

MINERVA Project

Machine learnINg for the rElapse Risk eValuation in Acute biliary pancreatitis

Code: 202273A4YP

Brief description

Acute pancreatitis is the most common pancreatic disorder, with a global incidence of 34 cases per 100,000 people and over 1.5 million new cases annually. Mortality approaches 1%. Patients with biliary acute pancreatitis (BAP) can be treated either with index cholecystectomy or conservatively. Although current guidelines advise performing cholecystectomy in all hospitalized patients, this approach is often impractical in everyday clinical settings due to competing demands within hospitals. Consequently, individuals initially deemed at “lower risk” are commonly treated conservatively or face extended delays before undergoing cholecystectomy. While conservative management can alleviate symptoms, up to 35% of these patients develop relapse (RAP) within 30 days, frequently necessitating emergency surgery under more critical clinical circumstances. RAP is linked to increased risks of chronic pancreatitis, pancreatic cancer, postoperative complications, and death, along with higher healthcare costs driven by prolonged and repeated hospital stays. Previously, no standardized approach existed to determine which patients were at higher or lower risk of RAP, meaning patient prioritization lacked a defined criterion. Since assessing RAP risk is a crucial outcome that clinicians must evaluate carefully in every BAP patient before choosing a management strategy, this evaluation should be rapid, reliable, accurate, and not require additional costs or procedures.

The MINERVA (*Machine learnINg for the rElapse Risk eValuation in Acute biliary pancreatitis*) project represents the first prospective multicenter study designed to estimate RAP risk and identify its predictive factors. It aims to prospectively gather data from more than 340 patients across multiple Italian centers to develop a machine learning and AI-based predictive score for estimating relapse risk in individuals admitted with mild BAP. The model will undergo both internal and external validation, with the overarching aim of enhancing clinical decision-making. The resulting score is intended to assist clinicians in prioritizing patients, improving waiting list management, and minimizing treatment delays, ultimately helping to prevent emergency scenarios. Specifically, the primary objective of the MINERVA study is to combine prospectively collected data from 13 centers during the project period with nearly 2,000 retrospective patient records, identify predictors of RAP, and create an effective optimized model capable of estimating relapse risk after a first BAP episode in conservatively treated patients, using only data collected at admission. Recurrence will be assessed at 30, 60, and 90 days, as well as at 1-year follow-up, to ensure a thorough evaluation. A novel AI-based risk score is going to be developed and validated, delivering accurate results where previously published traditional statistical models had failed to identify dependable predictive variables. The MINERVA score relies exclusively on patients' demographic and laboratory data, along with other variables that can be easily obtained and recorded at initial admission. It therefore provides clinicians with a validated, standardized tool to guide patient prioritization and estimate relapse risk.

The MINERVA project is carried out through a set of coordinated parallel activities across all participating research units, encompassing data collection, model development, validation, and dissemination.

In particular, the units responsible for carrying out the project activities are:

- Responsible local research unit University of Naples Federico II (UniNa)
- Local Research Unit University of Cagliari (UniCa)
- Local Research Unit University of Campania Luigi Vanvitelli (Vanvitelli)

The recruitment centers list is:

- Unità di Chirurgia Generale, Ospedale Maggiore di Bologna
- Unità di Chirurgia e Terapie Oncologiche Avanzate, Ospedale Morgagni-Pierantoni di Forlì
- IRCCS A.O.U. di Bologna, Policlinico Sant'Orsola Malpighi di Bologna
- Clinica Chirurgica, Università di Trieste
- Chirurgia Generale e d'Urgenza, Ospedale San Paolo di Civitavecchia
- U.O. Chirurgia Generale, Nuovo Ospedale "S.Stefano" di Prato
- U.O. Chirurgia Generale Ospedale San Marco-Zingonia di Bergamo
- Struttura Semplice Dipartimentale Chirurgia d'Urgenza, Arcispedale Sant'Anna di Ferrara
- U.O.C. di Chirurgia Generale e d'Urgenza, Ospedale M.Bufalini di Cesena
- Dipartimento di Scienze della Salute, Università del Piemonte Orientale di Novara
- Unità di Chirurgia Generale, Dipartimento di Chirurgia, AO Annunziata di Cosenza
- U.O.C. di Chirurgia Generale, Ospedale Madonna del Soccorso di San Benedetto del Tronto

Objectives and expected outcomes

Primary aim

The primary aim of the MINERVA project is the prediction of risk relapse of Recurrent Acute Pancreatitis (RAP) in patients after a first episode of mild Biliary Acute Pancreatitis (BAP) according to the 2012 Revised Atlanta Classification. This outcome is reached by the development and validation of a novel risk score.

The MINERVA score is going to be grounded upon a machine learning (ML) model that takes into account patients' demographic and hematological information along with data that can be easily collected and recorded at index patient admission.

Secondary aims

Secondary aims of the MINERVA project are:

- compare the accuracy of the MINERVA machine learning model to other traditional machine learning models previously adopted in literature (such as ANN and SVM) and with statistical models (such as multiple regression)
- apply the MINERVA machine learning model to the prediction of BAP complications

In summary, the MINERVA project aims to reach the following objectives and results:

- 1) Propose a novel methodology for the assessment of the risk of relapse in patients with Biliary Acute Pancreatitis;
- 2) Propose a machine learning predictive model using a deep learning architecture applied to easily collectable data;
- 3) Validate the MINERVA score on an extensive, multicentric, prospective cohort;
- 4) Allow national and international clinicians, medical staff, researchers and the general audience to freely and easily access the MINERVA score computation and use it in their daily clinical practice.

Results

The MINERVA Project objectives were met as described below:

1) Propose a novel methodology for the assessment of the risk of relapse in patients with Biliary Acute Pancreatitis;

This aim was achieved by several consecutive steps. First of all, during the MINERVA project the first prospective study protocol was written and published (<https://doi.org/10.1186/s13017-025-00594-7>) so that it could be shared among all the recruitment centers for the submissions to local ethical committees and for the subsequent recruitment of the patients along with the informed consent. Then, a shared REDCap database was structured and created. REDCap was chosen as it provides an intuitive platform for designing anonymized customizable electronic GDPR-compliant data collection forms. The database included built-in data validation to ensure accuracy, user access controls to protect privacy, and strong security measures such as encryption and audit trails. Data could be easily exported via a Python code pipeline created on purpose. The credentials were shared among all the recruitment centers so that they could add the patients independently. An ad hoc rulebook was provided to all centers specifying the details of each patient variable and data to be inserted and the complete step by step procedure to ensure correctness. After recruitment started, local investigators were closely monitored by the project units through informal check-ins and scheduled six-month meetings to ensure consistent and accurate data entry throughout the follow-up period. In case of missing data or delayed collection, individual centers were notified via email about the patient code fields to be updated. In parallel, a literature review was conducted to identify variables previously associated with the outcome, and the findings were published as an open-access report (<https://doi.org/10.1186/s13017-025-00601-x>). The prospectively collected data (344 patients) were then analyzed to identify potential predictors and exclude variables that might introduce bias. After removing outliers, drop-out cases and lost to follow-up, the prospective multicentric cohort with a recurrence ≤ 90 days or without recurrence comprised 297 patients.

2) Propose a machine learning predictive model using a deep learning architecture applied to easily collectable data and validate the MINERVA score on an extensive, multicentric, prospective cohort;

During the data collection period, using already available retrospective data from the international MANCTRA study (2413) patients, several machine learning models were tested extensively and compared to reach the outcome of developing a risk score expressed in probability that could take into account solely of patients' demographic and laboratory variables and other data that can be easily collected and recorded at index patient admission. The list of features that were included in the final models are the following: as continuous variables the patient's age, BMI, WBC neutrophil, platelet, INR, reactive protein c-values, AST, ALT, total bilirubin, direct conjugated bilirubin, GGT, lipases, LDH; as categorical variables sex, previous episodes of pancreatitis, clinical history of diabetes, clinical history of chronic lung disease, clinical history of ischemic disease, hypertension, atrial fibrillation, chronic kidney disease, disease of the hematopoietic system, immunosuppressive drugs at the time of admission, choledocholithiasis, acute cholangitis, and Endoscopic Retrograde Cholangiography-Pancreatography. To account for the class imbalance problem, several approaches were employed such as oversampling of the minority class by repetition and using the ADASYN algorithm as well as tentative undersampling of the majority class. Imputation of missing data was conducted using multiple imputation (MICE). Merged (retrospective + prospective) and preprocessed data were divided into training (70%), validation (15%) and test sets for external validation (15%), and AUC, F1-score and accuracy were used for evaluation. Models' optimization concerned three different types of algorithms which are briefly summarized below:

- Machine learning models: Four supervised classifiers, i.e. Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), and SVM with RBF kernel (SVM-RBF) were trained. All models reached a training AUC > 0.80.
- Ensemble models: Eight ensemble strategies combined predictions from MLP, LR, RF, and SVM-RBF; all weights were optimised on the validation set, with test metrics computed independently; thresholds were selected by maximising F2-score ($\beta=2$) to prioritize recall over precision in the clinical screening context.
- Deep learning models: Two deep learning architectures were evaluated (MLP and CNN). Both best models of the two architectures reached a training AUC > 0.80, but the best performance of all the models tested was achieved by the CNN (AUC=0.79), confirming the superiority of this architecture.

3) Allow national and international clinicians, medical staff, researchers and the general audience to freely and easily access the MINERVA score computation and use it in their daily clinical practice.

The MINERVA project is one of the few AI-powered clinical studies that allowed the developed model to be transferred easily to the clinical practice thanks to how easy and intuitive the platform user interface is and how little data available at admission it needs to provide a reliable estimate of the risk of relapse. The diffusion of the MINERVA project purposes, platform and results along with the use of the MINERVA risk score was carried out through extensive dissemination activities that were conducted throughout all the project duration. Dissemination activities included:

- the development of a fully dedicated website (<https://www.minervaproject.org/>), including a subscription-based newsletter and integration with social media platforms (X, Facebook, Instagram, and LinkedIn) to disseminate project updates. The website also provides direct access to the MINERVA score platform, enabling free and user-friendly use of the developed risk prediction model <https://minerva-score.onrender.com/>; setting up of a dedicated GitHub repository (<https://github.com/adrianodesimone3/minerva-score>) to share the open-source code of the MINERVA platform, ensuring transparency and accessibility;
- dissemination within the scientific community including presentation of the MINERVA project at national and international conferences, open-access publication of protocol with results made freely available to both the scientific community and the public, promoting reproducibility and broad dissemination of knowledge, with all publications released under open-access policies.

The final MINERVA project workshop represented an incredible opportunity during which it was possible to present a demo of the MINERVA online platform to clinicians and healthcare researchers, to collect relevant and important feedback on its enhancements from stakeholders and to discuss opportunities to improve the platform both considering the aspects pertaining the user interface and the intuitiveness of the platform presentation but also considering the clinical aspects and accessibility of the variables to be considered for the risk assessment model. An important emphasis was put on detailing the easiness of access to each variable when patients are first admitted and evaluated so to make sure that the MINERVA project model only made use of variables easy to obtain at the first visit. Additionally, while the MINERVA risk score is extensively validated using multicentric (prospective and retrospective) national and international data, additional collaborations with international researchers were actively pursued to expand the validation dataset, with the aim of further improving the model's accuracy, robustness, and generalizability in future applications in particular with researchers with previous experience with acute pancreatitis management at the Hospital Clínico Universitario Lozano Blesa and Hospital General Universitario de Alicante.