Journal of Biosciences and Medicines, 2026, 14(8), 110-124
https://www.scirp.org/journal/jbm
ISSN Online: 2327-509X
ISSN Print: 2327-5081

The Prediction of Severe Acute Pancreatitis: From Traditional Scoring Systems to Emerging Artificial Intelligence

Hao Li[1], Chuanxin Zou[1*], Lingyan Hu[2]

[1]Department of Gastroenterology, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou, China
[2]Department of Endocrinology, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou, China
Email: *1714745884@qq.com

How to cite this paper: Li, H., Zou, C.X. and Hu, L.Y. (2026) The Prediction of Severe Acute Pancreatitis: From Traditional Scoring Systems to Emerging Artificial Intelligence. *Journal of Biosciences and Medicines*, **14**, 110-124.
https://doi.org/10.4236/jbm.2026.148011

Received: July 12, 2026
Accepted: August 9, 2026
Published: August 12, 2026

Abstract

Severe acute pancreatitis (SAP) is a severe complication of acute pancreatitis (AP), with a mortality rate approaching 30%, bringing a heavy burden to the global healthcare system. Early identification of high-risk patients within the first 48 hours of admission is crucial for timely implementing intensive care, reasonably triaging patients, and carrying out targeted treatments. In the past, various scoring systems were mainly used in clinical practice to evaluate disease severity, including the Ranson score, Glasgow/Imrie score, Acute Physiology and Chronic Health Evaluation II (APACHE II), and BISAP score. Although the above scoring tools can stratify patients by risk, there are issues such as delayed calculation time and suboptimal sensitivity and specificity. Apart from this, although serological indicators such as blood urea nitrogen, procalcitonin, C-reactive protein, and interleukins can provide certain prognostic reference value, the effectiveness of assessing each indicator individually varies significantly. In addition, imaging assessment methods represented by the Balthazar CT severity index can provide detailed information on organ morphology, but this examination requires a delayed scanning time to obtain optimal imaging results, which also limits its clinical application value. In recent years, a variety of advanced machine learning algorithms—including Extreme Gradient Boosting, Random Forests, Support Vector Machines, and Artificial Neural Networks—have developed rapidly. These models can simultaneously integrate patients' demographic information, serological test results, and imaging data, and they demonstrate consistently stable and excellent performance in overall prognosis prediction, with AUC values generally exceeding 0.90. Therefore, this paper systematically reviews the development of predictive models for severe acute pancreatitis, from traditional scoring systems to emerging machine learning models, summarizing the advantages and disad-

vantages of each method one by one, while also discussing the challenges that still need to be overcome before such machine learning tools can be effectively applied in clinical practice.

Keywords

Severe Acute Pancreatitis, Prediction, Scoring Systems, Biomarkers, CT Severity Index, Artificial Intelligence, Machine Learning, XGBoost, Random Forest

1. Introduction

Acute pancreatitis (AP) is an inflammatory disorder of the pancreas with an annual global incidence ranging from 13 to 45 cases per 100,000 population, and its incidence has been steadily increasing over the past three decades [1] [2]. Although the majority of AP episodes are self-limiting and resolve within one week, approximately 15% - 20% of patients progress to severe acute pancreatitis (SAP), which is characterized by persistent organ failure lasting beyond 48 hours [3].

According to the Revised Atlanta Classification of 2012, SAP is defined by the presence of persistent organ failure (cardiovascular, respiratory, or renal failure exceeding 48 hours), which may involve single or multiple organ systems [3]. This revision distinguishes transient from persistent organ failure and identifies the latter as the primary determinant of mortality. When evaluating predictive models, four outcomes must be distinguished: 1) overall SAP, 2) persistent organ failure, 3) death, and 4) local complications (e.g., pancreatic necrosis, encapsulated necrosis). SAP carries a mortality rate of up to 30% and accounts for most AP-related deaths, often accompanied by local complications and prolonged ICU stays [4] [5].

The first 24 - 48 hours after onset represent the critical window for treatment decisions. During this period, personalized nutritional support, rapid ICU triage, and goal-directed therapy directly influence clinical outcomes [6]. Early identification of high-risk SAP patients enables clinicians to move beyond standardized treatment toward individualized interventions. An ideal prediction tool must allow rapid assessment at admission, combining high accuracy with broad applicability. Therefore, accurate early risk stratification remains a major clinical challenge. In this review, we discuss the evolution of SAP prediction strategies, including traditional scoring systems, biomarkers, imaging approaches, and emerging artificial intelligence-based models.

2. Traditional Clinical Scoring Systems

2.1. Ranson Criteria

The Ranson Criteria, first proposed by John H.C. Ranson and his team in 1974, was the first scale developed to classify the severity of acute pancreatitis [7]. This

scoring system was developed based on clinical data from 100 patients and comprises 11 indicators, which are assessed at two time points: "on admission" and "48 hours after admission". Five indicators are assessed at admission: age > 55 years, white blood cell count > 16,000/μL, blood glucose > 11.1 mmol/L, aspartate aminotransferase (AST) > 250 IU/L, and lactate dehydrogenase (LDH) > 350 IU/L. The assessment at 48 hours after admission included the following six criteria: a decrease in hematocrit > 10%, an increase in blood urea nitrogen (BUN) > 1.8 mmol/L, serum calcium < 2.0 mmol/L, arterial oxygen partial pressure (PaO_2) < 60 mmHg, a base deficit > 4 mmol/L, and an estimated fluid accumulation > 6 L. The risk of death increases as the Ranson score rises: the mortality rate is less than 3% for scores of 0 - 2, 10% - 20% for scores of 3 - 4, and as high as 40% or more for scores of 5 or higher [7] [8].

Although the Ranson score is widely used, it has significant shortcomings. On one hand, it requires 48 hours of data, making early risk assessment difficult; on the other hand, it was developed for people with alcoholic pancreatitis, so it can't be directly applied to patients with other causes [9]. Thirdly, the scores need to be collected in two separate periods for lab results, which is a complicated process and prone to missing data; its AUC for predicting severe acute pancreatitis within 48 hours is 0.81 - 0.95 [10]-[13].

2.2. Glasgow/Imrie Score

Clement W. Imrie first proposed the Glasgow Scale in 1970, also known as the Imrie criteria. By 1984, Blamey *et al.* refined this framework [14]. The system includes eight factors: age > 55 years, white blood cell count > 15,000/uL, blood glucose > 10 mmol/L, serum urea > 16 mmol/L, arterial PO_2 < 60 mmHg, serum calcium < 2.0 mmol/L, serum albumin < 32 g/L, and LDH > 600 IU/L. Like Ranson, the Glasgow score requires 48 hours for full assessment. Its AUC for predicting SAP at 48 hours ranges from 0.80 to 0.85 15. Its advantage is a relatively concise indicator set and more extensive validation in European populations. However, its 48-hour window limits early utility; when only admission data are used, sensitivity drops to 68% - 71% [15].

2.3. Apache II

The APACHE II score, introduced by Knaus *et al.* in 1985 for general ICU severity assessment [16], was adapted for AP by Larvin and McMahon in 1989 [17]. It includes 12 acute physiologic variables, age, and chronic health status, with scores from 0 to 71. Unlike Ranson and Glasgow, APACHE II can be assessed at any time, including at admission. A cutoff of 8 is typically used for predicting SAP. During the first 24 hours, sensitivity ranges from 63% to 74% and specificity from 73% to 81% [13] [17]. Serial APACHE II monitoring can track disease progression and treatment response.

APACHE II has several practical limitations. Its calculation requires 12 physiologic parameters, many not readily available in the emergency setting. It was not

designed specifically for AP, and some parameters have low relevance to pancreatitis. The calculation is cumbersome for routine clinical use. Mounzer *et al.* compared nine scoring systems and found APACHE II had an AUC of 0.78 for predicting persistent organ failure using 24-hour data, no better than simpler tools such as BISAP [15].

2.4. BISAP Score

The Bedside Index for Severity in Acute Pancreatitis (BISAP) was developed by Wu *et al.* in 2008 using classification and regression tree (CART) analysis of approximately 18,000 patients from 212 hospitals [18]. The BISAP score incorporates five parameters assessed within the first 24 hours: Blood urea nitrogen > 25 mg/dL, Impaired mental status (Glasgow Coma Scale < 15), Systemic inflammatory response syndrome (SIRS), Age > 60 years, and Pleural effusion on imaging. Each parameter is assigned one point, yielding a total score from 0 to 5. A BISAP score of 3 or higher is associated with significantly increased mortality risk [18] [19].

The BISAP score has gained considerable popularity due to its simplicity, use of readily available variables, and ability to be calculated entirely at the bedside within the first 24 hours. In the initial validation cohort, the BISAP score demonstrated an AUC of 0.82 for predicting mortality, comparable to APACHE II but with substantially greater ease of use [18] [19]. Subsequent studies have confirmed these findings, with AUC values for predicting SAP ranging from 0.77 to 0.85 at 24 hours [13] [18]. However, BISAP performs poorly in predicting persistent organ failure: at a cutoff of 2, sensitivity drops to 38% [15]. Additionally, the system assigns disproportionate weight to age, reducing its effectiveness in elderly patients.

2.5. Comparative Performance and Limitations

Papachristou *et al.* compared scoring systems in 185 AP patients and found the Ranson score had the best predictive ability (AUC 0.94), outperforming BISAP (0.81) and APACHE II (0.78). However, Ranson requires 48 hours, limiting early clinical application [13]. The predictive performance of traditional scores varies across populations and settings, with limited external generalizability [20]. As a result, a lot of research has started focusing on cutting-edge machine learning models, new biomarkers and advanced imaging technologies.

3. Biomarkers

3.1. Established Biomarkers

C-reactive protein (CRP) is the most widely used single biomarker for predicting progression to SAP. A meta-analysis reported an AUC of 0.85, sensitivity of 0.76, and specificity of 0.79 at 48 hours after onset [21].CRP ≥ 150 mg/L is generally used to predict severe acute pancreatitis, but CRP only peaks 48 - 72 hours after inflammation starts, which isn't great for early risk assessment. CRP is a non-spe-

cific acute phase reactant and can be affected by infections and other inflammatory diseases, which limits its accuracy in predicting severe acute pancreatitis [21] [22].

Procalcitonin (PCT) is a popular alternative marker to CRP; it can rise within 6 to 12 hours of inflammation, signalling systemic inflammation earlier. A meta-analysis of 18 studies (1,764 patients) reported a pooled sensitivity of 0.80, specificity of 0.84, and AUC of 0.89 for SAP prediction within 24 hours of admission [23]. Many studies show that PCT predicts SAP is better than CRP, but this indicator is costly to test and not widely available, so it can't be used routinely on a large scale. [23] [24].

Blood urea nitrogen (BUN) has gained significant attention in recent years due to its simplicity, low cost, and wide availability. In a study by Wu *et al.* that included 3 prospective cohorts totaling 1,043 patients with AP, a BUN level of ≥20 mg/dL at admission was found to be significantly associated with an increased risk of death (OR = 4.6) [25]. Also, consistently high BUN levels within 24 hours of hospital admission are an independent risk factor for increased mortality. Monitoring BUN dynamically can help assess disease progression and prognosis. The BISAP score includes BUN ≥ 25 mg/dL as a key evaluation criterion, further showing its role in early risk stratification of acute pancreatitis [18].

3.2. Emerging and Novel Biomarkers

Interleukin-6 (IL-6) is a key pro-inflammatory cytokine that induces CRP synthesis. Its levels rise rapidly in early AP, making it useful for early risk prediction. A meta-analysis of 13 studies reported an AUC of 0.85 within 24 hours of onset and a diagnostic odds ratio of 14 (95% CI: 7 - 27) [26]. Despite its promise, clinical application is limited by a lack of standardized testing methods, high costs, and the absence of a uniform cutoff value 26,27. Other interleukins such as IL-8 and IL-10 are also involved in the inflammatory response but show inferior predictive ability compared to IL-6 [27].

Immune cell ratio indices derived from complete blood counts have been applied to early AP risk assessment due to their simplicity and low cost. The neutrophil-to-lymphocyte ratio (NLR) reflects both inflammatory response and immune status. A meta-analysis by Kong *et al.* reported a pooled sensitivity of 0.79, specificity of 0.71, and AUC of 0.82 for SAP prediction [28]. The platelet-to-lymphocyte ratio (PLR) also has predictive value; combining NLR and PLR can further improve performance [29]. However, these markers lack sufficient accuracy to be used alone for clinical decision-making.

Trypsinogen activation peptide (TAP) directly reflects the premature activation of trypsinogen within the pancreas, and premature trypsinogen activation is one of the key pathophysiological mechanisms underlying the onset of AP.A multicenter study involving 246 patients showed that urinary TAP had a sensitivity of 58% and 83% for predicting severe acute pancreatitis at 24 and 48 hours after onset, respectively, with a specificity of approximately 72% to 73% [30]. However,

due to the complexity of the testing process, the susceptibility of TAP in urine to degradation, and the fact that commercially available test kits are not yet fully developed, its clinical application remains somewhat limited.

4. Radiological Scoring Systems

4.1. Balthazar CT Severity Index

Contrast-enhanced computed tomography (CECT) remains the primary imaging modality for morphologic assessment in AP. In 1985, Balthazar *et al.* described a CT grading system based on pancreatic morphology [31], later combined with a necrosis scoring component in 1990 to form the CT Severity Index (CTSI) [32]. The CTSI uses a 10-point system: the Balthazar index (grades A - E, 0 - 4 points) plus a necrosis score (0 points for no necrosis, 2 for <30%, 4 for 30% - 50%, 6 for >50%). Assessed after 72 hours, a score greater than or equal to 7 is associated with high risk of purulent necrosis, prolonged hospitalization, and increased mortality [33].

4.2. Modified CT Severity Index

The Modified CT Severity Index (MCTSI), introduced by Mortele *et al.* in 2004, simplified the original CTSI by using a 10-point scale that incorporates inflammation (0 - 4 points), necrosis (0 - 4 points), and extrapancreatic complications (2 points) [34]. The MCTSI demonstrated improved correlation with patient outcomes, including length of hospital stay, need for intervention, and organ failure. In a comparative analysis by Bollen *et al.*, various CT scoring systems, including the CTSI and MCTSI, demonstrated comparable predictive accuracy, with no statistically significant differences among them [35].

4.3. Limitations of Radiological Scoring

CT-based scoring systems have fundamental limitations for early prediction. CECT is not recommended within the first 48 - 72 hours of admission because early imaging may underestimate necrosis and expose patients to contrast during a period of potential renal compromise [6] [36]. CT scores cannot reliably predict early mortality or systemic organ failure before necrosis develops. The 2026 iLATAM-AP guidelines recommend optimal CECT timing for severity assessment at least 72 hours after pain onset [37]. MRI is an alternative but is limited by prolonged scanning times, high cost, and limited availability [5].

5. Emerging Tools: Artificial Intelligence and Machine Learning

5.1. The Paradigm Shift

Traditional scoring tools and individual biomarkers have inherent analytical limitations. Machine learning (ML) and artificial intelligence (AI) can capture complex, non-linear interactions across multiple data types. Unlike standard regres-

sion, which requires manual specification of cross-variable relationships, ML algorithms learn data hierarchies autonomously and can identify predictive patterns not apparent through conventional analysis [38] [39].

5.2. Machine Learning Models for SAP Prediction

Extreme Gradient Boosting (XGBoost) performs steadily in SAP prognosis assessment. Thapa and others built an XGBoost model based on 61,894 electronic medical records with 24-hour clinical data, achieving an AUC of 0.921, outperforming HAPS and BISAP scores [40]. Luo and others compared five machine learning models, and the results showed that the random forest (RF) performed best at admission, with internal and external validation AUCs of 0.969 and 0.961, clearly better than the Ranson score (0.847) and BISAP score (0.796) [41].

Hong *et al.* applied a random forest model with 16 variables to 648 AP cases, achieving a cross-validated AUC of 0.89 and a test-phase AUC of 0.96 [42]. Feature importance analysis picked out blood sugar, blood calcium, HDL, LDL, albumin, creatinine, and BUN as key predictors; researchers used the LIME method to visually explain individual risks [42].

Litvin *et al.* developed an artificial neural network (ANN) that outperformed logistic regression for admission-day SAP prediction (AUC 0.92 vs. 0.84) [43]. The multicenter PANCREATIA study tested models using only emergency room vital signs and demographic data, without laboratory or imaging results, and achieved an AUC of 0.849 for mortality prediction [44], demonstrating that AI can extract clinically useful signals from minimal data.

5.3. Multimodal and Multi-Omics Approaches

Current research tends to integrate multiple data sources to build a unified prediction model. Yin and others developed the PrismSAP model framework, combining CT deep learning features, radiomics, laboratory results, and clinical indicators, achieving a 72-hour prediction AUC of 0.916 [45]. Li and colleagues used a voting ensemble to sort AP into three severity levels (MAP, MSAP, SAP), getting a C-statistic of 0.902 and using SHAP to make it easier to understand [46]. These findings suggest that integrating multidimensional clinical data enables more accurate risk stratification than any single data type alone.

5.4. Evidence from Systematic Reviews

Qian *et al.* synthesized 33 studies of ML models for SAP prediction and reported pooled C-indices of 0.88 (95% CI: 0.86 - 0.90) in external validation and 0.87 (95% CI: 0.84 - 0.89) internally [47]. However, due to differences in study populations, predictor variables, study endpoints and model frameworks, there is considerable heterogeneity in current research methodologies, so the results should be interpreted with caution. Further comparisons of machine learning models and traditional scores are needed within the same prospective cohort to better verify the clinical superiority of the former. Critelli and others summarised that machine

learning models have an average AUC of 0.91, but most studies are single-centre retrospective analyses, lacking external validation and with limited generalizability [48]. The author emphasised that the relevant models need to be validated through prospective, multicentre studies.

6. Future Directions and Clinical Challenges

6.1. Data Generalizability and Validation

Even if the AUC values are high in retrospective studies, they often can't be directly applied to real clinical work. Most predictive models are built using data from a single centre, making it hard to reliably use them in other population cohorts [47]. Predictive models that are optimised for specific groups (like East Asians or Caucasians) usually work poorly when applied to other groups. This is because the genetic risk factors, disease causes, and medical resource conditions differ between populations. On top of that, variations in how hospitals record medical histories, imaging protocols, and testing methods can further make these models perform worse during external validation [48] [49].

6.2. Interpretability and the "Black Box" Problem

One big hurdle for applying machine learning models in clinics is their lack of interpretability. Traditional scoring systems have clear logic and the derivation process is easy to follow. In contrast, gradient boosting models and deep neural networks are 'black box models', making it hard to explain the prediction for each patient from a biological perspective [50]. These black box issues make clinicians responsible for diagnostic decisions a bit wary. Currently, the academic world has proposed quite a few methods, like SHAP, LIME and attention mechanisms, which can explain model predictions after the fact [40] [42], but these methods have their own limitations, including instability and vulnerability to manipulation.

6.3. Clinical Validation and Prospective Studies

Most machine learning-based severe acute pancreatitis (SAP) prediction models haven't been prospectively validated or tested in randomised controlled trials. Most of the existing evidence comes from retrospective analyses, which have inherent biases like selection bias, information bias, and confounding bias [48]. So far, no model has met the strict standards for clinical application: none have been proven to be practically useful in clinics, nor shown to improve patient outcomes, and there's a lack of evidence on cost-effectiveness. Regulators still haven't put forward a clear set of rules to assess AI clinical decision support tools [51].

A very common methodological flaw in machine learning research is data leakage: that is, using data when training a model that wouldn't be available at the moment of prediction, such as later trends in test indicators, CT scan results 72 hours after onset, or treatment-related data [52]. This could lead to inflated performance metrics in the model reports. Any further research must ensure that the

variables included in the model are limited to data available at the initial triage assessment [53].

Prospective studies should clearly define their clinical targets. Models must distinguish among persistent organ failure, overall SAP, local complications, and all-cause mortality, as optimal predictors and risk thresholds differ across these outcomes. Composite endpoints are insufficient for guiding targeted interventions [54].

6.4. Beyond Discrimination: Calibration and Clinical Utility

Although most machine learning studies show ideal AUC results, just having good discriminatory ability is not enough to prove that the model can safely guide clinical treatment [55]. The model also needs to be well calibrated, meaning its predicted probabilities should match the actual event occurrence rates [56]. If the model calibration is poor, it can lead to unreasonable clinical interventions, unnecessary treatments, or missed necessary interventions. Therefore, decision curve analysis (DCA) is needed to verify the model's actual clinical value at different risk threshold levels [57]. The best cutoff value can't be generalised. When screening in general wards for SAP or ruling out severe cases, try not to miss any diagnoses (high sensitivity); when deciding whether a patient needs to be transferred to the ICU, try not to get it wrong (high specificity). Before the model is officially used in clinics, it must pass both calibration and DCA clinical net benefit validation [58].

6.5. Integration into Clinical Workflow

It's still really tricky to use machine learning models at the bedside. A lot of hospital IT systems don't support real-time calculations; in the busy pace of the emergency department, extra manual data entry is hard to make happen [59] [60]. There's still no clear rule on who's responsible when AI-assisted diagnosis goes wrong. If the model underestimates the risk for high-risk patients and their condition gets worse, it's still unclear what kind of legal responsibility everyone should bear [61]. For a model to really take off in clinics, it needs to tick three boxes: smoothly connect with existing hospital electronic medical records, have a simple and user-friendly interface, and clearly define the division of labour and decision-making rules between doctors and AI.

6.6. Future Horizons

There are currently several cutting-edge directions to further optimise SAP prediction models. By combining bedside indicators, imaging, and multi-omics data, we can find new prognostic markers and explain the molecular mechanisms behind necrosis progression and organ failure. Federated learning can enable multi-centre joint modelling without compromising patient privacy, while also improving model generalisability. Natural language processing can extract useful information from medical records, enriching the model's input features. Finally, the

model needs to evolve from a single static assessment to a dynamic risk prediction system that can be updated in real time.

7. Conclusion

The risk prediction for severe acute pancreatitis has moved from traditional scoring tools to a new stage that combines lab results, imaging findings and AI models. While traditional scores can be used for early stratification, their predictive accuracy and clinical usefulness are quite limited. Individual lab tests and imaging scores also can't fully cover the disease's complex pathological changes. Machine learning can integrate multidimensional data and explore nonlinear relationships, offering better prediction results. But at present, models generally face issues like poor generalisation, lack of interpretability, absence of prospective validation and difficulty in practical implementation. In the future, multicentre prospective studies are needed to optimise model performance and interpretability and to ensure the models are truly integrated into clinical practice.

Author Contributions

Study conception and methodology design: Hao Li, Chuanxin Zou; systematic literature retrieval and screening: Hao Li; data extraction, quality assessment and evidence interpretation: Hao Li, Chuanxin Zou; manuscript drafting, revision and polishing: Hao Li; final approval of the submitted version: Hao Li, Chuanxin Zou.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Li, T., Qin, C., Zhao, B., Li, Z., Zhao, Y., Lin, C., *et al.* (2024) Global and Regional Burden of Pancreatitis: Epidemiological Trends, Risk Factors, and Projections to 2050 from the Global Burden of Disease Study 2021. *BMC Gastroenterology*, **24**, Article No. 398. https://doi.org/10.1186/s12876-024-03481-8

[2] Xiao, A.Y., Tan, M.L.Y., Wu, L.M., Asrani, V.M., Windsor, J.A., Yadav, D., *et al.* (2016) Global Incidence and Mortality of Pancreatic Diseases: A Systematic Review, Meta-Analysis, and Meta-Regression of Population-Based Cohort Studies. *The Lancet Gastroenterology & Hepatology*, **1**, 45-55. https://doi.org/10.1016/s2468-1253(16)30004-8

[3] Banks, P.A., Bollen, T.L., Dervenis, C., Gooszen, H.G., Johnson, C.D., Sarr, M.G., *et al.* (2013) Classification of Acute Pancreatitis—2012: Revision of the Atlanta Classification and Definitions by International Consensus. *Gut*, **62**, 102-111. https://doi.org/10.1136/gutjnl-2012-302779

[4] Mederos, M.A., Reber, H.A. and Girgis, M.D. (2021) Acute Pancreatitis. *JAMA*, **325**, 382-390. https://doi.org/10.1001/jama.2020.20317

[5] Tenner, S., Vege, S.S., Sheth, S.G., Sauer, B., Yang, A., Conwell, D.L., *et al.* (2024) American College of Gastroenterology Guidelines: Management of Acute Pancreatitis. *American Journal of Gastroenterology*, **119**, 419-437. https://doi.org/10.14309/ajg.0000000000002645

[6] Crockett, S.D., Wani, S., Gardner, T.B., Falck-Ytter, Y., Barkun, A.N., Crockett, S., *et al.* (2018) American Gastroenterological Association Institute Guideline on Initial Management of Acute Pancreatitis. *Gastroenterology*, **154**, 1096-1101. https://doi.org/10.1053/j.gastro.2018.01.032

[7] Ranson, J.H.C., Rifkind, K.M., Roses, D.F., Fink, S.D., Eng, K. and Spencer, F.C. (1974) Prognostic Signs and the Role of Operative Management in Acute Pancreatitis. *Surgery, Gynecology & Obstetrics*, **139**, 69-81.

[8] Basit, H., Ruan, G.J. and Mukherjee, S. (2025) Ranson Criteria. StatPearls. https://pubmed.ncbi.nlm.nih.gov/29493970/

[9] Corfield, A.P., Williamson, R.C.N., Mcmahon, M.J., Shearer, M.G., Cooper, M.J., Mayer, A.D., *et al.* (1985) Prediction of Severity in Acute Pancreatitis: Prospective Comparison of Three Prognostic Indices. *The Lancet*, **326**, 403-407. https://doi.org/10.1016/s0140-6736(85)92733-3

[10] Shi, P.N., Song, Z.Z., He, X.N. and Hong, J.N. (2025) Evaluation of Scoring Systems and Hematological Parameters in the Severity Stratification of Early-Phase Acute Pancreatitis. *World Journal of Gastroenterology*, **31**, Article No. 105236. https://doi.org/10.3748/wjg.v31.i15.105236

[11] Zhu, J., Wu, L., Wang, Y., Fang, M., Liu, Q. and Zhang, X. (2024) Predictive Value of the Ranson and BISAP Scoring Systems for the Severity and Prognosis of Acute Pancreatitis: A Systematic Review and Meta-Analysis. *PLOS ONE*, **19**, e0302046. https://doi.org/10.1371/journal.pone.0302046

[12] Metri, A., Bush, N. and Singh, V.K. (2024) Predicting the Severity of Acute Pancreatitis: Current Approaches and Future Directions. *Surgery Open Science*, **19**, 109-117. https://doi.org/10.1016/j.sopen.2024.03.012

[13] Papachristou, G.I., Muddana, V., Yadav, D., O'Connell, M., Sanders, M.K., Slivka, A., *et al.* (2010) Comparison of BISAP, Ranson's, APACHE-II, and CTSI Scores in Predicting Organ Failure, Complications, and Mortality in Acute Pancreatitis. *American Journal of Gastroenterology*, **105**, 435-441. https://doi.org/10.1038/ajg.2009.622

[14] Blamey, S.L., Imrie, C.W., O'Neill, J., Gilmour, W.H. and Carter, D.C. (1984) Prognostic Factors in Acute Pancreatitis. *Gut*, **25**, 1340-1346. https://doi.org/10.1136/gut.25.12.1340

[15] Mounzer, R., Langmead, C.J., Wu, B.U., Evans, A.C., Bishehsari, F., Muddana, V., *et al.* (2012) Comparison of Existing Clinical Scoring Systems to Predict Persistent Organ Failure in Patients with Acute Pancreatitis. *Gastroenterology*, **142**, 1476-1482. https://doi.org/10.1053/j.gastro.2012.03.005

[16] Knaus, W.A., Draper, E.A., Wagner, D.P. and Zimmerman, J.E. (1985) Apache II: A Severity of Disease Classification System. *Critical Care Medicine*, **13**, 818-829. https://doi.org/10.1097/00003246-198510000-00009

[17] Larvin, M. and Mcmahon, M. (1989) Apache-II Score for Assessment and Monitoring of Acute Pancreatitis. *The Lancet*, **334**, 201-205. https://doi.org/10.1016/s0140-6736(89)90381-4

[18] Wu, B.U., Johannes, R.S., Sun, X., Tabak, Y., Conwell, D.L. and Banks, P.A. (2008) The Early Prediction of Mortality in Acute Pancreatitis: A Large Population-Based Study. *Gut*, **57**, 1698-1703. https://doi.org/10.1136/gut.2008.152702

[19] Singh, V.K., Wu, B.U., Bollen, T.L., Repas, K., Maurer, R., Johannes, R.S., *et al.* (2009) A Prospective Evaluation of the Bedside Index for Severity in Acute Pancreatitis Score in Assessing Mortality and Intermediate Markers of Severity in Acute Pancreatitis.

The American Journal of Gastroenterology, **104**, 966-971.
https://doi.org/10.1038/ajg.2009.28

[20] Capurso, G., Ponz de Leon Pisani, R., Lauri, G., Archibugi, L., Hegyi, P., Papachristou, G.I., *et al.* (2023) Clinical Usefulness of Scoring Systems to Predict Severe Acute Pancreatitis: A Systematic Review and Meta-Analysis with Pre and Post-Test Probability Assessment. *United European Gastroenterology Journal*, **11**, 825-836.
https://doi.org/10.1002/ueg2.12464

[21] Li, Y., Zhao, X. and Meng, F. (2024) Diagnostic Value of CRP for Predicting the Severity of Acute Pancreatitis: A Systematic Review and Meta-Analysis. *Biomarkers*, **29**, 494-503. https://pubmed.ncbi.nlm.nih.gov/39417604/
https://doi.org/10.1080/1354750X.2024.2415463

[22] Khanna, A.K., Meher, S., Prakash, S., Tiwary, S.K., Singh, U., Srivastava, A., *et al.* (2013) Comparison of Ranson, Glasgow, MOSS, SIRS, BISAP, APACHE-II, CTSI Scores, IL-6, CRP, and Procalcitonin in Predicting Severity, Organ Failure, Pancreatic Necrosis, and Mortality in Acute Pancreatitis. *HPB Surgery*, **2013**, 1-10.
https://doi.org/10.1155/2013/367581

[23] Chen, L. and Jiang, J. (2022) The Diagnostic Value of Procalcitonin in Patients with Severe Acute Pancreatitis: A Meta-Analysis. *Turkish Journal of Gastroenterology*, **33**, 722-730. https://doi.org/10.5152/tjg.2022.22098

[24] Rau, B.M., Kemppainen, E.A., Gumbs, A.A., *et al.* (2007) Early Assessment of Pancreatic Infections and Overall Prognosis in Severe Acute Pancreatitis by Procalcitonin (PCT). *Annals of Surgery*, **245**, 745-754.
https://doi.org/10.1097/01.sla.0000252443.22360.46
https://pmc.ncbi.nlm.nih.gov/articles/PMC1877072/

[25] Wu, B.U., Bakker, O.J., Papachristou, G.I., Besselink, M.G., Repas, K., van Santvoort, H.C., *et al.* (2011) Blood Urea Nitrogen in the Early Assessment of Acute Pancreatitis: An International Validation Study. *Archives of Internal Medicine*, **171**, 669-676.
https://doi.org/10.1001/archinternmed.2011.126

[26] Kumar, S., Aziz, T., Kumar, R., Kumar, P., Kumar, A., Saha, A., *et al.* (2025) Diagnostic Accuracy of Interleukin-6 as a Biomarker for Early Prediction of Severe Acute Pancreatitis: A Systematic Review and Meta-Analysis. *Journal of Family Medicine and Primary Care*, **14**, 667-674. https://doi.org/10.4103/jfmpc.jfmpc_1366_24

[27] Mititelu, A., Grama, A., Colceriu, M., Bența, G., Popoviciu, M. and Pop, T.L. (2024) Role of Interleukin 6 in Acute Pancreatitis: A Possible Marker for Disease Prognosis. *International Journal of Molecular Sciences*, **25**, Article 8283.
https://doi.org/10.3390/ijms25158283

[28] Kong, W., He, Y., Bao, H., Zhang, W. and Wang, X. (2020) Diagnostic Value of Neutrophil-Lymphocyte Ratio for Predicting the Severity of Acute Pancreatitis: A Meta-Analysis. *Disease Markers*, **2020**, Article ID: 9731854.
https://doi.org/10.1155/2020/9731854

[29] Kaplan, M., Ates, I., Oztas, E., *et al.* (2018) A New Marker to Determine Prognosis of Acute Pancreatitis: PLR and NLR Combination. *Journal of Medical Biochemistry*, **37**, 21-30. https://doi.org/10.1515/jomb-2017-0039
https://pubmed.ncbi.nlm.nih.gov/30581338/

[30] Neoptolemos, J.P., Kemppainen, E.A., Mayer, J., Fitzpatrick, J., Raraty, M., Slavin, J., *et al.* (2000) Early Prediction of Severity in Acute Pancreatitis by Urinary Trypsinogen Activation Peptide: A Multicentre Study. *The Lancet*, **355**, 1955-1960.
https://doi.org/10.1016/s0140-6736(00)02327-8

[31] Balthazar, E.J., Ranson, J.H., Naidich, D.P., Megibow, A.J., Caccavale, R. and Cooper,

M.M. (1985) Acute Pancreatitis: Prognostic Value of CT. *Radiology*, **156**, 767-772. https://doi.org/10.1148/radiology.156.3.4023241

[32] Balthazar, E.J., Robinson, D.L., Megibow, A.J. and Ranson, J.H. (1990) Acute Pancreatitis: Value of CT in Establishing Prognosis. *Radiology*, **174**, 331-336. https://doi.org/10.1148/radiology.174.2.2296641

[33] Balthazar, E.J. (2002) Acute Pancreatitis: Assessment of Severity with Clinical and CT Evaluation. *Radiology*, **223**, 603-613. https://doi.org/10.1148/radiol.2233010680

[34] Mortele, K.J., Wiesner, W., Intriere, L., Shankar, S., Zou, K.H., Kalantari, B.N., *et al.* (2004) A Modified CT Severity Index for Evaluating Acute Pancreatitis: Improved Correlation with Patient Outcome. *American Journal of Roentgenology*, **183**, 1261-1265. https://doi.org/10.2214/ajr.183.5.1831261

[35] Bollen, T.L., Singh, V.K., Maurer, R., Repas, K., van Es, H.W., Banks, P.A., *et al.* (2012) A Comparative Evaluation of Radiologic and Clinical Scoring Systems in the Early Prediction of Severity in Acute Pancreatitis. *American Journal of Gastroenterology*, **107**, 612-619. https://doi.org/10.1038/ajg.2011.438

[36] Thoeni, R.F. (2012) The Revised Atlanta Classification of Acute Pancreatitis: Its Importance for the Radiologist and Its Effect on Treatment. *Radiology*, **262**, 751-764. https://doi.org/10.1148/radiol.11110947

[37] Cárdenas-Jaén, K., Guilabert, L., Martínez-Moneo, E., Mancilla, C., Bispo, M., Moutinho-Ribeiro, P., *et al.* (2026) AEG-AESPANC-OPGE-SIED-SPG Ibero-Latin American Guidelines on Acute Pancreatitis (iLATAM-AP). *United European Gastroenterology Journal*, **14**, e70190. https://doi.org/10.1002/ueg2.70190

[38] Zhou, Y., Ge, Y., Shi, X., Wu, K., Chen, W., Ding, Y., *et al.* (2022) Machine Learning Predictive Models for Acute Pancreatitis: A Systematic Review. *International Journal of Medical Informatics*, **157**, Article 104641. https://doi.org/10.1016/j.ijmedinf.2021.104641

[39] Tan, Z., Li, G., Zheng, Y., Li, Q., Cai, W., Tu, J., *et al.* (2024) Advances in the Clinical Application of Machine Learning in Acute Pancreatitis: A Review. *Frontiers in Medicine*, **11**, Article ID: 1487271. https://doi.org/10.3389/fmed.2024.1487271

[40] Thapa, R., Iqbal, Z., Garikipati, A., Siefkas, A., Hoffman, J., Mao, Q., *et al.* (2022) Early Prediction of Severe Acute Pancreatitis Using Machine Learning. *Pancreatology*, **22**, 43-50. https://doi.org/10.1016/j.pan.2021.10.003

[41] Luo, Z., Shi, J., Fang, Y., Pei, S., Lu, Y., Zhang, R., *et al.* (2023) Development and Evaluation of Machine Learning Models and Nomogram for the Prediction of Severe Acute Pancreatitis. *Journal of Gastroenterology and Hepatology*, **38**, 468-475. https://doi.org/10.1111/jgh.16125

[42] Hong, W., Lu, Y., Zhou, X., Jin, S., Pan, J., Lin, Q., *et al.* (2022) Usefulness of Random Forest Algorithm in Predicting Severe Acute Pancreatitis. *Frontiers in Cellular and Infection Microbiology*, **12**, Article ID: 893294. https://doi.org/10.3389/fcimb.2022.893294

[43] Litvin, A., Korenev, S., Rumovskaya, S., Sartelli, M., Baiocchi, G., Biffl, W.L., *et al.* (2021) WSES Project on Decision Support Systems Based on Artificial Neural Networks in Emergency Surgery. *World Journal of Emergency Surgery*, **16**, Article No. 50. https://doi.org/10.1186/s13017-021-00394-9

[44] Villasante, S., Fernandes, N., Perez, M., Cordobés, M.A., Piella, G., Martínez, M., *et al.* (2026) Prediction of Severe Acute Pancreatitis at a Very Early Stage of the Disease Using Artificial Intelligence Techniques, without Laboratory Data or Imaging Tests. *Annals of Surgery*, **284**, e31-e40. https://doi.org/10.1097/sla.0000000000006579

[45] Yin, M., Lin, J., Wang, Y., Liu, Y., Zhang, R., Duan, W., *et al.* (2024) Development and Validation of a Multimodal Model in Predicting Severe Acute Pancreatitis Based on Radiomics and Deep Learning. *International Journal of Medical Informatics*, **184**, Article 105341. https://doi.org/10.1016/j.ijmedinf.2024.105341

[46] Li, J., Mu, D., Zheng, S., Tian, W., Wu, Z., Meng, J., *et al.* (2023) Machine Learning Improves Prediction of Severity and Outcomes of Acute Pancreatitis: A Prospective Multi-Center Cohort Study. *Science China Life Sciences*, **66**, 1934-1937. https://doi.org/10.1007/s11427-022-2333-8

[47] Qian, R., Zhuang, J., Xie, J., Cheng, H., Ou, H., Lu, X., *et al.* (2024) Predictive Value of Machine Learning for the Severity of Acute Pancreatitis: A Systematic Review and Meta-Analysis. *Heliyon*, **10**, e29603. https://doi.org/10.1016/j.heliyon.2024.e29603

[48] Critelli, B., Hassan, A., Lahooti, I., Noh, L., Park, J.S., Tong, K., *et al.* (2025) A Systematic Review of Machine Learning-Based Prognostic Models for Acute Pancreatitis: Towards Improving Methods and Reporting Quality. *PLOS Medicine*, **22**, e1004432. https://doi.org/10.1371/journal.pmed.1004432

[49] López Gordo, S., Ramirez-Maldonado, E., Fernandez-Planas, M.T., Bombuy, E., Memba, R. and Jorba, R. (2025) AI and Machine Learning for Precision Medicine in Acute Pancreatitis: A Narrative Review. *Medicina*, **61**, Article 629. https://doi.org/10.3390/medicina61040629

[50] Deng, J., Elghobashy, M.E., Zang, K., Patel, S.K., Guo, E. and Heybati, K. (2025) So You've Got a High AUC, Now What? An Overview of Important Considerations When Bringing Machine-Learning Models from Computer to Bedside. *Medical Decision Making*, **45**, 640-653. https://doi.org/10.1177/0272989x251343082

[51] Abulibdeh, R., Celi, L.A. and Sejdić, E. (2025) The Illusion of Safety: A Report to the FDA on AI Healthcare Product Approvals. *PLOS Digital Health*, **4**, e0000866. https://doi.org/10.1371/journal.pdig.0000866

[52] Kapoor, S. and Narayanan, A. (2023) Leakage and the Reproducibility Crisis in Machine-Learning-Based Science. *Patterns*, **4**, Article 100804. https://doi.org/10.1016/j.patter.2023.100804

[53] Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V.X., Doshi-Velez, F., *et al.* (2019) Do No Harm: A Roadmap for Responsible Machine Learning for Health Care. *Nature Medicine*, **25**, 1337-1340. https://doi.org/10.1038/s41591-019-0548-6

[54] Wolff, R.F., Moons, K.G.M., Riley, R.D., Whiting, P.F., Westwood, M., Collins, G.S., *et al.* (2019) PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, **170**, 51-58. https://doi.org/10.7326/m18-1376

[55] Collins, G.S., Reitsma, J.B., Altman, D.G. and Moons, K.G.M. (2015) Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Annals of Internal Medicine*, **162**, 55-63. https://doi.org/10.7326/m14-0697

[56] Van Calster, B., McLernon, D.J., van Smeden, M., Wynants, L. and Steyerberg, E.W. (2019) Calibration: The Achilles Heel of Predictive Analytics. *BMC Medicine*, **17**, Article No. 230. https://doi.org/10.1186/s12916-019-1466-7

[57] Vickers, A.J. and Elkin, E.B. (2006) Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making*, **26**, 565-574. https://doi.org/10.1177/0272989x06295361

[58] Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G. and King, D. (2019) Key Challenges for Delivering Clinical Impact with Artificial Intelligence. *BMC Medicine*,

17, Article No. 195. https://doi.org/10.1186/s12916-019-1426-2

[59] Hu, Q., Hu, Y., Tan, C., Yang, Y., Su, H., Huang, Z., *et al.* (2026) Acute Pancreatitis: Mechanisms and Therapeutic Approaches. *Signal Transduction and Targeted Therapy*, **11**, Article No. 15. https://doi.org/10.1038/s41392-025-02394-6

[60] Finkelstein, J., Gabriel, A., Schmer, S., Truong, T. and Dunn, A. (2024) Identifying Facilitators and Barriers to Implementation of AI-Assisted Clinical Decision Support in an Electronic Health Record System. *Journal of Medical Systems*, **48**, Article No. 89. https://doi.org/10.1007/s10916-024-02104-9

[61] Mello, M.M. and Guha, N. (2024) Understanding Liability Risk from Using Health Care Artificial Intelligence Tools. *New England Journal of Medicine*, **390**, 271-278. https://doi.org/10.1056/nejmhle2308901