Revista Eletrônica
Acervo Saúde
Electronic Journal Collection Health ISSN 2178-2091

Aplicações da inteligência artificial no diagnóstico diferencial do abdômen agudo

Applications of artificial intelligence in the differential diagnosis of acute abdomen

Aplicaciones de la inteligencia artificial en el diagnóstico diferencial del abdomen agudo

Laura Gava Moretti[1], Sofia Alves Pignataro[1], Vitória Chies Colassiol[2], Fernanda Juliani Cardia[1], Maria Vitória Teixeira Vesaro[1], Júlia Karoline Viana Fabi[3], Maria Júlia Biazon Alves[3], Alessandro Gabriel Macedo Veiga[4].

RESUMO

Objetivo: Sintetizar sobre o uso da inteligência artificial e do machine learning no diagnóstico diferencial, na estratificação de risco e no suporte à decisão clínica e cirúrgica em pacientes com abdômen agudo. **Métodos:** Trata-se de uma revisão integrativa conduzida conforme as diretrizes PRISMA. A busca foi realizada na base de dados PubMed utilizando descritores relacionados à inteligência artificial, machine learning, diagnóstico e abdômen agudo. Foram incluídos estudos publicados entre 2023 e 2026 que avaliaram a aplicação dessas tecnologias em pacientes com abdômen agudo. **Resultados:** Foram selecionados 11 estudos, os quais demonstraram resultados promissores na identificação de doenças abdominais agudas, estratificação de risco, predição de sepse, mortalidade e complicações cirúrgicas, além do suporte à interpretação de exames de imagem. Os modelos apresentaram bom desempenho em diferentes contextos clínicos e cirúrgicos. **Considerações finais:** A inteligência artificial e o machine learning apresentam potencial para auxiliar o diagnóstico diferencial e a tomada de decisão no manejo do abdômen agudo, atuando como ferramentas complementares à avaliação clínica. Entretanto, são necessários estudos adicionais para validação clínica e externa desses modelos.

Palavras-chave: Inteligência artificial, Machine learning, Abdômen agudo, Diagnóstico diferencial, Tomografia computadorizada.

ABSTRACT

Objective: To synthesize the evidence on the use of artificial intelligence and machine learning in differential diagnosis, risk stratification, and clinical and surgical decision support in patients with acute abdomen. **Methods:** This is an integrative review conducted in accordance with the PRISMA guidelines. The search was performed in the PubMed database using descriptors related to artificial intelligence, machine learning,

[1] Fundação Educacional do Município de Assis (FEMA), Assis – SP.
[2] Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS), Porto Alegre – RS.
[3] Universidade de Marília (UNIMAR), Marília – SP.
[4] Universidade do Oeste Paulista (UNOESTE), Jaú – SP.

SUBMETIDO EM: 8/2026 | ACEITO EM: 9/2026 | PUBLICADO EM: 10/2026

diagnosis, and acute abdomen. Studies published between 2023 and 2026 that evaluated the application of these technologies in patients with acute abdomen were included. **Results:** Eleven studies were selected, demonstrating promising results in the identification of acute abdominal diseases, risk stratification, prediction of sepsis, mortality, and surgical complications, as well as support for the interpretation of imaging examinations. The models showed good performance across different clinical and surgical settings. **Final considerations:** Artificial intelligence and machine learning have the potential to assist in differential diagnosis and decision-making in the management of acute abdomen, serving as complementary tools to clinical assessment. However, further studies are needed for the clinical and external validation of these models.

Keywords: Artificial intelligence, Machine learning, Acute abdomen, Differential diagnosis, Computed tomography.

RESUMEN

Objetivo: Sintetizar la evidencia sobre el uso de la inteligencia artificial y el aprendizaje automático en el diagnóstico diferencial, la estratificación del riesgo y el apoyo a la toma de decisiones clínicas y quirúrgicas en pacientes con abdomen agudo. **Métodos:** Se trata de una revisión integradora realizada de acuerdo con las directrices PRISMA. La búsqueda se llevó a cabo en la base de datos PubMed utilizando descriptores relacionados con inteligencia artificial, aprendizaje automático, diagnóstico y abdomen agudo. Se incluyeron estudios publicados entre 2023 y 2026 que evaluaron la aplicación de estas tecnologías en pacientes con abdomen agudo. **Resultados:** Se seleccionaron 11 estudios, los cuales demostraron resultados prometedores en la identificación de enfermedades abdominales agudas, la estratificación del riesgo, la predicción de sepsis, mortalidad y complicaciones quirúrgicas, así como en el apoyo a la interpretación de estudios de imagen. Los modelos presentaron un buen desempeño en diferentes contextos clínicos y quirúrgicos. **Consideraciones finales:** La inteligencia artificial y el aprendizaje automático presentan potencial para contribuir al diagnóstico diferencial y a la toma de decisiones en el manejo del abdomen agudo, actuando como herramientas complementarias a la evaluación clínica. Sin embargo, se requieren estudios adicionales para la validación clínica y externa de estos modelos.

Palabras clave: Inteligencia artificial, Aprendizaje automático, Abdomen agudo, Diagnóstico diferencial, Tomografía computarizada.

INTRODUÇÃO

O abdômen agudo representa uma das principais causas de atendimento em serviços de emergência em âmbito mundial, englobando um conjunto heterogêneo de condições inflamatórias, obstrutivas, perfurativas e vasculares que frequentemente demandam intervenção cirúrgica urgente (SARTELLI M, et al., 2017). Estudos epidemiológicos apontam que a dor abdominal aguda é responsável por aproximadamente 7% a 10% de todas as consultas em departamentos de pronto-atendimento (BRODER J, et al., 2022). Diante dessa elevada prevalência, a condição se configura como um crítico problema de saúde pública, cujo potencial de gravidade exige uma abordagem imediata para evitar desfechos desfavoráveis e complicações severas.

A elevada morbimortalidade associada ao abdômen agudo está diretamente relacionada à rapidez com que o diagnóstico definitivo e a terapêutica cirúrgica são estabelecidos (SARTELLI M, et al., 2017). Condições de evolução complexa, tais como apendicite aguda complicada, isquemia mesentérica, perfuração de víscera oca e obstrução intestinal possuem a capacidade de progredir rapidamente para quadros graves de sepse, falência múltipla de órgãos e óbito caso o controle do foco infeccioso ou isquêmico sofra atrasos (SINGER M, et al., 2016). Entretanto, a investigação diagnóstica permanece desafiadora na rotina médica, uma vez que diferentes etiologias compartilham manifestações clínicas inespecíficas, enquanto subgrupos vulneráveis - incluindo idosos, gestantes e indivíduos imunossuprimidos

- frequentemente manifestam sintomas atípicos que mascaram apresentações clássicas da doença (BJM BEST PRACTICE, 2026).

Nesse contexto, a tomada de decisão clínica exige a integração criteriosa de dados da anamnese, achados laboratoriais e exames de imagem. Embora métodos avançados como ultrassonografia e tomografia computadorizada tenham ampliado a precisão diagnóstica nas últimas décadas, a interpretação correta desses exames ainda depende substancialmente da experiência do profissional (WOLFE JM, et al, 2022).

Consequentemente, esse processo está sujeito a uma considerável variabilidade interobservador, uma vulnerabilidade que se acentua criticamente em cenários de alta demanda assistencial e privação de sono nos plantões de emergência (NEWMAN-TOKER DE, et al., 2022). Além disso, a ocorrência de atrasos nesse fluxo interpretativo correlaciona-se diretamente com o pior prognóstico dos pacientes e o aumento substancial dos custos em saúde (NEWMAN-TOKER DE et al., 2022).

Paralelamente, a inteligência artificial (IA) tem emergido como uma das abordagens tecnológicas mais promissoras e disruptivas na área da saúde. Por meio de técnicas avançadas de aprendizado de máquina (machine learning) e aprendizado profundo (deep learning), algoritmos computacionais são capazes de processar grandes volumes de dados e identificar padrões complexos em registros clínicos, exames laboratoriais e dados radiológicos (TOPOL EJ, 2019).

Essa capacidade analítica atua de forma sinérgica com a prática médica, fornecendo subsídios para otimizar o suporte à tomada de decisão assistencial em cenários de alta complexidade. Evidências recentes indicam que o desempenho dessas ferramentas apresenta resultados promissores na classificação de doenças, na estratificação de risco e interpretação ágil de exames de imagem em cenários de urgência e emergência (KIRUBARAJAN A, et al., 2020).

No âmbito específico das condições abdominais agudas, modelos de IA vêm sendo investigados para auxiliar o diagnóstico de quadros como apendicite, obstrução intestinal, colecistite, isquemia mesentérica. Estudos recentes demonstram que essas ferramentas frequentemente apresentam níveis de acurácia, sensibilidade e especificidade diagnóstica comparáveis ou até mesmo superiores aos métodos convencionais de interpretação clínica e radiológica (HEEN AF, et al., 2023).

Apesar do crescente volume de publicações, as evidências permanecem fragmentadas entre diferentes patologias, métodos computacionais e contextos clínicos. Essa heterogeneidade metodológica e a falta de consenso sobre a aplicabilidade prática desses algoritmos criam uma lacuna na compreensão do real impacto assistencial e da segurança dessas tecnologias na rotina médica (RAGAB H, et al., 2026).

Diante desse cenário de fragmentação de evidências e da necessidade de otimização dos fluxos de urgência, torna-se essencial investigar as inovações tecnológicas direcionadas ao manejo do paciente crítico. Portanto, esta revisão integrativa tem como objetivo realizar uma síntese sobre o uso da inteligência artificial e do machine learning no diagnóstico diferencial, na estratificação de risco e no suporte à decisão clínica e cirúrgica em pacientes com abdômen agudo. Assim, busca-se avaliar o desempenho e a acurácia dessas ferramentas no diagnóstico diferencial, bem como seu potencial clínico no suporte à decisão clínica e cirúrgica, na predição de risco perioperatório e no manejo das complicações. Com isso, este estudo visa fornecer um panorama consolidado que fundamente a incorporação segura, eficaz e baseada em evidências dessas tecnologias na prática médica.

MÉTODOS

Desenho do estudo

Trata-se de uma revisão bibliográfica integrativa, que foi conduzida de acordo com as diretrizes Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Foram contempladas as etapas de formulação da pergunta, definição dos critérios de elegibilidade, estratégia de busca, seleção dos estudos, extração dos dados e síntese dos resultados.

Pergunta de pesquisa

Para estruturação da pergunta de pesquisa foi utilizada a estratégia PICO (acrônimo para population, intervention, comparison e outcomes) apresentada no **Quadro 1**, que levou a construção da seguinte pergunta norteadora: Quais são as aplicações da inteligência artificial no diagnóstico diferencial, na conduta, no risco perioperatório e complicações no abdômen agudo?"

Quadro 1 - Definição dos termos para estruturação da pergunta de pesquisa pelo acrônimo PICO.

P - População	**Pacientes com abdômen agudo**
I - Intervenção/ exposição	Aplicação de inteligência artificial ou machine learning
C- Comparador	Não se aplica
O – Outcome (desfecho)	Diagnóstico diferencial, acurácia diagnóstica, estratificação prognóstica, suporte à decisão clínica/cirúrgica, risco perioperatório e complicações pós-operatórias

Fonte: Moretti LG, et al., 2026.

Estratégia de busca

Inicialmente, foi realizada consulta aos Descritores em Ciências da Saúde (DeCS), para conhecimento dos descritores universais, bem como a tradução dos termos para o inglês contidos no Medical Subject Headings (MeSH), sendo selecionados os descritores: "machine learning", "diagnosis", "abdomen, acute", " surgical abdomen", " artificial intelligence", " deep learning", "diagnostic accuracy", e "differential diagnosis".

Após a seleção dos termos de busca, foram construídas as seguintes estratégias de busca utilizando os operadores booleanos "AND" e "OR", sendo a primeira: (machine learning) AND (diagnosis) AND (Abdomen, Acute) e a segunda: ((Acute Abdomen) OR (surgical abdomen)) AND ((Artificial Intelligence) OR (machine learning) OR (deep learning)) AND ((Diagnosis) OR (diagnostic accuracy) OR (differential diagnosis)).

Para busca, foi considerada a base de dados do United States National Library of Medicine – PubMed. Foram aplicados os filtros "free full text" e "2023-2026".

Critérios de seleção

Foram incluídos artigos delineados como ensaio clínico e ensaio clínico randomizado, além de estudos observacionais, retrospectivos e de coorte, sendo excluídos artigos de revisões da literatura, relatos e séries de caso, teses e livros. Ademais, outros critérios de inclusão foram o uso de Inteligência Artificial ou Machine Learning, bem como pacientes com abdômen agudo.

Dois avaliadores realizaram o processo de seleção dos artigos, de forma independente, a fim de realizar uma seleção mais eficaz. Esse processo foi realizado em duas etapas, sendo, na primeira, a seleção dos artigos pela leitura dos títulos e resumos, na segunda fase foi realizada a seleção dos artigos após a leitura dos textos completos para extração dos dados. Não foram utilizadas ferramentas de automação, sendo todas as exclusões realizadas manualmente.

Extração de dados e apresentação dos resultados

Para extração dos dados foram consideradas informações sobre: autores; ano de publicação; características da amostra; desenho do estudo; desfecho e dados estatísticos para o desfecho de interesse para cada grupo intervenção/ exposição e comparador/controle. Esses dados foram apresentados de forma descritiva em uma tabela (**Quadro 2**).

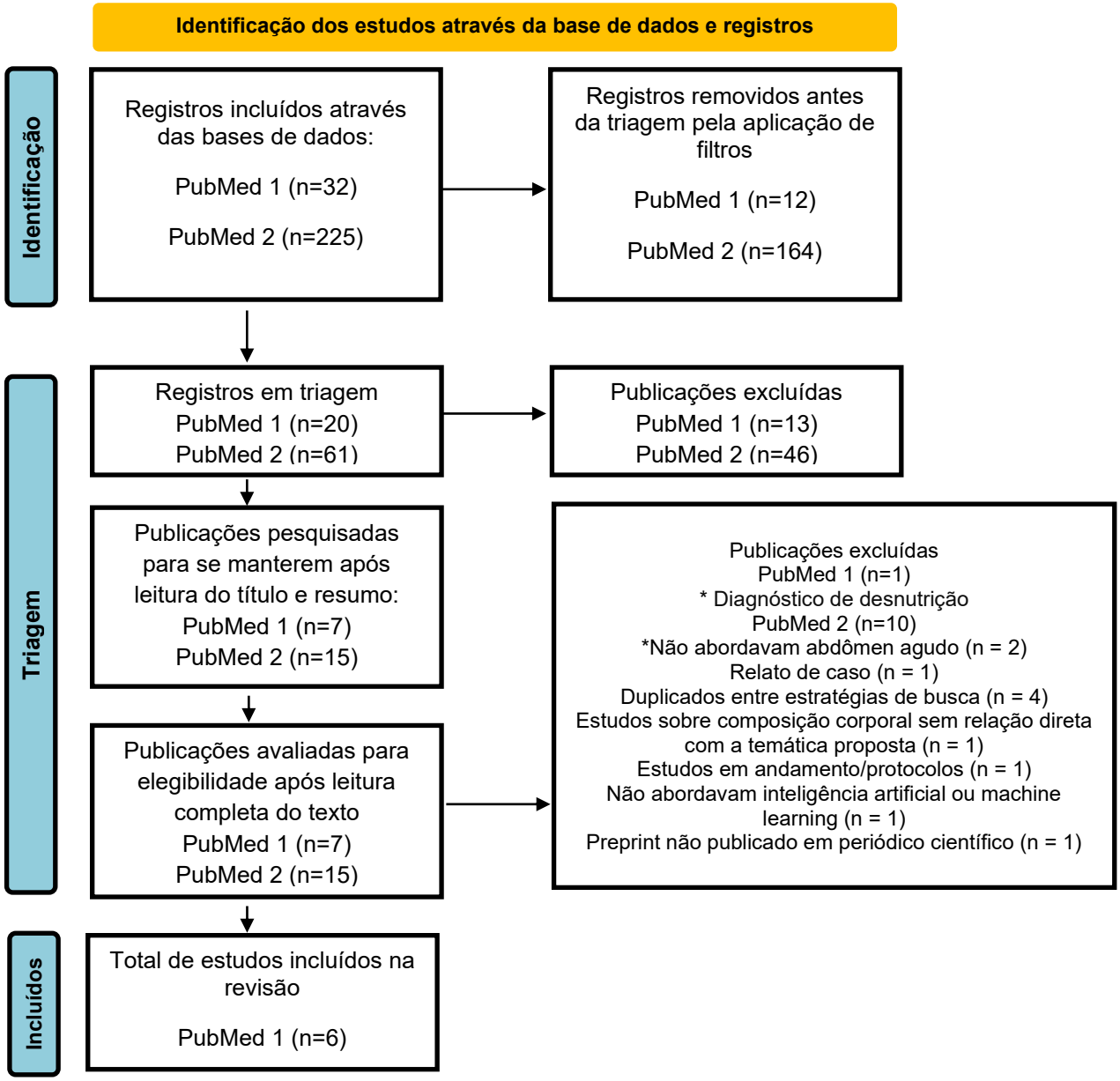
Além disso, foi realizada a classificação do nível de evidência científica dos estudos incluídos, de acordo com os critérios propostos pelo Joanna Briggs Institute (JBI), considerando o tipo de pergunta clínica e o desenho metodológico correspondente.

Os estudos foram categorizados em níveis hierárquicos de evidência, variando de nível 1 (maior nível de evidência, incluindo revisões sistemáticas e estudos experimentais ou de acurácia diagnóstica) a nível 5 (menor nível de evidência, incluindo opinião de especialistas e pesquisas básicas), conforme a estrutura específica para efetividade, diagnóstico, prognóstico e demais domínios aplicáveis.

RESULTADOS

Foram selecionados 11 estudos, com critérios de seleção e etapas expostas na **Figura 1**. O fluxograma de seleção foi adaptado conforme o modelo PRISMA.

Figura 1 - Fluxograma do processo de seleção dos artigos.



Fonte: Moretti LG, et al., 2026 com elaboração com base no protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).

A caracterização dos estudos foi reunida em forma de quadro-síntese no **Quadro 2**, e os desfechos explorados no **Quadro 3**.

Quadro 2 - Caracterização geral dos estudos selecionados.

Nº de referência e Citação	Desenho de Estudo	Nível de Evidência	Amostra	Comparador	Modelo de IA
(BETTE S, et al., 2024)	Caso-controle	3b	267 pacientes ≥18 anos com pancreatite aguda diagnosticada pelos critérios de Atlanta e TC abdominal contrastada de qualidade adequada.	745 pacientes de base de dados local submetidos à tomografia abdominal contrastada por dor abdominal superior de etiologia incerta, sem diagnóstico de pancreatite aguda.	Modelo híbrido radiômica + machine learning/statistical model.
(LIU Z, et al., 2025)	Estudo observacional retrospectivo	3b	2.541 pacientes (I-level), 2.628 pacientes (II-level) e 2.654 pacientes (III-level), ≥18 anos com abdômen agudo não traumático e duração dos sintomas por <7 dias.	Comparação interna entre cinco algoritmos de machine learning (Regressão Logística, DNN, SVM, Random Forest e XGBoost) por validação cruzada de 5 folds.	Algoritmos de machine learning (Regressão Logística, Redes Neurais Profundas, SVM, Random Forest e XGBoost) com seleção de atributos por Recursive Feature Elimination with Cross-Validation (RFECV).
(DU L, et al., 2026)	Estudo observacional retrospectivo	3b	383 pacientes ≥18 anos com abdômen agudo não traumático admitidos entre 2019 e 2024.	comparação interna via validação cruzada entre o grupo de treino, com n=269, e o grupo de validação, com n=114.	Seleção de variáveis por LASSO com validação cruzada de 10 folds integrada a regressão logística multivariada para construção de nomograma preditivo.
(WEI W, et al., 2024)	Estudo observacional retrospectivo	3b	322 pacientes ≥65 anos com apendicite aguda confirmada cirurgicamente (219 com apendicite não complicada e 103 com apendicite complicada) atendidos entre 2012 e 2022.	Comparação interna de desempenho entre 9 algoritmos diferentes de Machine Learning (LR, CART, RF, SVM, Bayes, KNN, NN, FDA e GBM).	Modelos de Machine Learning (Logistic Regression, CART, Random Forest, SVM, Naive Bayes, KNN, Neural Network, FDA e Gradient Boosting Machine [GBM]) com interpretação por SHAP.
(TAKAISHI T, et al., 2025)	Estudo observacional retrospectivo	3b	567 tomografias computadorizadas de pacientes com apendicite (2011–2020) para treinamento/validação, além de 100 casos de apendicite e 100 controles com abdômen agudo para teste externo temporal.	Avaliação independente por dois radiologistas experientes utilizando escala Likert de 5 pontos para classificação das detecções do modelo.	Deep Learning baseado em Faster Region-Based Convolutional Neural Network (Faster R-CNN) adaptada para detecção tridimensional (3D).

Nº de referência e Citação	Desenho de Estudo	Nível de Evidência	Amostra	Comparador	Modelo de IA
(XIANG K, SHANG D, 2025)	Estudo observacional retrospectivo	3b	1311 pacientes com pancreatite aguda submetidos à TC na admissão; 934 pacientes possuíam dados clínicos completos e 772 possuíam simultaneamente dados clínicos e de imagem para construção do modelo multimodal.	Comparação entre modelos de Deep Learning (DNN, CNN/VGG16 com mecanismo de atenção e modelo multimodal) e algoritmos tradicionais de Machine Learning (SVM).	Modelo Multi-View Deep Learning combinando Deep Neural Network (DNN) para dados clínicos e Convolutional Neural Network (VGG16 com mecanismo de atenção) para imagens de TC.
(CHEN B, et al., 2024)	Retrospectivo multicêntrico	3b	10.993 pacientes com cirurgia de emergência geral (EGS) abdominal de 5 centros na China; coorte de teste com 3.298 pacientes (30%); pacientes ≥18 anos com EGS abdominal.	Comparação interna entre cinco algoritmos de machine learning (XGBoost, Random Forest, Regressão Logística, SVM e Naive Bayes) por validação em coorte de teste independente	Algoritmos de machine learning (XGBoost, Random Forest [RF], Regressão Logística [LR], SVM, Naive Bayes [NB]) com seleção de features por análise de importância e RF como modelo ótimo.
(CHOI HU, et al., 2024)	Retrospectivo	3b	180 pacientes (média 29±9 anos; 91 feminino) com suspeita de apendicite aguda.	Comparação interna entre ultralow-dose CT (ULDCT) sem denoising e ULDCT com algoritmo de denoising baseado em imagem (D-ULDCT/IDLDA), avaliada por 6 radiologistas (3 board-certified + 3 residents).	Algoritmo de denoising baseado em imagem (IDLDA/QuantaStream Denoising) aplicado a CT de dose reduzida para melhoria de qualidade de imagem e performance diagnóstica.
(GAN T, et al., 2024)	Retrospectivo	3b	Pacientes ≥18 anos com dor abdominal aguda atendidos.	Comparação interna entre Logistic Regression e múltiplos algoritmos de machine learning incluindo Artificial Neural Network (ANN) por validação cruzada.	Modelos de machine learning incluindo Logistic Regression e Artificial Neural Network (ANN) com seleção de features para predição de triagem de dor abdominal aguda.
(PIEBPIEN P, et al., 2024)	Retrospectivo	3b	12.596 cirurgias abdominais; 404 casos de SSI (3.21%; 95%CI 2.91%–3.53%).	Comparação interna entre 4 algoritmos de machine learning (Naive Bayes, XGBoost, Random Forest, Decision Tree) e modelos estatísticos tradicionais (SENIC original e updated-SENIC-LR).	Algoritmos de machine learning (Naive Bayes [NB], XGBoost, Random Forest [RF], Decision Tree [DT]) com 14 predictors vs modelos estatísticos tradicionais (SENIC com 6 predictors).
(HAMEDI Z, et al., 2025)	Retrospectivo	3b	286 imagens de ultrasound abdominal com segmentation masks e organ-level labels do Bern University Hospital Inselspital.	Avaliação qualitativa do desempenho do SAMed fine-tuned vs abordagem tradicional de segmentação manual por operadores medicos.	SAMed (Segment Anything Model for medical imaging) fine-tuned com transfer learning, pré-trained em larger public US dataset antes de fine-tuning no dataset específico.

Fonte: Moretti LG, et al., 2026.

Quadro 3 - Síntese dos desfechos avaliados, métricas de desempenho e aplicabilidade clínica dos modelos de IA.

Ref	Aplicação clínica	Desfecho	Desempenho	Conclusão
1	Diagnóstico de pancreatite aguda em pacientes com dor abdominal superior de etiologia incerta	Diagnóstico de PA por detecção de alterações radiômicas imperceptíveis à avaliação visual convencional (textura de imagem)	AUC=0,932 (IC95% 0,852–1,00)	Bom desempenho diagnóstico, comparável à lipase elevada e quadro clínico clássico
2	Pré-diagnóstico e triagem médica automatizada de pacientes com abdômen agudo não traumático em serviços de emergência	Classificação hierárquica de doenças em três níveis diagnósticos utilizando apenas dados clínicos iniciais e sinais físicos	AUC >0,90 para categorias diagnósticas principais (I-level) e AUC >0,80 para subgrupos diagnósticos (II-level)	Excelente capacidade de triagem inicial, com desempenho semelhante entre modelos complexos e regressão logística
3	Predição precoce do risco de sepse em pacientes internados com quadro de abdômen agudo	Identificação de risco de sepse (Sepsis-3) com base em 6 preditores independentes: AKI, choque, TCO_2, hipoalbuminemia, NLR elevado e linfocitopenia	AUC = 0,929 (IC95% 0,898–0,959) no treino e AUC = 0,934 (IC95% 0,891–0,977) na validação	Excelente capacidade de estratificação precoce do risco de sepse em pacientes com abdômen agudo
4	Diferenciação pré-operatória entre apendicite complicada (CA) e não complicada (UA) em pacientes idosos com apendicite aguda	Classificação de apendicite complicada utilizando 12 variáveis clínicas, laboratoriais e radiológicas, incluindo duração da dor abdominal, peritonite, temperatura corporal e bilirrubina total	O algoritmo GBM apresentou sensibilidade de 0,9167 e alta precisão discriminatória geral	Excelente desempenho na identificação de apendicite complicada em idosos, com o modelo GBM demonstrando maior robustez e potencial para auxiliar a tomada de decisão clínica precoce
5	Detecção e localização tridimensional automatizada de apendicite em tomografias computadorizadas de pacientes com abdômen agudo	Identificação da localização do apêndice inflamado por meio de delineamento de boxes tridimensionais em exames de TC	Precisão = 0,66 (87/132); Sensibilidade (Recall) = 0,87 (87/100); taxa de falsos positivos = 0,23 por paciente	Bom desempenho discriminatório na detecção de lesões locais, demonstrando alta sensibilidade diagnóstica e robustez anatômica, embora apresente precisão moderada devido a falsos positivos em alças intestinais vizinhas
6	Classificação precoce da gravidade da pancreatite aguda (leve, moderadamente grave e grave) utilizando dados clínicos e tomografia computadorizada na admissão	Predição da gravidade da pancreatite aguda (MAP, MSAP e SAP) por meio de modelo multimodal integrando informações clínicas e radiológicas	Modelo multimodal apresentou acurácia global de 80,26%. Por subgrupo, a acurácia foi de 91,69% para MAP, 64,90% para MSAP e 75,56% para SAP, superando os modelos de visão única e algoritmos tradicionais	Modelo multimodal apresentou melhor desempenho que modelos clínicos ou radiológicos isolados na estratificação da gravidade da pancreatite aguda
7	Calculador de risco peri-operatório para mortalidade pós-operatória em pacientes com	Predição de mortalidade pós-operatória utilizando 35 variáveis pré-operatórias incluindo exames	Random Forest apresentou AUC = 0.8961, F1 = 0.1284, Sensibilidade = 0.9500 e Especificidade = 0.8422, superando XGBoost (AUC = 0.8504), LR	Excelente capacidade de predição de mortalidade pós-operatória em EGS abdominal, com Random

Ref	Aplicação clínica	Desfecho	Desempenho	Conclusão
	cirurgia de emergência geral (EGS) abdominal em serviços de emergência.	laboratoriais, scores de severidade e características demográficas.	(AUC = 0.7893), SVM (AUC = 0.7705) e NB (AUC = 0.8024). RF superou significativamente scores tradicionais: ESS (AUC = 0.7510), APACHE II (AUC = 0.7147) e ASA (AUC = 0.6747). Top 5 features: BUN level, preoperative sepsis, AAST EGS grading, ALB, age.	Forest demonstrando superioridade sobre outros algoritmos de ML e scores tradicionais de risco cirúrgico.
8	Diagnóstico de apendicite aguda utilizando CT de dose ultralow com algoritmo de denoising para reduzir exposição à radiação ≥50% sem perda de qualidade diagnóstica.	Classificação diagnóstica de apendicite aguda e avaliação de qualidade de imagem (noise, SNR, CNR, acceptability diagnóstica) entre ULDCT e D-ULDCT.	D-ULDCT apresentou AUC = 0.94–0.97 vs ULDCT AUC = 0.90–0.97. Melhora significativa de AUC para residents: +0.06 (p=0.022) e +0.05 (p=0.046). D-ULDCT demonstrou noise significativamente menor (p<0.001), SNR e CNR significativamente maiores (p<0.001), e melhor acceptability diagnóstica para todos 6 radiologistas.	Excelente melhoria na qualidade de imagem e performance diagnóstica de CT ultralow-dose para apendicite aguda, com benefício particular para radiologistas menos experientes, permitindo redução de dose ≥50% sem comprometer diagnóstico.
9	Triagem rápida e pré-diagnóstico de pacientes com dor abdominal aguda em serviços de emergência, especialmente em contextos com recursos limitados.	Classificação de triagem de dor abdominal aguda utilizando 7 preditores principais: histórico de diabetes, histórico de stroke, pulse, blood pressure, pale appearance, bowel sounds e localização da dor.	Artificial Neural Network (ANN) apresentou accuracy de 97.92%, superando Logistic Regression e outros modelos de ML testados.	Excelente capacidade de triagem rápida de dor abdominal aguda com ANN demonstrando superior acurácia, permitindo decisão clínica rápida em emergency departments com recursos limitados.
10	Predição de risco de infecção de sítio operacional (SSI) em pacientes após cirurgia abdominal para monitoramento preventivo.	Identificação de SSI (surgical site infection) utilizando predictors clínicos tradicionais (SENIC) vs 14 predictors incluindo novos fatores.	Updated-SENIC-LR (6 predictors) apresentou C-statistic = 0.768 (95%CI 0.745–0.790), superando significativamente todos algoritmos de ML: NB (C = 0.679), XGBoost (C = 0.675), RF (C = 0.656) e DT (C = 0.651). SENIC original teve C = 0.668. Overfitting detectado especialmente em DT, RF e XGBoost.	Modelos de machine learning NÃO superaram modelos estatísticos tradicionais para predição de SSI em cirurgia abdominal. Updated-SENIC-LR e NB demonstraram melhor performance para monitoramento de risco SSI, com MLs apresentando overfitting problemático.
11	Segmentação e classificação automatizada de órgãos e free fluid em ultrasound abdominal para suporte diagnóstico em emergency departments.	Identificação, delineamento e classificação de múltiplos órgãos abdominais e detecção de free fluid em imagens de US para decisão clínica rápida em emergência.	SAMed fine-tuned demonstrou capacidade de identificar e classificar múltiplos órgãos com sucesso. Segmentação de free fluid apresentou oportunidades para melhoria (métricas específicas não relatadas, avaliação apenas qualitativa). Transfer learning funcionou como solução confiável para lidar com dataset pequeno.	Bom desempenho do SAMed fine-tuned para segmentação de ultrasound abdominal com identificação multifocal de órgãos, mas segmentação de free fluid necessita melhoria. Transfer learning resolve efetivamente problema de datasets pequenos em aplicações médicas de IA.

Fonte: Moretti LG, et al., 2026.

DISCUSSÃO

A Inteligência Artificial (IA) tem sido incorporada progressivamente à prática médica, gerando discussões sobre seu papel como ferramenta de apoio ou potencial substituta de determinadas atividades. No contexto da Radiologia, Rowe SP, et al. (2022) destaca que a principal questão não é a substituição do profissional radiologista, mas a capacidade da IA de aumentar a eficiência e a precisão na execução de tarefas. Esse cenário também se estende a outras especialidades médicas, principalmente na cirúrgica.

Os estudos incluídos nesta revisão demonstram que técnicas de aprendizado de máquina (Machine Learning) e aprendizado profundo (Deep Learning) apresentam potencial para auxiliar o diagnóstico diferencial do abdômen agudo, embora as evidências utilizadas sejam predominantemente observacionais e retrospectivas. Ao integrar informações da anamnese, exame físico, exames laboratoriais e métodos de imagem, esses algoritmos podem aumentar a acurácia diagnóstica, aprimorar a estratificação de risco e fornecer suporte à tomada de decisão clínica (BETTE S, et al., 2024; DU L, et al., 2026; TAKAISHI T, et al., 2025).

Dentre os 11 artigos selecionados, foi demonstrado desempenho promissor, embora com limitações metodológicas, no uso de IA associada a Machine Learning em casos de abdômen agudo, desde diagnósticos clínicos até predição cirúrgica. Contudo, apenas um obteve alta taxa de detecção; (CHEN B, et al., 2024) o único estudo multicêntrico com uma das maiores amostras, o qual desenvolveu uma calculadora de risco perioperatório baseada em Inteligência Artificial e Machine Learning para pacientes que foram submetidos à cirurgia geral de emergência (CGE) por doenças abdominais. Seu principal objetivo foi construir um modelo capaz de predizer precocemente a gravidade clínica e o risco de complicações pós-operatórias, auxiliando cirurgiões na tomada de decisões antes e durante o procedimento cirúrgico.

Nesse estudo, foram analisados mais de 10 mil pacientes provenientes de 5 centros hospitalares chineses e, entre os desfechos avaliados, estavam mortalidade hospitalar, pneumonia pós-operatória, infecção de sítio cirúrgico, trombose e necessidade de ventilação mecânica por mais de 48 horas. Ademais, para chegar a elevada acurácia, foi necessário comparar cinco algoritmos de Machine Learning, dos quais apenas um se sobressaiu, o modelo Random Forest, o qual demonstrou maior capacidade discriminatória para identificar pacientes com maior risco de evolução desfavorável.

Um dos aspectos mais relevantes deste estudo foi a comparação do Machine Learning com ferramentas prognósticas tradicionais amplamente utilizadas na prática clínica – Emergency Surgery Score (ESS), APACHE II e ASA, além de integrar dados clínicos, laboratoriais e a classificação anatômica da doença, conseguindo, assim, um estudo completo sobre os casos, evidenciando que a IA possui capacidade de auxiliar a medicina tradicional. Além do auxílio na tomada de decisão em casos dúbios e complexos, a calculadora vem a ser uma opção de embasamento na comunicação com familiares, subsidiando decisões em casos em que não há sinais físicos clássicos ou imagens patognomônicas referentes à condição apresentada.

Em contrapartida, Wei W, et al. (2024) desenvolve uma análise unicêntrica de um modelo de Machine Learning que prevê apendicite complicada em pacientes idosos, com foco em auxiliar a decisão clínica sobre a gravidade da doença e a necessidade de intervenção cirúrgica precoce, sendo sua principal motivação o alto risco de mortalidade da apendicite complicada.

O estudo incluiu dados clínicos, laboratoriais e de imagem, assim como Chen B, et al. (2024). Ambos alimentaram nove algoritmos com variáveis, no entanto enquanto em um o Random Florest se destacou (CHEN B, et al., 2024), em outro o modelo Gradient Boosting Machine (GBM) se sobressaiu (WEI W, et al, 2024), posto que conseguiu identificar com elevada precisão os pacientes com maior probabilidade de apresentar apendicite complicada, cujo grupo tende a necessitar de cirurgia de urgência. Com isso, ainda que com diferentes caminhos de desenho de estudo e modelos de Machine Learning, ambos (WEI W, et al, 2024; CHEN B, et al., 2024) chegam a resultados satisfatórios, uma vez que auxiliam os médicos a realizarem cirurgias de forma mais detalhada e com decisão precisa.

Já outro estudo (WEI W, et al, 2024) priorizou pacientes de alto risco, indicando cirurgias mais rapidamente, reduzindo diagnósticos tardios e, o melhor de todos os resultados, individualizando o tratamento do paciente. Chen B, et al. (2024), em seus estudos, possibilita que o cirurgião adapte a estratégia terapêutica, priorize pacientes mais graves, planeje recursos hospitalares, indique monitorização quando necessário e personalize o manejo perioperatório.

Outrossim, outro estudo (PIEBPIEN P, et al 2024) avaliou um modelo predição de risco de infecção do sítio cirúrgico após cirurgias abdominais, comparando seu desempenho com um modelo clínico tradicional (SENIC) atualizado, no qual foram incorporados novos fatores de risco, além dos originalmente utilizados. Foram desenvolvidos 4 modelos de Machine Learning, os quais utilizaram apenas dados clínicos e perioperatórios, excluindo dados de exames laboratoriais e de imagem. Em todos os algoritmos, o tempo cirúrgico superior a 2 horas foi consistentemente o fator mais importante, exercendo maior influência sobre o risco de infecção cirúrgica. Por fim, o modelo Naïve Bayes (índice C 0,679) apresentou o melhor desempenho.

Contudo, nenhum algoritmo de Machine Learning superou o modelo clínico tradicional atualizado SENIC-LR, cujo índice C atingiu 0,768, indicando que, neste trabalho, os métodos tradicionais ainda apresentam excelentes desempenhos quando bem calibrados. Além disso, sabe-se que, na prática, os exames de imagem desempenham um papel fundamental no diagnóstico e planejamento cirúrgico, por isso é suposto que aliar o Machine Learning e a IA, para leitura de imagens, com os modelos tradicionais, pode auxiliar na formação de modelos preditivos mais completos, precisos e úteis para o planejamento cirúrgico e a prevenção de complicações.

Ademais, a análise realizada por Gan T, et al. (2024) utiliza apenas sinais clínicos, facilmente obtidos na avaliação inicial, permitindo uma triagem rápida sem depender de exames laboratoriais ou de imagem, para prever precocemente a gravidade de pacientes com dor abdominal aguda, com o fito de identificar, rapidamente, aqueles com maior risco de evolução crítica e possível intervenção cirúrgica. O estudo de (GAN T, et al 2024) também defende, assim como na análise de Piebpien P, et al. (2024), que, mesmo com resultados excelentes da Machine Learning apenas com sinais clínicos, é fundamental o auxílio de exames de imagens e laboratoriais. Todavia, os resultados Gan T, et al. (2024) identificam que apenas com a Rede Neural Artificial (ANN), utilizando somente dados clínicos, podem prever, com elevada precisão, a gravidade de pacientes com dor abdominal aguda, evitando realizações de laparoscopias desnecessárias, atrasos nas indicações cirúrgicas e elevação da mortalidade.

Em contraste, o estudo realizado por Liu Z, et al. (2025), que desenvolveu uma estrutura de Machine Learning para pré-diagnóstico do abdômen agudo não traumático utilizando apenas informações clínicas superficiais, conseguiu identificar rapidamente doenças potencialmente graves, como apendicite, obstrução intestinal e gravidez ectópica, sendo possível priorizar pacientes com predição cirúrgica rapidamente, sem necessitar de exames de imagem ou laboratoriais. Assim como as análises de Gan T, et al. (2024) e de Piebpien P, et al. (2024) não elimina totalmente a utilização de exames complementares para melhorar o diagnóstico e as decisões clínicas, mas os algoritmos utilizados conseguiram, com alta precisão, reconhecer condições cirúrgicas ainda durante a triagem inicial, funcionando como uma ferramenta de estratificação precoce da gravidade, capaz de auxiliar na organização do fluxo de pacientes e na definição de prioridades assistenciais.

Entre os artigos relacionados que associam IA e análise de exames de imagem (BETTE S, et al 2024; CHOI HU, et al 2024; HAMEDI Z, et al 2025; TAKAISHI T, et al 2025) todos utilizaram Deep Learning, uma subárea da Inteligência Artificial e do Machine Learning que utiliza redes neurais profundas. A utilização desse modelo fez com que esses estudos obtivessem uma elevada taxa de precisão, uma vez que a análise dos exames de imagem, com o auxílio de Deep Learning, consegue identificar estruturas anatômicas complexas, apontar uma localização objetiva da lesão, melhorar a qualidade das imagens e interpretá-las automaticamente aumentando a precisão diagnóstica, diferentemente da observação convencional, baseada exclusivamente na interpretação visual do radiologista, o que torna o exame dependente da compreensão de cada observador, gerando uma variabilidade. Isto é, a análise tradicional não possui o mesmo nível de precisão da realizada com o auxílio de Deep Learning, posto que cada radiologista, segundo os estudos, pode ter uma visão variada de uma mesma imagem.

Como concluído por Issaiy M, et al. (2023), a utilização da IA, no contexto de abdômen agudo, demonstrou seu potencial para melhorar a precisão diagnóstica e as capacidades prognósticas, exibindo uma mudança na forma como se aborda as condições médicas críticas, posto que, com a IA e o Machine Learning, é possível tomar decisões com mais veracidade e criar um atendimento individualizado para cada paciente. Além disso, o acurado desempenho de vários modelos de IA ressalta sua relevância na clínica e na cirurgia, excedendo os métodos diagnósticos tradicionais.

Ademais, em consonância ao editorial de Rowe SP, et al. (2022), a Inteligência Artificial, em conjunto com o Machine Learning, auxilia as atividades médicas, principalmente no campo cirúrgico, corroborando com os estudos analisados. Os resultados desta revisão indicam que os modelos de aprendizado de máquina podem reduzir erros na abordagem do abdômen agudo ao fornecer suporte à tomada de decisão clínica e cirúrgica. Conforme discutido por Heen AF, et al. (2023), esses algoritmos são capazes de integrar informações clínicas, laboratoriais e de imagem para identificar pacientes com maior probabilidade de necessitar de intervenção cirúrgica. Esse suporte torna-se especialmente relevante em serviços de emergência, onde a elevada demanda assistencial pode favorecer atrasos diagnósticos e falhas na priorização dos casos.

Embora os resultados deste estudo demonstrem o potencial da IA para aprimorar o diagnóstico e auxiliar na tomada de decisão cirúrgica, sua incorporação à prática clínica deve ser acompanhada de importantes considerações éticas. Segundo a World Health Organization (2021) Ethics and Governance of Artificial Intelligence for Health, o desenvolvimento e a utilização dessas tecnologias devem priorizar princípios como transparência, proteção dos direitos humanos, privacidade dos dados e supervisão humana. Dessa forma, a implementação de sistemas de inteligência artificial no manejo do abdômen agudo deve ocorrer de forma responsável, garantindo que o suporte fornecido pelos algoritmos complemente, e não substitua, o julgamento clínico do cirurgião.

Nesse contexto, a inteligência artificial deve ser compreendida como uma ferramenta de apoio à decisão, cuja utilização depende da supervisão contínua do cirurgião. A decisão terapêutica deve permanecer fundamentada na avaliação clínica individualizada, utilizando as informações fornecidas pelos algoritmos como complemento ao raciocínio médico, e não como substituição da responsabilidade profissional.

Apesar do elevado desempenho demonstrado pelos modelos de inteligência artificial na identificação e estratificação de pacientes com abdômen agudo, sua implementação na prática clínica ainda representa um desafio. De acordo com o artigo de Ragab H, et al (2026), a incorporação dessas tecnologias depende da padronização dos fluxos assistenciais, da transparência dos algoritmos, da definição de diretrizes regulatórias e da participação ativa dos profissionais de saúde durante sua utilização. Além disso, o desenvolvimento de sistemas capazes de integrar informações clínicas, laboratoriais e de imagem, aliado ao respeito aos princípios éticos, poderá ampliar a aplicabilidade da inteligência artificial e favorecer sua incorporação de forma segura à tomada de decisão cirúrgica.

Limitações

Embora os resultados desta revisão demonstrem o potencial da inteligência artificial para apoiar o diagnóstico diferencial e a tomada de decisão cirúrgica no abdômen agudo, a interpretação desses achados deve considerar algumas limitações metodológicas dos estudos incluídos. Observa-se o predomínio de estudos observacionais retrospectivos, classificados como nível de evidência 3b, o que reduz a capacidade de estabelecer relações causais e aumenta a suscetibilidade a vieses inerentes a esse delineamento.

Além disso, muitos modelos foram desenvolvidos a partir de bases de dados de instituições únicas e submetidos apenas à validação interna, limitando sua aplicabilidade em diferentes contextos clínicos. A heterogeneidade dos algoritmos empregados, das populações avaliadas e dos desfechos analisados também dificulta comparações diretas entre os estudos e a padronização das evidências disponíveis. Também vale ressaltar que a busca foi aplicada apenas a uma única base de dados, sendo dada preferência para artigos disponíveis gratuitamente, fazendo com que se estabelecesse uma limitação metodológica do presente estudo.

Dessa forma, embora os resultados sejam promissores, sua incorporação à prática clínica deve ser acompanhada por estudos prospectivos, multicêntricos e com validação externa, capazes de confirmar a reprodutibilidade e a segurança desses sistemas em diferentes cenários assistenciais.

CONSIDERAÇÕES FINAIS

A revisão integrativa evidenciou que a inteligência artificial e o machine learning apresentam potencial como ferramentas complementares no diagnóstico diferencial, na estratificação de risco e no suporte à decisão clínica e cirúrgica em pacientes com abdômen agudo. Os estudos analisados demonstraram desempenho promissor na identificação de doenças abdominais agudas, predição de sepse, mortalidade e complicações, além de apoio à interpretação de exames de imagem. No entanto, a predominância de estudos retrospectivos, com validação interna e heterogeneidade metodológica, limita a generalização dos achados. Portanto, são necessários estudos prospectivos, multicêntricos e com validação externa antes da incorporação ampla dessas tecnologias na prática assistencial. A partir da análise desses estudos, a inteligência artificial poderá ser incorporada de maneira mais segura e eficaz à prática assistencial no manejo do abdômen agudo.

REFERÊNCIAS

1. BETTE S, et al. Radiomics-Based Machine Learning Model for Diagnosis of Acute Pancreatitis Using Computed Tomography. Diagnostics (Basel), 2024; 14(7): 718.
2. BMJ Best Practice. Abdominal pain. BMJ Best Practice, 2026.
3. BRODER J, et al. Guidelines for Reasonable and Appropriate Care in the Emergency Department 2 (GRACE-2): Low-risk, recurrent abdominal pain in the emergency department. Academic Emergency Medicine, 2022; 29(7): 869-882.
4. CHEN B, et al. Machine learning based peri-surgical risk calculator for abdominal related emergency general surgery: a multicenter retrospective study. International Journal of Surgery, 2024; 110(6): 3527-3535.
5. CHOI HU, et al. Diagnostic performance and image quality of an image-based denoising algorithm applied to radiation dose-reduced CT in diagnosing acute appendicitis. Abdominal Radiology, 2024; 49(6): 1839-1849.
6. DU L, et al. Development and validation of a machine learning-based nomogram model integrating LASSO regression and multiple evaluation methods for predicting early sepsis risk in acute abdomen. European Journal of Medical Research, 2026; 31: 464.
7. GAN T, et al. Machine learning based prediction models for analyzing risk factors in patients with acute abdominal pain: a retrospective study. Frontiers in Medicine, 2024; 11: 1354925.
8. HAMEDI Z, et al. AI-Based Analysis of Abdominal Ultrasound Images to Support Medical Diagnosis in Emergency Departments. Studies in Health Technology and Informatics, 2025; 325: 16-21.
9. HEEN AF, et al. Machine Learning for Decision-Support in Acute Abdominal Pain: Proof of Concept and Central Considerations. Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren, 2023.
10. ISSAIY, M., et al., Artificial Intelligence and Acute Appendicitis: A Systematic Review of Diagnostic and Prognostic Models. World Journal of Emergency Surgery, 2023; 18(1): 59.
11. KIRUBARAJAN A, et al. Artificial intelligence in emergency medicine: A scoping review. Journal of the American College of Emergency Physicians Open, 2020; 1(6): 1691-1702.
12. LIU Z, et al. Intelligent prediagnosis for nontraumatic acute abdomen with surface-level information using machine learning. Science Progress, 2025; 108(2): 368504251350763.
13. NEWMAN-TOKER DE, et al. Diagnostic Errors in the Emergency Department: A Systematic Review. Diagnosis, 2022.

14. PIEBPIEN P, et al. Can machine learning models improve the prediction of surgical site infection in abdominal surgery than traditional statistical models? Journal of International Medical Research, 2024; 52(11): 3000605241293696.
15. RAGAB H, et al. Artificial Intelligence in Abdominal MRI Diagnostics: Current Applications, Challenges, and Future Perspectives. Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren, 2026.
16. ROWE SP, et al. The future of radiology: What if artificial intelligence is really as good as predicted? Diagnostic and Interventional Imaging, Volume 103, Issue 9, 2022.
17. SARTELLI M, et al. The management of intra-abdominal infections from a global perspective: 2017 WSES guidelines for management of intra-abdominal infections. World Journal of Emergency Surgery, 2017; 12: 29.
18. SINGER M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA, 2016; 315(8): 801-810.
19. TAKAISHI T, et al. Deep learning for appendicitis: development of a three-dimensional localization model on CT. Japanese Journal of Radiology, 2025; 43(11): 1859-1869.
20. TOPOL EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 2019; 25(1): 44-56.
21. XIANG K, SHANG D. Construction of a Multi-View Deep Learning Model for the Severity Classification of Acute Pancreatitis. Discovery Medicine, 2025; 37(192): 73-92.
22. WEI W, et al. Construction of a clinical prediction model for complicated appendicitis based on machine learning techniques. Scientific Reports, 2024; 14(1): 16473.
23. WORLD HEALTH ORGANIZATION. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: World Health Organization, 2021.
24. WOLFE JM, et al. Abdominal Pain in the Emergency Department: How to Select the Correct Imaging for Diagnosis. Open Access Emergency Medicine, 2022; 14: 789-800.