broadest possible number of routinely available variables in the early phase of AP. The iterative predictive model obtained was externally validated in an independent retrospective dataset.

**Statistical analysis**

All data processing, modeling and assessment of performances was done using R [version 4.0.4 (2021-02-15, “Lost Library Book”)] and visualized in R-studio (version 1.3.9.59). No unique algorithm was developed for this study. All data R scripts or functions are available online at the following link: https://github.com/mayerlelab/microlithiasisPredict. P-values of < 0.05 were considered statistically significant if appropriate for the tests used.

**RESULTS**

**Microlithiasis predictive score - results of the identification cohort**

Between January 1, 2015 and October 1, 2020, 218 patients with AP received an EUS during their initial admission with AP at LMU University Hospital meeting the study inclusion criteria (Figure 1). In 47 of 218 pancreatitis patients, no causal pancreatitis etiology other than endosonographically detected biliary microconcretions/sludge was found during the respective inpatient stay. 171 out of 218 pancreatitis patients with EUS, 52.6% (90/171) were classified as ‘idiopathic’, 21.6% (37/171) as acute on chronic and 15.2% (27/171) with macrolithiasis as of biliary etiology (Supplementary Table 1). Mean age in the microlithiasis/sludge cohort was 59.1 (SD 18.8) in comparison to AP patients of other etiologies with 54.6 (SD 17.1; \( P = 0.122 \)) years. Gender distribution was not statistically different in both cohorts, with a male predominance in both cohorts (31/47; 66% of microlithiasis patients) and 103/171 (60.2%; \( P = 0.475 \)); Table 1. 76.6% of
microlithiasis-AP patients were assessed as mild pancreatitis cases according to the revised Atlanta classification [36/47; 19.1% moderate (9/47), 4.3% severe (2/47)]. In the other-AP cohort, 71.9% of patients were assessed as mild pancreatitis cases according to the revised Atlanta classification [123/171; 25.7% moderate (44/171), 2.3% severe (4/171)]. A total of 29 variables from serum samples and 5 from urine were used to develop the ML-based microlithiasis prediction algorithm. All variables listed corresponded to the values measured at admission for each individual pancreatitis inpatient (see Table 1 for list of variables used). To move from the base ML to the iterated ML model, weighting was done taking scale variance into account. For the LMU identification cohort, age, triglycerides, sodium, glutamic pyruvic transaminase, erythrocytes, potassium, thyrotropin, protein (total) and leukocytes in descending order were of greatest importance predicting microlithiasis/sludge. Using the iterated learner-based model an accuracy of 0.8361 [95% confidence interval (CI): 0.791-0.9185; odds ratio = 20.88 (95%CI: 2.08-209.27)] with a sensitivity of 97.92% and a positive predictive value (PPV) of 83.93% could be achieved for the prediction of microlithiasis as trigger of pancreatitis [net present value (NPV) = 0.80; specificity: 0.31; Table 2).

**Microlithiasis predictive score - validation cohort**

Data from two large-volume university pancreas centers were used for score validation. In total, a validation cohort of 56 patients with microlithiasis and 81 non-microlithiasis AP patients were retrieved from the clinical data base at the University Hospital of the Technical University Munich (22 × microlithiasis AP, 51 × other AP) as well as the University Hospital Göttingen (14 × microlithiasis AP, 30 × other-AP; Figure 1, Table 3). In the Technical University Munich cohort, the group of other-AP patients was mainly alcohol-related [31/51 (60.8%)], while in the Göttingen cohort biliary macrolithiasis was held responsible for the majority of AP patients [16/33 (53.3%)]. Idiopathic etiology was named as the second most frequent etiology group in both external cohorts with app. 30% each [Technical University Munich 17/51 (33.3%), Göttingen 10/33 (33.3%)]; Supplementary Table 1. Microlithiasis patients in the validation cohort were on average
60.1 (SD 18.4) years old, patients from the other-AP cohort were respectively 55.3 (SD 16.8) years old. In both groups (microlithiasis + other-AP), the majority of patients were male [24/36 (66.7%) and 46/81 (56.8%), respectively] resembling the identification cohort. 63.9% of microlithiasis-AP patients were assessed as mild pancreatitis cases according to the revised Atlanta classification [23/36; 27.7% moderate (10/36), 8.3% severe (3/36)]. In the other-AP cohort, 59.2% of patients were assessed as mild pancreatitis cases according to the revised Atlanta classification [48/81; 27.2% moderate (22/81), 12.5% severe (11/81)]. Using automated ML, the best-fitting model for iterative reduction of variables was used to achieve external validation of the microlithiasis predictive score using the optimized iterative ML model. For the validation cohort, based on the variables ordered by scaled importance in Figure 3, an accuracy of 0.7607 (95% CI: 0.673-0.8347), a PPV of 0.7573 and a NPV of 0.7857 was achieved (sensitivity: 0.96, specificity: 0.31; Table 2). 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 2).

**DISCUSSION**

Previous and more recent studies on idiopathic pancreatitis still report a proportion of idiopathic pancreatitis stably at 20%-30%[12,13]. However, it has been suspected for decades and is increasingly supported by evidence that a large proportion of pancreatitis patients classified primarily as idiopathic actually suffer from a biliary etiology and that detecting these patients during the first episode of pancreatitis is restricted due to the lack of availability of timely and high quality EUS exams[14]. Furthermore, there is a lack of reliable data on when during an inpatient stay of an IAP-labeled patient an EUS could detect biliary microconcrements as triggers for pancreatitis without causing an unnecessary burden for the patient through overdiagnosis. This is an important questions as we know from Oría et al[15] that CBD stones usually pass within 48 h suggesting that microconcrements might even pass
more rapidly and might not be detected on EUS. Prospective data generated by the study that reported a variance of EUS-based biliary concrement detection rates of 19% in the low risk group, but 58% in the moderate risk group and 50% in the high risk group (grouping according to ASGE recommendation[16]). Risk stratification in terms of pre-test probability for EUS use < 48 h after hospital admission to rule in or out the presence of biliary concrements is warranted before intervention to overcome lack of availability and reduce costs and side effects[12]. Diagnostic evaluation is complicated by the fact that biliary microconcrements could be a coincidental finding in the context of pancreatitis-induced gallbladder hypomotility, and therefore must always be understood in the individual patient’s setting, taking into account a PPV of a biliary pancreatitis origin greater than 85% with elevation of the alanine transaminase (ALT) above three times the upper limit of normal[17]. However, no causally effective drug for pancreatitis therapy is available in 2022 and the detection of causally remediable pancreatitis causes such as biliary microlithiasis or sludge will continue to play a decisive role in the prevention of further pancreatitis attacks. The efficacy of cholecystectomy in the cohort of IAP patients was shown in a meta-analysis with a recurrence rate of 11% compared to 38.9% in conservatively treated patients (risk ratio 0.41; 0.16-1.07)[18]. Our ML-based approach of predicting biliary microlithiasis and sludge should therefore be understood as an approach to make up for the lack of evidence from prospective studies on the optimal timing of EUS in IAP patients as this score is based on widely available laboratory values and can be used to determine the probability of the presence of biliary microconcrements at admission. Our score helps to select patients for EUS with a high sensitivity and a very high negative predictive values and thus will reduce costs and complications of unnecessary EUS exams as well as allow to subject patients to further treatment to prevent recurrence of biliary pancreatitis at the time of presentation in the emergency department. Preliminary work on ML-based algorithms and prediction models in the context of AP has focused on severity assessment and prediction of complication[15]. A multicenter retrospective study used an auto-ML-based approach to predict pancreatitis severity, comparable to
our ML approach, achieved an area under the curve (AUC) > 0.90 in the GBM model with specificity and accuracy > 0.95 in the early detection of patients with a subsequently severe course of pancreatitis[19]. Outperforming clinically established non-ML-based scoring systems such as BISAP or Ranson underlying the relevance of ML approach over an educated guess[20,21]. ML-based prediction scores with regard to biliary microconcrements have not yet been published. Non-ML-based multivariate logistic regression models using widely available laboratory values have previously shown that an ALT level more than three times above the norm at patients’ admission [specificity of 82%, sensitivity of 60%, receiver operating characteristic (ROC)-AUC 0.733; \( P < 0.001 \)] and an age > 69.5 years (specificity 92%, sensitivity of 57%, ROC-AUC 0.759; \( P < 0.001 \)) act as the best predictors of biliary etiology[17,22]. Here, our ML-based prediction score achieves higher sensitivity values (96.30%), whereby ALT and, above all, age also rank 4th and 1st in the weighting of our score, thus confirming the existing evidence in the area of non-ML laboratory value-based prediction of biliary etiology of pancreatitis (Figure 3). Contrary to previously published studies on laboratory-based prediction of biliary pancreatitis etiology, our prediction tool is based specifically on microlithiasis and sludge and not primarily on gallstones and occult microlithiasis/sludge subsumed in this cohort as in previous studies.

Our study has several limitations: First, the retrospective study approach did not allow us to generate a uniform definition of the two entities microlithiasis and sludge. Even after extensive literature research we were unable to delineate a uniform but distinct definition of biliary microlithiasis and sludge. We thus decided to use the terms as synonyms between the endoscopy centers of the three participating university hospitals. This might impose a significant bias. The macrolithiasis, which was again clearly listed in the endoscopy findings across the universities, ensured quality of EUS. Likewise, the patient cohort declared as other-AP in terms of etiology varied greatly between the participating centers (Supplementary Table 1). Ultimately, this probably reflects the individual diagnostic scope and the question of whether EUS can generate added value in the context of the individual patient. Also, due to the retrospective study
design, no attempt could be made to increase the degree of purity of biliary (microlithiasis and sludge) triggered pancreatitis by uniformly fulfilling laboratory chemistry tests prior to EUS. This resulted in a proportion of patients of 36.6% with, for example, missing calcium values in the laboratory chemistry pancreatitis work-up.

Our study is convincing in presenting for the first time a robust ML-based and externally validated prediction model for pancreatitis patients declared idiopathic early in the diagnostic work-up and may be helpful as a noninvasive decision tool by combining simple and widely used laboratory values to decide for or against EUS. In order to make the Microlithiasis Predictive Score and the relatively high number of underlying variables usable, a user-friendly interface is available online at the following link for use in the research context: https://github.com/mayerlelab/microlithiasisPredict. To illustrate the performance of the Microlithiasis Predictive Score, we designed a graphical user interface for a quick entry of the values of the necessary patient variables, followed by the prediction of the need for EUS. The user friendly-interface (Video core tip, currently not deployed on Web) provides the user with the model-based estimated probability of the patient stratification to microlithiasis/sludge and other-pancreatitis. Moreover, it provides several graphical presentations to illustrate the impact of the specific variables on the decision. A multicenter prospective score validation with harmonised predefinition of biliary sludge and microlithiasis is currently being planned.

CONCLUSION
We present for the first time a ML based tool, externally validated in two sets of data from tertiary pancreatic referral centers, to predict the presence of biliary sludge und microlithiasis in patients with an initial label of idiopathic pancreatitis with an accuracy of 0.7607 (95%CI: 0.673-0.8347), a PPV of 0.7573 and a NPV of 0.7857. Upon prospective validation, the prediction score will aid in decision-making which patient to subject to EUS for diagnostic work-up at a first episode of pancreatitis.
ARTICLE HIGHLIGHTS

Research background
About 30% of acute pancreatitis (AP) cases classified as idiopathic actually have a biliary and thus mono causally treatable origin.

Research motivation
To date, there is no predictive score to differentiate between idiopathic and sludge- and microlithiasis-triggered acute biliary pancreatitis. Undiagnosed biliary pancreatitis aetiology poses the risk of overdiagnosis and additional patient burden. AP triggered by small biliary concrements (microlithiasis and sludge) is a particularly challenging diagnosis.

Research objectives
The aim of this study was to develop a machine-learning based prediction score for the presence of microlithiasis and sludge in AP patients. External score validation was performed at two university pancreas centres.

Research methods
The clinical and laboratory parameters of 218 AP patients were used to calculate a machine-learning based prediction model for the presence of sludge and microlithiasis. 47 patients with endosonographic evidence of sludge and microlithiasis (and no other possible underlying pancreatitis aetiology) were used in the identification cohort and compared with 171 AP patients without endosonographic evidence of sludge and microlithiasis. We trained supervised machine learning classifiers using H2O.ai automatically selecting the best suitable predictor model to predict microlithiasis/sludge. An external pancreatitis cohort from two university pancreas centres with 117 patients was used for validation.
The score, constructed from a total of 28 simple variables to be collected in the early phase of pancreatitis-associated hospitalisation and validated externally at two university pancreatic centres, can predict the presence of biliary sludge and microlithiasis with an accuracy of 0.7607 (95% confidence interval: 0.673-0.8347), a positive predictive value of 0.7573 and a net present value of 0.7857.

Research conclusions
For the first time, we present a machine-learning based prediction score to differentiate between sludge- and microlithiasis-triggered AP and idiopathic pancreatitis. By using it in the early phase of pancreatitis-related hospitalisation, patient selection for or against the use of endosonography can support clinical decision-making.

Research perspectives
Upon prospective validation, the prediction score will aid in decision-making which patient to subject to endosonography for diagnostic work-up at a first episode of pancreatitis specifically to differentiate between sludge/microlithiasis triggered and idiopathic AP.
Michał Żorniak, Simon Sirtl, Georg Beyer, Ujjwal Mukund Mahajan et al. "Consensus definition of sludge and microlithiasis as a possible cause of pancreatitis", Gut, 2023