World Journal of Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

World J Gastroenterol 2025 November 21; 31(43): 112000

DOI: 10.3748/wjg.v31.i43.112000

ISSN 1007-9327 (print) ISSN 2219-2840 (online)

REVIEW

Artificial intelligence in acute appendicitis: A comprehensive review of machine learning and deep learning applications

Sami Akbulut, Zeynep Kucukakcali, Cemil Colak

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's classification

Scientific Quality: Grade A, Grade A, Grade A, Grade B, Grade B

Novelty: Grade A, Grade A, Grade B, Grade B, Grade B

Creativity or Innovation: Grade A, Grade B, Grade B, Grade B, Grade B

Scientific Significance: Grade B, Grade B, Grade B, Grade B, Grade B

P-Reviewer: Chen XY, PhD, Professor, China; Pogorelic Z, PhD, Professor, Croatia; Xu M, PhD, China

Received: July 15, 2025

Revised: August 26, 2025

Accepted: October 14, 2025

Published online: November 21, 2025

Processing time: 128 Days and 10.3 Hours

Sami Akbulut, Surgery and Liver Transplantation, Inonu University Faculty of Medicine, Malatya 44280, Türkiye

Sami Akbulut, Zeynep Kucukakcali, Cemil Colak, Biostatistics and Medical Informatics, Inonu University Faculty of Medicine, Malatya 44280, Türkiye

Co-corresponding authors: Sami Akbulut and Cemil Colak.

Corresponding author: Sami Akbulut, MD, Professor, Surgery and Liver Transplantation, Inonu University Faculty of Medicine, Elazig Yolu 10. Km, Malatya 44280, Türkiye.
akbulutsami@gmail.com

Abstract

Acute appendicitis (AAp) remains one of the most common abdominal emergencies, requiring rapid and accurate diagnosis to prevent complications and unnecessary surgeries. Conventional diagnostic methods, including medical history, clinical assessment, biochemical markers, and imaging techniques, often present limitations in sensitivity and specificity, especially in atypical cases. In recent years, artificial intelligence (AI) has demonstrated remarkable potential in enhancing diagnostic accuracy through machine learning (ML) and deep learning (DL) models. This review evaluates the current applications of AI in both adult and pediatric AAp, focusing on clinical data-based models, radiological imaging analysis, and AI-assisted clinical decision support systems. ML models such as random forest, support vector machines, logistic regression, and extreme gradient boosting have exhibited superior diagnostic performance compared to traditional scoring systems, achieving sensitivity and specificity rates exceeding 90% in multiple studies. Additionally, DL techniques, particularly convolutional neural networks, have been shown to outperform radiologists in interpreting ultrasound and computed tomography images, enhancing diagnostic confidence. This review synthesized findings from 65 studies, demonstrating that AI models integrating multimodal data including clinical, laboratory, and imaging parameters further improved diagnostic precision. Moreover, explainable AI approaches, such as SHapley Additive exPlanations and local interpretable model-agnostic explanations, have facilitated model transparency, fostering clinician trust in AI-driven decision-making. This review highlights the advancements in AI for AAp diagnosis, emphasizing that AI is used not only to establish the diagnosis of AAp but also to differentiate complicated from uncomplicated cases. While preliminary results are promising, further prospective, multicenter studies are required for

large-scale clinical implementation, given that a great proportion of current evidence derives from retrospective designs, and existing prospective cohorts exhibit limited sample sizes or protocol variability. Future research should also focus on integrating AI-driven decision support tools into routine emergency care workflows.

Key Words: Acute appendicitis; Complicated appendicitis; Artificial intelligence; Machine learning; Deep learning; Decision support systems; Explainable artificial intelligence; Predictive modeling; Diagnosis

©The Author(s) 2025. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: This comprehensive review explores the emerging role of artificial intelligence (AI), including machine learning and deep learning techniques, in diagnosing acute appendicitis (AAp). Despite advancements in imaging and clinical scoring, diagnosing AAp remains challenging, particularly in atypical cases. AI models such as random forests, support vector machines, and convolutional neural networks have demonstrated promising results in enhancing diagnostic accuracy and decision-making. In addition to aiding in the differential diagnosis of AAp from other causes of acute abdominal pain, AI approaches have also been applied to distinguish between complicated and uncomplicated appendicitis, thereby supporting risk stratification and guiding management strategies. The review discusses current evidence, potential benefits, and limitations of integrating AI-based decision support into clinical practice. These insights may pave the way for more precise, timely, and individualized management of AAp.

Citation: Akbulut S, Kucukakcali Z, Colak C. Artificial intelligence in acute appendicitis: A comprehensive review of machine learning and deep learning applications. *World J Gastroenterol* 2025; 31(43): 112000
URL: https://www.wjgnet.com/1007-9327/full/v31/i43/112000.htm
DOI: https://dx.doi.org/10.3748/wjg.v31.i43.112000

INTRODUCTION

Acute appendicitis (AAp) is one of the most common acute abdominal emergencies worldwide, and timely and accurate diagnosis is crucial for optimal patient management[1-3]. AAp usually develops as a result of obstruction of the appendiceal lumen, which subsequently initiates the inflammatory cascade. This blockage may arise from appendicoliths, appendiceal tumors, intestinal parasites, or hypertrophied lymphoid tissue. The underlying etiology varies with age: In children, lymphoid hyperplasia related to heightened immune activity is the most frequent trigger, whereas in older adults, fecalith-induced obstruction is most often observed, although other causes should not be disregarded[4,5]. These age-dependent etiologies not only shape how the disease presents clinically but also influence how it should be evaluated and managed across different patient populations. Epidemiological data indicate a lifetime incidence ranging from 5% to 20%, with a risk of 8.6% in men and 6.7% in women[6-9]. Interestingly, while men are more likely to develop the disease, women have a higher probability of undergoing appendectomy (23% *vs* 12%), underscoring possible diagnostic discrepancies and gender-specific clinical pathways[1,7,10]. This discrepancy may be attributed to several factors, including a more atypical symptom presentation observed in women, the diagnostic challenges arising from overlapping gynecological conditions such as ovarian cysts, pelvic inflammatory disease, and ectopic pregnancy, and the tendency toward lower surgical thresholds when diagnostic uncertainty persists in female patients[11,12]. These factors collectively contribute to higher false diagnostic rates in female patients compared to males[13].

Traditional diagnostic approaches encompass a combination of patient history, physical examination, and evaluation of biochemical markers such as white blood cell count, bilirubin, and C-reactive protein. Imaging modalities, including ultrasonography (US) and computed tomography (CT), are routinely employed, while magnetic resonance imaging (MRI) is preferentially used in specific populations such as pregnant patients. In addition, clinical scoring systems serve as valuable adjuncts to improve diagnostic accuracy and guide clinical decision-making[3,14-17]. These scoring systems include: Alvarado; Eskelinen; Ohmann; Appendicitis inflammatory response (AIR); Raja Isteri Pengiran Anak Saleha Appendicitis; Pediatric appendicitis score (PAS); Adult appendicitis score; Tzanakis; Lintula; Fenyo-Lindberg; Karaman, and others[14,16,18-23]. For instance, among established clinical scoring systems, the AIR score has demonstrated utility in severity stratification, with a recent pediatric study showing that a score ≥ 9 distinguishes perforated from non-perforated AAp with 89.5% sensitivity, 71.9% specificity, and an area under the curve (AUC) of 0.80, establishing it as a clinically accessible reference standard for complicated AAp assessment[24]. In adults, the Alvarado score (cutoff ≥ 8) demonstrated moderate accuracy in distinguishing AAp from negative appendectomy, with reported sensitivity of 72.9%, specificity of 70.6%, and an AUC of 0.782[25]. When applied to distinguish complicated from uncomplicated AAp, the Alvarado score with a cutoff of ≥ 6 achieved 80.6% sensitivity, 44.5% specificity, and an AUC of 0.605[26].

Despite the widespread use of traditional diagnostic tools, accurately distinguishing between a normal appendix and the spectrum of appendiceal inflammation- ranging from uncomplicated to complicated (perforated) AAp- remains a significant clinical challenge, particularly in atypical presentations, pediatric patients, and pregnant women, where diagnostic nuances and imaging limitations complicate decision-making[3,9,15]. While conventional approaches are often effective, their sensitivity and specificity vary widely across patient populations. To address these limitations, artificial

intelligence (AI), including its subdomain machine learning (ML), and more specifically deep learning (DL), models have emerged as promising tools, capable of integrating multimodal clinical, laboratory, and radiological data to improve diagnostic accuracy and risk stratification[16,17,27-30]. Importantly, false-negative or false-positive diagnostic outcomes whether they originate from conventional clinical tools or from AI-supported systems can lead to serious clinical consequences such as perforation or unnecessary surgery, thereby increasing patient morbidity and healthcare burden [31-34].

In recent years, AI and ML techniques have gained significant importance not only in the initial diagnosis of AAp but also in accurately differentiating uncomplicated cases from those that are complicated such as perforated or gangrenous AAp as well as in assessing disease severity, guiding treatment planning, and predicting postoperative complications[3,8,16,27-30,33,35-84]. AI-based models, particularly ensemble learning approaches, not only improve diagnostic accuracy but also support clinical decision-making by reducing unnecessary surgeries and identifying high-risk perforation cases earlier. DL algorithms can accurately diagnose AAp from radiological images, while ML-based models effectively analyze laboratory data and patient characteristics to predict the risk of perforation[31,33,43,60]. Notably, ML and DL models such as random forest (RF), support vector machine (SVM), gradient boosting machine (GBM), and convolutional neural networks (CNNs) have achieved higher accuracy rates than traditional diagnostic methods in AAp diagnosis[47,85].

Recent literature highlights the diagnostic superiority of various ML models across diverse populations and clinical settings. Erman *et al*[35] developed ML models for pediatric patients that achieved 76.4% accuracy and an area under the receiver operating characteristic (AUROC) of 0.79 for detecting perforation, and 70.1% accuracy with an AUROC of 0.77 for grading severity. Gollapalli *et al*[38] demonstrated that bagging and stacking ensemble methods, particularly k-nearest neighbors (KNN) and decision tree-based models, achieved up to 92.6% accuracy and F1 scores above 90% when combined with up sampling techniques. In a resource-limited setting, Phan-Mai *et al*[46] showed that GBM distinguished complicated AAp with an AUROC of 0.858, performing better than other classifiers like SVM or artificial neural networks (ANN). Additional high-performance models include the gaussian naive bayes used by Roshanaei *et al*[36], yielding 95% accuracy, and the GBM-based model in Wei *et al*'s study[40], which reached 95.56% accuracy with a high sensitivity (91.67%) and specificity (97.39%). These findings reinforce the growing utility of ML in clinical AAp decision-making, particularly when tailored with proper sampling and algorithm selection. Issaiy *et al*[28] conducted a systematic review of 29 studies, most of which addressed diagnostic applications of AI in AAp. ANNs were frequently used and demonstrated a high performance, with accuracy rates commonly exceeding 80%, and AUROC values reaching up to 0.985. Despite these promising results, most studies suffered from selection bias and lacked internal validation. This reinforces the growing utility of ML in clinical AAp decision-making, particularly when tailored with proper sampling and algorithm selection. While the study provides a broad overview of AI integration across the entire spectrum of AAp management, our review offers a more focused and technically detailed analysis specifically on ML and DL methodologies for diagnosis, with in-depth evaluation of model architectures, performance metrics, imaging modalities, and emerging areas such as radiomics, explainable AI (XAI), and multimodal data fusion[16].

This review evaluates AI applications in AAp diagnosis, highlighting their clinical impact, comparative performance, and future implications for emergency medicine and surgical decision-making. In particular, ensemble learning techniques and hybrid models have been highlighted as key approaches to improving diagnostic sensitivity[38]. The integration of AI into radiological imaging (US, CT, MRI) and clinical data is analyzed in terms of diagnostic performance metrics such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and future research directions are discussed.

Since 2023, six systematic reviews[5,16,27,28,81,82] and two narrative reviews[9,80] have examined the application of AI in AAp, addressing diagnostic, therapeutic, and prognostic dimensions. In contrast, our study differs by providing a more focused and technically detailed evaluation of ML and DL methodologies, particularly for the diagnosis of AAp. Furthermore, it distinctly incorporates the correct use of AI, ML, and DL terminologies, an aspect often overlooked in prior reviews. Unlike these earlier studies, our study emphasizes algorithmic architectures, imaging methods, and XAI approaches, thereby offering a complementary perspective tailored for researchers and clinicians interested in diagnostic applications. Finally, this study uniquely considers the application of AI architectures by analyzing a total of 65 studies covering both adult and pediatric AAp research[3,8,17,29,30,33,35-79,84,86-98]. Table 1 provides detailed information on the referenced studies, including the AI models and their performance metrics, serving as a consolidated reference for comparative evaluation.

FUNDAMENTAL TERMINOLOGY IN AI-BASED MODELS

AI is transforming modern medicine by enabling advanced data analysis, pattern recognition, and predictive decision-making across a wide range of clinical specialties. From diagnostic imaging to electronic health record analysis, personalized treatment planning, differential diagnosis, and disease classification, AI technologies have demonstrated growing utility in enhancing clinical workflows, improving diagnostic accuracy, and supporting evidence-based decisions. The key concepts and subfields of AI most commonly used or actively researched in clinical medicine are as follows[27,28,80-82,99].

ML models

ML enables computers to identify patterns in data and make decisions without explicit programming. It includes supervised, unsupervised, and reinforcement learning (RL) approaches. Supervised algorithms such as RF, SVM, GBMs, and extreme gradient boosting (XGBoost) have demonstrated strong performance in various medical classification tasks,

Table 1 Summary of artificial intelligence models for acute appendicitis: Methods and performance metrics

No.	Ref.	Year	Country	Dataset size	Variables used	AI methods	Performance metrics
1	Sibic *et al*[84]	2025	Turkey	AAp: 400; non-AAp: 400	Demographic, and radiological data [CT images (CNN architectures)]	MobileNet v2, ResNet v2, EfficientNet b2, Inception v3 (MobileNet v2 best results)	Accuracy: 79.1; precision: 82.0; sensitivity: 74.7; F1 score: 78.1; AUC: 0.877
2	Navaei *et al*[17]	2025	Iran	AAp: 465; non-AAp: 317	Demographic, clinical and biochemical data	DT, RF, SVM, KNN, GBM, AdaBoost, XGBoost, LightBoost, CatBoost (RF best results)	Accuracy: 94.6; sensitivity: 93.9; specificity: 95.7; F1 score: 93.6
3	Li *et al*[8]	2025	China	Compl AAp: 88; uncompl AAp: 213	Demographic, clinical and biochemical data	LR, SVM, RF, DT[1], GBM, KNN, GNB, MLP (RF best results)	Accuracy: 81.0; sensitivity: 76.0; specificity: 83.0; F1 score: 74.0; AUC: 0.840
4	Kucukakcali *et al* [86]	2025	Turkey	Compl AAp: 34; uncompl AAp: 65; non-AAp: 41	Demographic and biochemical data	SGB (non-AAp *vs* AAp)	Accuracy: 96.3; sensitivity: 94.7; specificity: 100; F1 score: 97.3; AUC: 0.947
						SGB (uncompl *vs* compl AAp)	Accuracy: 78.9; sensitivity: 83.3; specificity: 76.9; F1 score: 71.4; AUC: 0.790
5	Kucukakcali *et al* [87]	2025	Turkey	Compl AAp: 183; uncompl AAp: 290; negative AAp: 117	Demographic and biochemical data	AdaBoost, XGBoost, SGB, bagged CART, RF (XGBoost best results)	Accuracy: 80.0; sensitivity: 70.8; specificity: 85.4; F1 score: 72.3
						AdaBoost, XGBoost, SGB, bagged CART, RF (XGBoost best results)	Accuracy: 90.7; sensitivity: 100; specificity: 61.5; F1 score: 94.3
6	Kim *et al*[29]	2025	South Korea	Compl AAp: 655; uncompl AAp: 2789; negative AAp: 55[1]; non-AAp: 3058	CT images (non *vs* uncomplicated)	3D-CNN (transfer learning, ResNet/DenseNet/EfficientNet) (DenseNet best results)	Accuracy: 79.5; sensitivity: 70.1; specificity: 87.6; AUC: 0.865
					CT images (complicated *vs* uncomplicated)	3D-CNN (transfer learning, ResNet/DenseNet/EfficientNet) (DenseNet best results)	Accuracy: 76.1; sensitivity: 82.6; specificity: 74.2; AUC: 0.827
7	Kendall *et al*[88]	2025		Compl AAp: 119[2]; uncompl AAp: 344; non-AAp: 317	Demographic, clinical, biochemical and radiological data	RF, LightGBM, LR, SGD, KNN, Dummy, GANDALF, RF + embedded LightGBM (best result)	Accuracy: 98.1; sensitivity: 97.8; specificity: 96.1; AUROC: 0.993
						RF, LightGBM, LR, SGD, KNN, Dummy, GANDALF, LightGBM + filter FS (best result)	Accuracy: 90.1; sensitivity: 78.8; specificity: 95.1; AUROC: 0.931
8	Erman *et al*[35]	2025	Canada	Compl AAp: 602; uncompl AAp: 1378	Demographic, clinical and biochemical data	ML pipeline	Accuracy: 70.1; NPV: 82.8; PPV: 56.4
9	Chen *et al*[3]	2025	China	Compl AAp: 357; uncompl AAp: 416	Demographic, clinical and biochemical data	XGBoost, RF, DT (CART), SVM (XGBoost best results)	Accuracy: 85.5; sensitivity: 86.5; specificity: 84.6; AUC: 0.914
10	Aydin *et al*[89]	2025	Turkey	Compl AAp: 296; uncompl AAp: 3658; non-AAp: 4632; validation: Compl AAp: 1580; Uncompl AAp: 1287; Non-AAp: 169	Demographic, clinical, biochemical and radiological data	LR, KNN, SVM, CART, RF (RF best results for AAp diagnosis)	Accuracy: 99.2; sensitivity: 99.8; specificity: 99.3; AUC: 0.996
						LR, KNN, SVM, CART, RF (RF best results for severity of AAp)	Accuracy: 99.2; sensitivity: 99.3; specificity: 99.1; AUC: 0.995

11	Zhao *et al*[90]	2024	China	Compl AAp: 258; uncompl AAp: 76	Demographic, clinical, biochemical and radiological data (CT images)	Radiomics model (CT images), CT model (clinical and CT features), combined model	Accuracy: 75.4; sensitivity: 74.6; specificity: 82.6; AUC: 0.817
12	Yazici *et al*[37]	2024	Turkey	Compl AAp: 142; uncompl AAp: 990	Demographic, clinical and biochemical data	KNN, DT, LR, SVM, MLP, GNB (LR best result)	Accuracy: 96.0; sensitivity: 60.0; specificity: 100
13	Wei *et al*[40]	2024	China	Compl AAp: 103; uncompl AAp: 219	Demographic, clinical and biochemical data	LR, CART, FR, SVM, Bayes, KNN, NN, FDA, GBM (GBM best result)	Accuracy: 95.6; sensitivity: 91.7; specificity: 97.4; F1 score: 93.0
14	Schipper *et al*[33]	2024	Netherlands	AAp: 167; non-AAp: 169	Data including physical examination	XGBoost	AUC: 0.919
					Data including physical examination and biochemical data	XGBoost	AUC: 0.923
15	Roshanaei *et al*[36]	2024	Iran	AAp: 138; non-AAp: 396	Demographic, clinical and biochemical data	GNB	Accuracy: 95.0; sensitivity: 87.2; specificity: 97.5; F1 score: 89.0
16	Marcinkevičs *et al*[52]	2024	Germany	Compl AAp: 97; uncompl AAp: 482	Radiological data (US images) (diagnosis)	CBM; MVCBM; SSMVCBM	AUROC: 0.800; AUPR: 0.920
					Radiological data (US images) (severity)	CBM; MVCBM; SSMVCBM	AUROC: 0.780; AUPR: 0.580
17	Males *et al*[39]	2024	Croatia	Compl AAp: 252; uncompl AAp: 252; negative AAp: 47 (pediatric cases)	Demographic, clinical and biochemical data	RF	Sensitivity: 99.7; specificity: 17.0
						XGBoost	Sensitivity: 99.8; specificity: 12.0
						LR	Sensitivity: 99.7; specificity: 5.2
18	Liang *et al*[91]	2024	China	Training cohort: Compl AAp: 236; uncompl AAp: 464; validation cohort: Compl AAp: 182; uncompl AAp: 283	Demographic, clinical, biochemical and radiological data	Conventional combined model (clinical + CT features); deep learning radiomics (DL + radiomics) our combined model (clinical + CT + DL + radiomics) radiologist's diagnosis	Accuracy: 79.0; sensitivity: 66.5; specificity: 85.3; AUC: 0.816
							Accuracy: 72.5; sensitivity: 70.2; specificity: 73.9; AUC: 0.799
19	Gollapalli *et al*[38]	2024	Saudi Arabia	411 patients[3]	Demographic, clinical and biochemical data	DT (experiment 1)	Accuracy: 75.0; sensitivity: 13.8; precision: 40.0; F1 score: 20.5
						KNN (experiment 1)	Accuracy: 83.1; sensitivity: 41.4; precision: 75.0; F1 score: 53.3
						DT (experiment 2)	Accuracy: 87.4; sensitivity: 91.2; precision: 83.8; F1 score: 87.4
						KNN (experiment 2)	Accuracy: 84.7; sensitivity: 84.6; precision: 83.7; F1 score: 84.2
						KNN bagging (experiment 3)	Accuracy: 92.1; sensitivity: 91.2; precision: 92.2; F1 score: 91.7
						DT bagging (experiment 3)	Accuracy: 89.5; sensitivity: 83.5;

							precision: 93.8; F1 score: 88.4
						Stacking (experiment 4)	Accuracy: 92.6; sensitivity: 89.0; precision: 95.3; F1 score: 92.0
20	Chadaga *et al*[42]	2024	India	AAp: 465; non-AAp: 317 (pediatric cases)	Demographic, clinical and biochemical data	RF, LR, DT, KNN, AdaBoost, CatBoost, LightGBM, XGBoost, APPSTACK. Bayesian optimization, hybrid bat algorithm, hybrid self-adaptive bat algorithm, firefly algorithm, grid search, randomized search (hybrid bat algorithm with APPSTACK best results)	Accuracy: 94.0; sensitivity: 74.0; precision: 85.0; F1 score: 78.0; AUC: 0.960
21	Abu-Ashour *et al* [41]	2024	Canada	AAp: 2100 (pediatric cases)	Ultrasound reports	Human	Precision: 57.3; sensitivity: 88.1; F score: 69.4
						ChatGPT (large language model)	Precision: 92.3; sensitivity: 68.4; F score: 78.5
					Operative reports	Human	Precision: 59.2; sensitivity: 95.3; F score: 73.1
						ChatGPT (large language model)	Precision: 97.1; sensitivity: 75.8; F score: 85.1
22	Phan-Mai *et al*[46]	2023	Vietnam	Compl AAp: 483; uncompl AAp: 1467	Demographic, clinical and biochemical data	SVM (SMOTE-adjusted)	Accuracy: 65.5; AUC: 0.730
						DT (SMOTE-adjusted)	Accuracy: 73.8; AUC: 0.738
						KNN (SMOTE-adjusted)	Accuracy: 74.1; AUC: 0.831
						LR (SMOTE-adjusted)	Accuracy: 72.9; AUC: 0.789
						ANN (SMOTE-adjusted)	Accuracy: 74.2; AUC: 0.810
						GBM (SMOTE-adjusted)	Accuracy: 82.0; AUC: 0.890
23	Pati *et al*[30]	2023	India	Compl AAp: 51[4]; uncompl AAp: 196; non-AAp: 183 (pediatric cases)	Demographic, clinical, biochemical and radiological data	LR, NB, KNN, SVM, DT, RF, MLP, AdaBoost (RF best for diagnostic)	Accuracy: 91.6; precision: 89.0; sensitivity: 92.0; specificity: 91.3; F1 score: 90.4
						LR, NB, KNN, SVM, DT, RF, MLP, AdaBoost (AdaBoost best for complication prediction)	Accuracy: 92.2; precision: 94.6; sensitivity: 96.3; specificity: 68.6; F1 score: 95.4
24	Park *et al*[45]	2023	South Korea	AAp: 246; non-AAp: 215; diverticulitis: 254	CT images	CNN-EfficientNet algorithm (single image method)	Accuracy: 86.1; precision: 85.4; sensitivity: 85.6; specificity: 86.5; AUC: 0.937
					CT images	CNN-EfficientNet algorithm (RGB method)	Accuracy: 87.9; precision: 87.1; sensitivity: 87.9; specificity: 88.1; AUC: 0.951
25	Lin *et al*[93]	2023	Taiwan	Compl AAp: 49; uncompl AAp: 362	Demographic, clinical, biochemical and radiological data	9 different MLP-ANN analyzed (Lin *et al*[93] ANN model best results)	AUC: 0.897; sensitivity: 85.7; specificity: 91.7
26	Li *et al*[92]	2023	China	Compl AAp: 141; uncompl AAp: 201 (pregnant patients)	Demographic, clinical, biochemical and	DT	AUC: 0.780

					radiological data		
27	Harmantepe *et al* [44]	2023	Turkey	AAp: 189; negative AAp: 156	Demographic and biochemical data	LR, SVM, NN, KNN, voting classifier (voting best result)	Accuracy: 86.2; sensitivity: 83.7; specificity: 88.6
28	Akbulut *et al*[43]	2023	Turkey	Compl AAp: 304; uncompl AAp: 1161; negative AAp: 332	Demographic and biochemical data	CatBoost + SHAP (non-AAp *vs* AAp)	Accuracy: 88.2; sensitivity: 84.2; specificity: 93.2; F1 score: 88.7; AUC: 0.947
						CatBoost + SHAP (compl *vs* uncompl AAp)	Accuracy: 92.0; sensitivity: 94.1; specificity: 90.5; F1 score: 91.1; AUC: 0.969
29	Xia *et al*[51]	2022	China	Compl AAp: 148; uncompl AAp: 150	Demographic and clinical data	SVM	Accuracy: 83.6; sensitivity: 81.7; specificity: 85.3; Matthews: 0.6732
30	Su *et al*[49]	2022	United States	AAp: 28002; non-AAp: 655 (adult cases)	Demographic and clinical data	LR	Accuracy: 96.0; sensitivity: 73.0; specificity: 68.0; AUC: 0.780
						RF	Accuracy: 97.0; sensitivity: 67.0; specificity: 71.0; AUC: 0.750
				AAp: 11128; non-AAp: 256 (pediatric cases)	Demographic and clinical data	LR	Accuracy: 95.0; sensitivity: 81.0; specificity: 78.0; AUC: 0.870
						RF	Accuracy: 96.0; sensitivity: 82.0; specificity: 75.0; AUC: 0.860
31	Shikha and Kasem[48]	2023	Brunei	Compl AAp: 25; uncompl AAp: 24; negative AAp: 97 (pediatric cases)	Demographic, Clinical, and biochemical data	AI pediatric appendicitis DT	Accuracy: 97.1; sensitivity: 96.7; specificity: 97.4
32	Mijwil and Aggarwal[47]	2022	Iraq	Appendectomy: 318[5]; medical: 307	Demographic, and biochemical data	RF, LR, NB, GLM, DT, SVM, GBT (RF best results)	Accuracy: 83.8; precision: 84.1; sensitivity: 81.1; specificity: 81.0
33	Akgül *et al*[50]	2021	Turkey	Compl AAp: 45; uncompl AAp: 147; negative AAp: 24; non-AAp: 106 (pediatric cases)	Demographic, clinical, biochemical and radiological data	ANN	Sensitivity: 89.8; specificity: 81.2; AUC: 0.910
34	Marcinkevics *et al* [53]	2021	Germany	Compl AAp: 51; uncompl AAp: 196; non-AAp: 183 (pediatric cases)	Demographic, clinical, biochemical and radiological data	LR (diagnostic)	Sensitivity: 88.0; specificity: 76.0; AUC: 0.910
						RF (diagnostic)	Sensitivity: 91.0; specificity: 86.0; AUC: 0.960
						GBM (diagnostic)	Sensitivity: 93.0; specificity: 86.0; AUC: 0.960
						LR (severity)	Sensitivity: 93.0; specificity: 42.0; AUC: 0.820
						RF (severity)	Sensitivity: 98.0; specificity: 45.0; AUC: 0.900
						GBM (severity)	Sensitivity: 97.0; specificity: 46.0; AUC: 0.900

35	Aparicio *et al*[79]	2021	Switzerland	AAp: 430 (pediatric cases)	Demographic, clinical, and biochemical data	SLIM risk model	AUC: 0.850; AUPR: 0.900
36	Hayashi *et al*[55]	2021	Japan	AAp: 70 videos (pediatric cases)	70 videos (between 85-347 images per video)	U-net-based CNN	Not indicated
37	Reismann *et al*[56]	2021	Germany	AAp: 29	Gene expression data (56.666 gene)	LR-based biomarker signature (4 genes)	AUC: 0.84
38	Ghareeb *et al*[54]	2021	Egypt	319	Clinical findings. Chronic diseases. Patient characteristics. Laboratory and imaging	Ensemble model (subspace KNN)	AUC: 0.82; accuracy: 91.1
39	Stiel *et al*[57]	2020	Germany	Compl AAp: 102; uncompl AAp: 234; negative AAp: 12; non-AAp: 115 (pediatric cases)	Demographic, clinical, biochemical and radiological data	Modified HAS based CART, AI score based RF (AAp *vs* nonoperative)	Sensitivity: 86.6; specificity: 70.9; AUC: 0.920
						Modified HAS based CART, AI score based RF (uncompl *vs* compl AAp)	Sensitivity: 97.1; specificity: 17.9; AUC: 0.710
40	Akmese *et al*[58]	2020	Turkey	AAp: 214; non-AAp: 214	Demographic and biochemical data	RF, CART, SVM, LR, KNN, ANN, GB (GB best results)	Accuracy: 95.3; sensitivity: 93.2; specificity: 97.1
41	Aydin *et al*[59]	2020	Turkey	Control: 4244; negative AAp: 169; compl AAp: 1559; uncompl AAp: 1272 (pediatric cases)	Demographic and biochemical data	KNN, NB, DT, SVM, GLM, RF (RF best results)	Accuracy: 97.5; sensitivity: 97.8; specificity: 97.2; AUC: 0.997
42	Rajpurkar *et al* [60]	2020	United States	AAp: 359; non-AAp: 287	CT images	Average of 2D Res-Net18, average of 2D Res-Net34, LRCN Res-Net18, LRCN Res-Net34, SE-ResNeXt-50, AppendiXNet (3D-ResNet CNN)	Accuracy: 72.5; sensitivity: 78.4; specificity: 66.7; AUC: 0.810
43	Park *et al*[61]	2020	United States	AAp: 215; non-AAp: 452	CT images	3D-CNN + grad-CAM	Accuracy: 91.5; sensitivity: 90.2; specificity: 92.0
44	Zhao *et al*[63]	2020	China	AAp: 48; non-AAp: 86	Midstream urine samples	Urinary proteomics + RF, SVM, NB (RF best results)	Accuracy: 83.6; sensitivity: 81.2; specificity: 84.4
45	Ramirez-garcialuna *et al* [62]	2020	Mexico	AAp: 51; non-AAp: 17; negative AAp: 3; control: 51	Demographic, clinical biochemical, radiological and infrared thermal data	Infrared thermography + RF classifier	Accuracy: 92.3; sensitivity: 90.0; specificity: 96.1; AUC: 0.906
46	Reismann *et al*[65]	2019	Germany	Compl AAp: 183; uncompl AAp: 290; negative AAp: 117 (pediatric cases)	Signature appendiceal diameter CRP leukocytes neutrophils	CRP, leukocytes, neutrophils, linear model (LBFGS) (AAp *vs* non-AAp)	Accuracy: 90.0; sensitivity: 93.0; specificity: 67.0; AUC: 0.910
						CRP, leukocytes, neutrophils, linear model (LBFGS) (compl *vs* uncompl AAp)	Accuracy: 51.0; sensitivity: 95.0; specificity: 33.0; AUC: 0.800
47	Kang *et al*[64]	2019	South Korea	AAp: 80; non-AAp: 164	Demographic, clinical biochemical and radiological data	Alvarado, AAS, Eskelinen, DT based CHAID algorithm	AUC: 0.850
48	Gudelis *et al*[66]	2019	Spain	AAp: 93; non-AAp: 159	Demographic, clinical biochemical and radiological data	ANN	AUC: 0.950; PCC: 93.5
						CHAID	AUC: 0.930; PCC: 81.7
49	Shahmoradi *et al* [67]	2018	Iran	AAp: 133; negative AAp: 48	Demographic, clinical and biochemical data	MLP	Accuracy: 92.9; sensitivity: 80.0; specificity: 97.5; AUC: 0.832
						RBFN	Accuracy: 77.6;

							sensitivity: 28.0; specificity: 87.8
						LR	Accuracy: 83.9; sensitivity: 58.3; specificity: 93.2; AUC: 0.808
50	Jamshidnezhad *et al*[69]	2017	Iran	NA	Demographic, clinical biochemical and radiological data	ACSS, MLNN, SVM, NN, hybrid fuzzy model, evolutionary–fuzzy + HBRC	Accuracy: 89.9
51	Afshari Safavi *et al*[68]	2015	Iran	Compl AAp: 24; uncompl: 59; negative AAp: 17	Demographic, and biochemical data	ANN (MLP)	Accuracy: 88.0; sensitivity: 97.6; AUC: 0.875
52	Park and Kim[70]	2015	South Korea	Compl AAp: 62; uncompl AAp: 143; non-AAp: 596	Demographic, clinical and radiological data	MLNN	Accuracy: 97.8; sensitivity: 96.6; specificity: 99.5
						RBF	AUC: 99.8; sensitivity: 99.7; specificity: 100
						PNN	AUC: 99.4; sensitivity: 98.1; specificity: 100
53	Lee *et al*[75]	2013	Taiwan	AAp: 464; negative-AAp: 110	Demographic, clinical and biochemical data	PEL, SVM, SMOTE, MCC, CM, WCUS, Alvarado (PEL best results)	Sensitivity: 57.3; specificity: 66.7; AUC: 0.619
54	Iliou *et al*[94]	2013	Greece	AAp: 71 Non-AAp: 236 (pediatric cases)	Demographic, clinical and biochemical data	K^{1}, JRip, bagging ensemble (majority voting)	Accuracy: 87.8
55	Deleger *et al*[95]	2013	United States	AAp: 534; control: 1566	Components of the pediatric appendicitis score	NLP	Sensitivity: 86.9; precision: 86.8; specificity: 93.8
56	Yoldaş *et al*[71]	2012	Turkey	AAp: 132; negative-AAp: 24	Demographic, clinical and biochemical data	ANN	Sensitivity: 100; specificity: 97.2; AUC: 0.950
57	Son *et al*[76]	2012	South Korea	AAp: 152; non-AAp: 174	Demographic, clinical and biochemical data	DT C5.0 model (univariate)	Accuracy: 80.2; sensitivity: 82.4; specificity: 78.3; AUC: 0.803
						DT C5.0 model (multivariate)	Accuracy: 73.5; sensitivity: 66.0; specificity: 80.0; AUC: 0.730
58	Malley *et al*[96]	2012	United States	AAp: 85; negative AAp: 21	Biochemical data	b-NN, class RF, Iboost, LR, KNN, regRF (regRF best results)	Brier score: 0.061; AUC: 0.976
59	Grigull and Lechner[74]	2012	Germany	AAp: 45 (pediatric cases)	Demographic, clinical and biochemical data	SVM, ANN, fuzzy logic, voting algorithm (combination best results)	Accuracy: 97.4
60	Hsieh *et al*[72]	2011	Taiwan	Compl AAp: 28; uncompl AAp: 87; negative AAp: 11; non-AAp: 65	Demographic, clinical and biochemical data	RF, SVM, ANN, LR (RF best results)	Accuracy: 96.0; sensitivity: 94.0; specificity: 100; AUC: 0.980
61	Ting *et al*[77]	2010	Taiwan	Compl AAp: 80; uncompl: 340; negative-AAp: 112	Demographic, clinical and biochemical data	DT	Sensitivity: 94.5; specificity: 80.5
62	Prabhudesai *et al* [73]	2008	United Kingdom	AAp: 24; non-AAp: 36	Demographic, clinical and biochemical data	Alvarado (≥ 7), Alvarado (≥ 6), clinical, ANN (ANN best results)	Sensitivity: 100; specificity: 97.2; PPV: 96.0; NPV: 100
63	Sakai *et al*[78]	2007	Japan	AAp: 86; negative AAp: 12; non-AAp: 71	Demographic, clinical and biochemical data	LR	Sensitivity: 21.4; specificity: 80.4; AUC: 0.719
						ANN	Sensitivity: 19.9; specificity: 78.5; AUC: 0.741

64	Pesonen *et al*[98]	1996	Finland	Suspected AAp: 911	Demographic, clinical and biochemical data	NN (ART1)	Sensitivity: 79.0; specificity: 78.0
						NN (SOM)	Sensitivity: 55.0; specificity: 83.0
						NN (LVQ)	Sensitivity: 87.0; specificity: 90.0
						NN (BP)	Sensitivity: 83.0; specificity: 92.0
65	Forsström *et al* [97]	1995	Finland	AAp: 145; negative AAp: 41	Biochemical data	LR	AUC: 0.678
						DiagaiD	AUC: 0.683
						NN (BP)	AUC: 0.622

[1]Three diagnostic models (clinical, computed tomography-radiomics, clinical-radiomics fusion) developed and validated using eight machine learning algorithms (logistic regression, support vector machine, random forest, decision tree, gradient boosting, k-nearest neighbor, gaussian naive bayes, multi-layer perceptron); best performance with random forest in the fusion model.
[2]I wish to emphasize that this manuscript is not related to medical terminology, and thus the percentages I reported are solely based on inferences.
[3]The authors stated that they made a distinction between 'complicated and non-complicated appendicitis' in this study; however, no information regarding the dataset was provided.
[4]This study contains serious clinical terminological errors, and it seems evident that the authors lack a proper understanding of appendicitis.
[5]Another major error, likely due to the lack of clinicians in the study, is the inconsistency of reporting 625 specimens examined while simultaneously stating that 307 patients were not operated. It is unclear how specimens could have been obtained from non-operated patients.
AAp: Acute appendicitis; Compl: Complicated; Uncompl: Uncomplicated; CRP: C-reactive protein; CT: Computed tomography; US: Ultrasonography; CNN: Convolutional neural network; DT: Decision tree; RF: Random forest; SVM: Support vector machine; KNN: K-nearest neighbor; GBM: Gradient boosting machine; AdaBoost: Adaptive boosting; XGBoost: Extreme gradient boosting; LightBoost: Light boosting; CatBoost: Categorical boosting; LR: Logistic regression; GNB: Gaussian naive bayes; MLP: Multi-layer perceptron; SGB: Stochastic gradient boosting; CART: Classification and regression trees; 3D: Three dimensional; LightGBM: Light gradient boosting machine; SGD: Stochastic gradient descent; FS: Feature selection; ML: Machine learning; FDA: Flexible discriminant analysis; NN: Neural network; CBM: Concept bottleneck models; MVCBM; Multiview concept bottleneck model; SSMVCBM: Semi-supervised Multiview concept bottleneck model; DL: Deep learning; SMOTE: Synthetic minority oversampling technique; NB: Naive bayes; RGB: Red, green, blue; ANN: Artificial neural networks; SHAP: SHapley Additive exPlanations; AI: Artificial intelligence; SLIM: Supersparse linear integer models; HAS: Heidelberg appendicitis score; GB: Gradient boosting; 2D: Two dimensional; LRCN: Long-term recurrent convolutional network; grad-CAM: Gradient-weighted class activation mapping; LBFGS: Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm; CHAID: χ^2d automatic interaction detection; RBFN: Radial basis function network; ACSS: Alvarado clinical scoring system; MLNN: Multilayer neural network structure; HBRC: Honeybee reproduction cycle; RBF: Radial basis function; PNN: Probabilistic neural network; PEL: Preclustering-based ensemble learning; MCC: Multiple-classifier committee; CM: Cluster-medoid; WCUS: Within-cluster under-sampling; NLP: Natural language processing; Iboost: Logit boost; regRF: Regression random forests; ART1: Binary adaptive resonance theory; SOM: Kohonen self-organizing map; LVQ: Learning vector quantization; BP: Back-propagation; DiagaiD: A connectionist approach to determine the diagnostic value of clinical data; AUC: Area under the curve; AUROC: Area under the receiver operating characteristic; NPV: Negative predictive value; PPV: Positive predictive value; PCC: Percent correctly classified; AUPR: Area under precision–recall; GLM: Generalized linear model.

including AAp diagnosis. Other models like adaptive boosting, light GBM (LightGBM), categorical boosting (CatBoost), and extra trees have also shown potential in clinical prediction and feature selection, although their use remains less frequent in AAp-specific studies[37,100,101].

Decision tree

Decision tree is a fundamental supervised learning algorithm that builds a hierarchical tree-like structure to classify data points based on feature values. Decision trees are widely used in medicine due to their transparency and interpretability, especially in clinical decision support systems. While they can function independently, they also serve as the foundational base for more complex ensemble learning methods such as RF and GBMs. Despite their simplicity, decision trees can effectively model non-linear relationships and are often favored in clinical settings where explainability is essential[102,103].

Lazy learning algorithms

KNN is a non-parametric, instance-based classification algorithm that predicts outcomes by comparing new data points to the most similar cases in the training set. It has been applied in various clinical tasks and has shown utility in AAp risk prediction using structured clinical data. While simple and interpretable, its application in high-dimensional datasets can be computationally expensive[38].

DL models

A specialized branch of ML, DL uses ANNs to process large and complex datasets, particularly for image and text analysis. Common architectures include multi-layer perceptron (MLP), CNNs for image classification, recurrent neural networks (RNNs) and long short-term memory networks for time-series and clinical note analysis, and generative adversarial networks for medical image augmentation. Vision transformers have shown promise in image segmentation tasks, while graph neural networks (GNNs) are discussed separately due to their unique capacity to model relational

data. Emerging self-supervised models have also improved representation learning in limited-labeled datasets. Hybrid DL models, such as CNN-RNN combinations, are increasingly used to enhance diagnostic accuracy[28,104,105].

Natural language processing

Natural language processing (NLP) focuses on the interaction between computers and human language, enabling AI systems to extract structured insights from unstructured clinical notes, discharge summaries, and radiology reports. It plays a crucial role in clinical information retrieval, temporal event extraction, and predictive modeling based on narrative patient data[106]. Abu-Ashour *et al*[41] integrated NLP into decision-support tools, significantly improving the triage efficiency for patients with suspected AAp.

Computer vision

This AI subfield enables machines to interpret and analyze visual data, making it particularly valuable for diagnostic imaging. In clinical medicine, computer vision is widely used for the classification, segmentation, and detection of anomalies in radiologic images such as US, CT, and MRI. Rajpurkar *et al*[60] developed AppendiXNet, a DL model that achieved an AUROC of 0.81 in identifying AAp from CT scans, demonstrating AI's potential in radiology-based triage. In addition, vision transformers have recently been explored for improving segmentation accuracy and enhancing the classification of complex medical images.

RL and XAI

RL is an AI paradigm where agents learn to make optimal sequential decisions by interacting with their environment and receiving feedback in the form of rewards. Although rarely applied to AAp, RL has been successfully used in other medical domains, such as insulin dosing in diabetes, sepsis management, and ventilator control, and holds promise for dynamic treatment optimization in surgical care pathways. XAI, on the other hand, addresses the interpretability of complex ML models. Techniques like SHapley Additive exPlanations (SHAP) and local interpretable model-agnostic explanations (LIME) are increasingly adopted to explain predictions made by ensemble and DL models, such as RF, XGBoost, and ANN. By improving transparency, XAI methods enhance clinician trust and facilitate regulatory approval in clinical AI deployment[43,107,108].

Federated learning

Federated learning enables AI models to be collaboratively trained across multiple healthcare institutions without centralizing sensitive patient data, thus enhancing privacy and data security. This approach could be particularly beneficial for future multi-center AAp research, allowing models to generalize across diverse populations while preserving institutional autonomy. However, challenges such as data heterogeneity, communication latency, and synchronization of model updates remain active areas of investigation[109].

Bayesian networks

Bayesian inference methods provide a probabilistic framework for modeling uncertainty in clinical data, particularly when information is incomplete or ambiguous. These networks use conditional dependencies among variables to support diagnostic reasoning and treatment decision-making. In AAp research, Bayesian networks could be applied to risk stratification by integrating prior knowledge from patient demographics, laboratory markers, and imaging findings[110]. However, several barriers hinder their widespread adoption in clinical practice. In particular, the need for expert knowledge to establish accurate prior probabilities, the challenges associated with constructing model structures, and the difficulties of real-time application in emergency settings stand out.

Transformer-based AI models (bidirectional encoder representations from transformers, generative pre-trained transformer)

These DL models are highly effective in processing and analyzing medical text, including clinical notes, discharge summaries, and radiology reports. They enable robust extraction of clinical features from unstructured data and have been used for tasks such as automated chart review, symptom classification, and clinical risk prediction. In the context of acute care, transformer-based models may support triage systems by identifying high-risk cases based on electronic health record narratives[111].

GNN

GNNs are DL architectures designed to process data that are structured as graphs, where entities (nodes) and their interactions (edges) are central to modeling. Unlike traditional neural networks, GNNs can capture complex relationships between clinical variables, making them suitable for representing patient comorbidities, disease trajectories, and treatment outcomes. Although not yet widely applied in AAp, GNNs could support risk stratification by integrating patient history and clinical interactions in a graph-based format[112].

Automated ML

Automated ML (AutoML) systems provide an automated framework for model selection, hyperparameter tuning, and feature engineering, thereby reducing the need for intensive manual intervention and allowing users with limited data science expertise to develop robust ML models[113]. In the context of AAp, AutoML could streamline the development of diagnostic and prognostic AI tools, particularly in clinical settings where dedicated data science resources are scarce, thus

facilitating broader clinical adoption.

Edge AI

Edge AI refers to executing AI algorithms directly on local devices, such as hospital bedside monitors, portable US scanners, or smartphones without relying on centralized cloud infrastructure. This enables real-time inference, reduces latency, enhances data privacy, and allows decision support in settings with limited or unstable internet access. Edge AI holds particular promise for emergency rooms, rural clinics, and prehospital environments where rapid, autonomous decision-making is essential[103].

AI's growing role in AAp diagnosis is evident across various ML and DL applications. By integrating multimodal data sources including patient history, physical examination findings, laboratory markers, and imaging studies, AI models enhance diagnostic precision. Studies have demonstrated that AI-powered decision support systems can aid clinicians in distinguishing complicated from uncomplicated AAp, optimizing treatment strategies, and reducing unnecessary surgeries. As AI continues to evolve, its clinical applications in AAp detection and management will likely expand, further enhancing precision medicine and individualized treatment approaches[28]. The summary of the terminology was provided in Table 2.

AI APPLICATIONS IN AAP DIAGNOSIS

Clinical data-based models

Traditional clinical scores assist in suspected AAp cases; however, they do not always provide sufficient sensitivity and specificity. AI methods leverage clinical data such as patient history, physical examination findings, and laboratory values to develop more accurate diagnostic models. As of 2025, numerous studies have demonstrated the success of AI-based clinical models in AAp diagnosis[3,8,16,27,28,33,35-84].

For example, in the study by Chadaga *et al*[42], ensemble learning methods outperformed traditional clinical scores in diagnosing AAp. The study reported an AUROC of 0.82 when using a combination of XGBoost and LightGBM models, highlighting a significant improvement over standard clinical evaluations. The systematic review by Rey *et al*[82] showed that AI models trained with multimodal data (clinical + laboratory + imaging) outperformed conventional diagnostic approaches for AAp. This review particularly emphasized the high sensitivity and specificity of models such as CatBoost, LightGBM, and RF. In another study, Phan-Mai *et al*[46] reported an AUROC of 89.4% in the prediction of perforation using DL models. Similarly, Akbulut *et al*[43] found that the CatBoost algorithm, when based solely on clinical data, achieved 92% specificity and 88% sensitivity in diagnosing AAp.

Clinical performance of ML models

Systematic reviews highlight the clinical performance of various ML models by comparing their sensitivity, specificity, and AUROC values. For example, the systematic review by Issaiy *et al*[28] reported that the RF model achieved 94% sensitivity and 96% specificity in diagnosing AAp. Similarly, the study by Rey *et al*[82] conducted a systematic review of pediatric AAp studies and found that all AI models many integrating clinical, laboratory, and imaging data achieved AUC/AUROC values above 0.9, demonstrating a superior diagnostic performance compared to conventional methods.

In another study, Yoldaş *et al*[71] evaluated a neural network model on 156 patients, reporting near-perfect results, *i.e.* 100% sensitivity and 97.2% specificity (PPV = 96%, NPV approaching 100%). However, such small-scale studies raise concerns about whether AI models will maintain the same accuracy when applied to larger datasets. However, such small-sample studies (< 200 cases in total) risk overfitting, particularly in DL, our analysis suggests a minimum target of 150 cases per severity class for ML and > 500 cases for DL architectures in AAp. For rare complications, prospective registries may be necessary to achieve an adequate statistical power.

ML has also demonstrated success in larger datasets. In a study conducted in Taiwan, a RF model trained on demographic and clinical data from 180 patients achieved 94% sensitivity, 100% specificity, 96% accuracy, and an AUC of 0.98 for AAp diagnosis, outperforming other algorithms such as SVM and ANN[72]. In a larger dataset from Turkey, which included 7244 patients, models trained on demographic and laboratory parameters showed that RF achieved the highest performance (AUC = 0.99), particularly in identifying complicated cases[59]. In this study, the decision tree model provided slightly lower accuracy (AUC = 0.94) but offered a more interpretable approach. XAI techniques facilitate the integration of such models into clinical practice.

Integration of ensemble learning into clinical decision support systems

Ensemble learning methods have been increasingly integrated into clinical decision support systems in recent years. According to the systematic review by Rey *et al*[82], AI models demonstrated consistently high diagnostic performance in pediatric AAp. In particular, models that combined clinical, laboratory, and imaging features achieved accuracies above 90% and AUROC values exceeding 0.9, substantially outperforming conventional diagnostic approaches. Similarly, Males *et al*[39] demonstrated that an XGBoost model incorporating SHAP and LIME techniques improved diagnostic performance, particularly in atypical cases.

Models trained solely on laboratory data have shown limited success. In a study by Mijwil and Aggarwal[47], the RF model, when trained exclusively on laboratory data, achieved 81% sensitivity, 81% specificity, and 84% accuracy. This finding suggests that ML models trained only on laboratory values may underperform in the absence of clinical and imaging data. However, Ghareeb *et al*[54] developed a hybrid model using ensemble learning techniques that achieved

Table 2 Definitions of artificial intelligence techniques employed in acute appendicitis research

Method	Definition	Relation to deep learning	Advantages
Deep learning	A subset of ML that uses multi-layered neural networks to automatically extract features from large datasets	DL is commonly used in image analysis text processing and predictive modeling. FL and edge AI can enhance the efficiency and privacy of DL models	High ACC strong capability in handling image and language data
Federated learning	A decentralized ML approach where models are trained across multiple institutions without sharing patient data	FL allows DL models to be trained across different centers while preserving patient privacy. It is useful for multi-center AI studies in appendicitis diagnosis	Enhances data privacy allows for cross-institutional AI model development
Edge AI	AI models that run directly on local hospital devices portable ultrasound scanners or mobile systems instead of relying on cloud computing	Edge AI enables DL models to operate in real-time on local devices reducing dependence on internet connectivity	Real-time processing improved data security reduced latency in decision-making
Bayesian networks	Probabilistic models that establish relationships between variables and handle uncertainty in data	Can be integrated with DL models to improve decision-making under incomplete information	Useful for risk prediction particularly in cases with missing clinical data
Transformer-based AI models (BERT, GPT)	Large language models capable of understanding and processing medical text	Can be used in combination with DL for automated triage systems and clinical note analysis	Efficient text processing potential for real-time clinical decision support
Graph neural networks	AI models that analyze relationships between data points in a structured graph format	GNNs can enhance DL models by incorporating complex patient relationships and comorbidities	Improves risk prediction models enhances interpretability of patient data interactions
Automated machine learning	AI systems that automatically optimize model selection hyperparameters and feature engineering	AutoML can generate optimized DL models without requiring manual tuning	Reduces the need for expert AI developers accelerates model deployment
Natural language processing	AI systems designed to interpret and extract information from human language including clinical notes and radiology reports	NLP models can be integrated with DL to analyze unstructured medical data	Enhances electronic health record analysis supports AI-assisted triage systems
Computer vision	AI field enabling machines to interpret visual data particularly useful in medical imaging	Computer vision models. including DL-based CNNs improve diagnostic ACC in radiology	Reduces diagnostic variability. increases ACC in CT and MRI interpretation
Reinforcement learning and explainable AI	AI models that learn optimal decision pathways based on cumulative rewards XAI ensures transparency in model predictions	Can optimize treatment strategies while SHAP and LIME techniques make AI models interpretable for clinicians	Improves AI adoption in healthcare enables better treatment planning
Machine learning	A broad AI field encompassing various algorithms including supervised and unsupervised learning	ML models, such as SVM, random forest and XGBoost form the foundation for AI in clinical decision-making	Provides adaptable and scalable models for medical data analysis
Vision transformers	A deep learning model specifically designed for image segmentation and classification	Enhances medical image analysis by capturing spatial relationships within radiology images	Improves segmentation ACC particularly in CT and MRI-based diagnosis
Lazy learning algorithms (KNN)	Classification method that identifies the closest data points in a dataset	Used in ML for patient clustering and classification	Simple yet effective but computationally expensive in large datasets
Extra trees classifier	A variant of random forest that introduces additional randomness to improve ACC	Works alongside ensemble learning to enhance classification performance	High ACC robustness in medical data analysis
Hybrid AI models	AI models combining ML and DL techniques to improve diagnostic performance	Used in multimodal AI-based appendicitis detection	Enhances ACC by integrating structured and unstructured data sources

AI: Artificial intelligence; BERT: Bidirectional encoder representations from transformers; GPT: Generative pre-trained transformer; KNN: K-nearest neighbor; ML: Machine learning; XAI: Explainable artificial intelligence; ACC: Accuracy; DL: Deep learning; FL: Federated learning; GNN: Graph neural network; AutoML: Automated machine learning; NLP: Natural language processing; CNN: Convolutional neural network; SHAP: SHapley Additive exPlanations; LIME: Local interpretable model-agnostic explanations; SVM: Support vector machine; XGBoost: Extreme gradient boosting; CT: Computed tomography; MRI: Magnetic resonance imaging.

93% accuracy even when trained only on laboratory data.

Differentiating complicated vs uncomplicated AAp

AI algorithms are not only used to detect AAp but also to differentiate between complicated and uncomplicated cases. In a study involving 1797 patients, Akbulut *et al*[43] developed a CatBoost algorithm based on demographic and biochemical data. This model achieved 84% sensitivity and 93% specificity (AUC = 0.94) for identifying AAp, while its

predictive performance for perforated AAp was even higher, with 94% sensitivity and 90% specificity (AUC = 0.97). Moreover, XAI techniques helped identify the most influential factors in predicting perforation, thereby guiding clinical decision-making[43].

Recent studies have demonstrated that integrating multimodal data such as clinical, laboratory, and imaging findings enhances the accuracy of AI models in differentiating complicated from uncomplicated AAp. According to the systematic review by Rey *et al*[82], AI models that incorporate multimodal inputs including CT-based features as well as inflammatory markers such as C-reactive protein and leukocytosis consistently achieved very high diagnostic accuracy (generally > 90%) or AUC values (> 0.9) for distinguishing AAp, indicating a strong predictive performance across diverse modeling approaches. Similarly, in a systematic review, Issaiy *et al*[28] reported that advanced ML models, such as RF, SVMs, ANNs, and XGBoost, achieved AUROC values ranging from 0.84 to 0.94, consistently outperforming traditional clinical assessment.

In addition to ML techniques, DL models have shown promising results in distinguishing complicated AAp. Phan-Mai *et al*[46] utilized a CNN trained on US and CT images, achieving an AUROC of 0.894, with 91% sensitivity and 88% specificity in detecting complicated AAp cases. Another study by Liang *et al*[91] reported that radiomics-based DL approaches improved prediction of perforation compared to standard radiological evaluations.

Furthermore, XAI techniques, such as SHAP and LIME, have been integrated into predictive models to enhance interpretability. According to Chadaga *et al*[42], SHAP-based feature analysis indicated that the most influential predictors for AAp detection in pediatric patients were length of hospital stay, visibility of the vermiform appendix on US, white blood cell count, and appendix diameter, thereby enhancing the interpretability and reliability of AI-driven diagnostic decision-making. These insights contribute to more reliable and interpretable AI-driven decision-making in clinical practice.

In conclusion, systematic review studies have shown that ML-based approaches provide higher accuracy than traditional clinical methods in diagnosing AAp. Particularly, ensemble learning models such as XGBoost, LightGBM, and CatBoost play a crucial role in diagnosing atypical cases. DL methods and radiomics-based AI models further enhance the differentiation between complicated and uncomplicated AAp. These technologies can be integrated into clinical decision support systems to improve diagnostic accuracy, reduce unnecessary surgeries, and optimize patient management strategies.

AI-ENHANCED MEDICAL IMAGING TECHNIQUES

AI applications in US imaging

US is typically the first-line imaging modality for diagnosing AAp. However, its diagnostic accuracy is operator-dependent and may vary significantly among less experienced users. AI assists in this area by enabling automatic appendix detection in US images, particularly benefiting less experienced practitioners. In the study by Abu-Ashour *et al*[41], ChatGPT-4 was applied to label free-text operative and US reports for grading pediatric AAp. Compared with human data abstractors, ChatGPT-4 substantially reduced misclassification rates (2.9% *vs* 28.2%) and prevented 59.2% of errors, while being nearly 40 times faster. Several studies have demonstrated that the addition of US findings to clinical and laboratory parameters improves diagnostic performance. For instance, Anandalwar *et al*[114] reported that incorporating white blood cell and polymorphonuclear leukocyte % into equivocal US results increased the NPV from 41.9% to 95.8%, while the PPV rose from 79.1% to 91.3% and from 89.1% to 96.8%[114]. Similarly, combining US with the PAS score provided higher specificity in distinguishing complicated from uncomplicated cases compared to either method alone[115]. Combining US with a PAS score may help distinguish complicated from uncomplicated AAp in a pediatric population.

According to a systematic review by Rey *et al*[82], AI integration into US-based assessment was reported to facilitate appendix visualization and improve diagnostic performance, particularly in case of low-resolution or challenging cases. In recent years, ensemble learning and transfer learning techniques have been increasingly applied in US analysis. Marcinkevičs *et al*[52] developed an interpretable ML framework for pediatric AAp detection using US images. Their best-performing model, a semi-supervised Multiview concept bottleneck model, achieved an AUROC of 0.80 and an area under precision-recall of 0.92, demonstrating competitive accuracy while maintaining model explainability. According to a recent systematic review by Rey *et al*[82], ensemble learning-based approaches such as XGBoost and LightGBM outperform traditional diagnostic methods in pediatric AAp, particularly when applied to US data. Although exact sensitivity and specificity values varied across studies, the review emphasizes the potential of these models to enhance diagnostic accuracy and supports clinical decision-making.

A recent study by Hayashi *et al*[55] further explored AI-assisted US in pediatric AAp diagnosis. Their research involved training a DL model with 70 US videos, evaluating its effectiveness in two phases. The first phase assessed AI performance in detecting the appendix, with successful identification in shallow scans but decreased accuracy in deeper scans (> 8 cm). Potential technical solutions to improve performance in deep tissue imaging include the integration of higher-frequency transducers with enhanced penetration capabilities, optimization of DL architectures through multi-scale feature extraction, implementation of contrast-enhanced US techniques, and development of hybrid models that combine AI with real-time operator feedback systems to guide probe positioning and optimize image acquisition parameters. The second phase analyzed AIs affect pediatricians' diagnostic confidence. Results indicated that AI assistance was beneficial when the appendix was at least partially detected but could negatively influence decision-making when the appendix was not identified. This highlights the need for further refinement in AI models, particularly in handling deep tissue scans and minimizing false negatives to avoid misleading clinicians.

AI applications in CT imaging

CT is considered the gold standard for diagnosing AAp due to its high sensitivity, especially in ambiguous cases, where low-dose abdominal CT is recommended. AI enhances CT image analysis by automatically detecting AAp features and complications, identifying details that may be missed by human observation. DL has played a pivotal role in this field. The AppendiXNet model developed by Rajpurkar *et al*[60] was trained on 438 patient CT images, achieving an accuracy of 72% (AUC = 0.81). The study by Gollapalli *et al*[38] found that AI-assisted CT analysis achieved up to 90% sensitivity in early AAp cases. Zhao *et al*[90] reported that radiomics models integrated with clinical information on CT images achieved higher diagnostic accuracy in differentiating simple from non-simple AAp compared with CT-based assessments alone.

Differentiating between complicated and uncomplicated AAp using CT images is crucial for determining surgical necessity. Traditional radiological criteria (*e.g.*, abscess, free air) do not always reliably indicate perforation. AI-enhanced imaging analysis can more accurately assess perforation risk. In the study by Liang *et al*[91], CT data from 1165 patients were analyzed using DL and radiomics techniques. A CatBoost model, when combined with radiologist evaluation, achieved an AUC of 0.79 for identifying complicated AAp. The model's sensitivity reached 70%, whereas traditional radiologist assessments achieved only 45%. Its NPV was 80%, which was 7% higher than that of radiologists. However, its specificity was 74%, lower than the 90% specificity of radiologists, indicating a potential tendency for overdiagnosis in non-complicated cases.

ML techniques are also effectively applied in CT analysis. Issaiy *et al*[28] conducted a systematic review of AI and ML models in AAp diagnosis and prognosis. They highlighted that ensemble learning models including XGBoost, CatBoost, and LightGBM consistently outperformed traditional diagnostic methods (*e.g.*, clinical scoring systems or standard imaging interpretation) across multiple studies, demonstrating higher accuracy, sensitivity, and specificity. According to Rey *et al*[82], ensemble learning and XAI techniques applied to CT analysis have demonstrated superior diagnostic performance compared with conventional interpretation methods, with several studies reporting high AUROC and sensitivity values across different patient cohorts. Dogan and Selcuk[31] developed a novel DL approach for AAp diagnosis, utilizing a hybrid CNN model integrated with ensemble learning techniques, including SVM, KNN, and RF. Their method demonstrated an independent diagnostic accuracy of 96% in cases with definitive CT-based radiological findings and 83.3% in cases with radiologically ambiguous CT findings, surpassing traditional radiologist-based evaluations. The hybrid model achieved a sensitivity of 95.7%, specificity of 69.7%, overall accuracy of 92.8%, and an F1 score of 94.2%, highlighting its robustness in diagnosing AAp, particularly in challenging cases where conventional imaging interpretation is difficult.

AI applications in MRI

MRI is primarily used for AAp diagnosis in pregnant women and pediatric patients, where radiation exposure must be minimized. However, its routine use in emergency settings remains limited. Despite its proven diagnostic accuracy, MRI's routine use in emergency settings remains limited due to longer acquisition times and limited availability compared to CT and US. Current literature on AI-assisted MRI analysis for AAp diagnosis is notably scarce, with existing AI research predominantly focusing on CT imaging, US, and clinical-laboratory parameter combinations. The vast majority of published AI studies in AAp diagnosis have utilized CT scans and US imaging as primary data sources, reflecting the more widespread availability and faster acquisition times of these modalities in emergency departments. However, emerging developments in pediatric MRI and AI demonstrate significant potential for future applications, particularly in image optimization, organ segmentation, and automated diagnosis. The integration of AI with MRI technology represents an underexplored frontier that could potentially enhance diagnostic accuracy and efficiency in AAp evaluation, especially in radiation-sensitive populations.

ML AND DL MODELS IN COMPARISON WITH TRADITIONAL METHODS

ML and DL approaches offer different advantages in the diagnosis of AAp. ML methods (*e.g.*, logistic regression, decision trees, RF, SVM, XGBoost, LightGBM, *etc.*) are models that can work with relatively smaller datasets and provide results that are somewhat interpretable. Indeed, many studies have demonstrated that ML models offer significantly higher accuracy than traditional clinical scores. For example, in a study by Gollapalli *et al*[38], the RF model outperformed other ML algorithms in both diagnosing AAp and predicting its complicated form, achieving an AUC of up to 99%. Similarly, Chadaga *et al*[42] reported that XGBoost and LightGBM algorithms, when integrated with clinical data, achieved an accuracy of 91% compared to traditional methods. ML-based systems also use XAI techniques such as SHAP to determine which clinical parameters are most influential in diagnosis, providing clinicians with more transparent models[116].

On the DL side, ANNs and, more specifically, CNNs for image analysis have gained significant attention in recent years. DL has the ability to autonomously learn features from raw data through multi-layered neural networks[117]. This allows it to detect complex data patterns that human experts may overlook. ANNs were first applied to AAp diagnosis in the late 2000s and have been shown to outperform clinical scores even in small-scale studies[73]. For example, in a study by Prabhudesai *et al*[73], an ANN model significantly outperformed clinicians, completely eliminating false-negative cases and diagnosing with 100% sensitivity (97% specificity). Similarly, in a study conducted in Turkey (Yoldaş *et al*[71]), an ANNs model achieved 100% sensitivity and 97% specificity, demonstrating excellent performance in preventing false negatives. These models statistically outperform scoring systems such as the Alvarado score.

The superiority of DL in image analysis is also becoming more evident. The aforementioned AppendiXNet study achieved a reasonable AUC (0.81) in AAp diagnosis using a limited number of CT samples[60]. However, more recent systematic reviews have shown that CNNs models become more robust in clinical diagnosis when trained on large datasets. In the study by Schipper *et al*[33] two ML models (history intake vitals examination and history intake vitals examination-laboratory tests) based on the XGBoost algorithm were developed to predict AAp in patients presenting with acute abdominal pain in the emergency department. The models demonstrated high discriminative performance (AUROC = 0.919 and 0.923, respectively), outperforming the Alvarado score (AUROC = 0.824) and showing comparable or superior accuracy to emergency physicians, particularly when laboratory results were incorporated. In a systematic review, Rey *et al*[82] noted that all included studies developed their own ML or CNN-based models and consistently reported diagnostic performances exceeding 90% accuracy or an AUC greater than 0.9.

Multimodal AI models, which integrate different types of data, are particularly effective in improving diagnostic success. Zhao *et al*[90] developed a radiomics model that integrated clinical data, laboratory parameters, and CT imaging and demonstrated that a combined model using both radiomics features and clinical information achieved a significantly higher AUC than a CT-only model (P = 0.041) in differentiating simple from complicated AAp. In a retrospective cohort study by Phan-Mai *et al*[46] various ML models including SVM, decision trees, logistic regression, KNN, ANNs, and gradient boosting were used to classify complicated *vs* uncomplicated AAp. AUC and accuracy values ranged from approximately 0.69 to 0.82 in raw data, with gradient boosting models achieving higher accuracy and an AUC ≥ 0.8 after balancing *via* synthetic minority oversampling technique.

However, it is important to note that some of the striking results reported in the literature were obtained in small and selectively chosen patient groups. Early ANN studies that reported 100% sensitivity and 97% specificity have been subject to methodological criticisms, particularly regarding small sample sizes and the risk of overfitting. While small-sample studies (< 200 total cases) risk overfitting, particularly in DL, our analysis suggests minimum a target of 150 cases per severity class for ML and > 500 cases for DL architectures in AAp. For rare complications, prospective registries should target ≥ 50 confirmed events per model class. A systematic review highlighted that many AI-based AAp diagnosis studies suffer from selection bias and inadequate model validation. Therefore, before DL models can be implemented in real-world practice, they must be tested on larger, multicenter, and heterogeneous datasets.

In the future, it will be crucial for these models to gain clinicians' trust by combining the interpretability of decision tree-based approaches with the accuracy of DL. Additionally, more prospective studies and randomized controlled trials are needed to integrate AI models into clinical practice. Particularly, the combination of ensemble learning techniques and DL models holds promise for developing the most accurate and reliable decision support systems in AAp diagnosis.

COMPARATIVE ANALYSIS OF DIAGNOSTIC ACCURACY AND PERFORMANCE OF AI MODELS

The performance of AI models in diagnosing AAp varies between studies but is generally high. Table 1 presents the performance metrics for distinguishing between normal appendix and AAp across various studies, while Table 2 summarizes the accuracy of AI models in detecting complicated (perforated) cases among AAp patients. These tables provide key performance indicators such as sensitivity, specificity, PPV, NPV, and AUROC.

Comprehensive systematic reviews indicate that DL and ensemble learning methods (*e.g.*, XGBoost, LightGBM, CatBoost) offer significant advantages over traditional clinical scoring methods in AAp diagnosis. While these reviews highlight promising performance, they also underscore a critical limitation: A great proportion of the underlying evidence derives from retrospective, single-center studies with heterogeneous definitions of 'complicated' AAp, which can introduce selection bias and limit generalizability. In the systematic review by Rey *et al*[82], ANN models were reported to reach an AUROC of up to 0.985. Models trained on large datasets, such as RF and XGBoost, achieved sensitivity rates of 90%-95%, while specificity values ranged between 85%-93%, minimizing false positives. For instance, in the study by Chadaga *et al*[42], a model combining XGBoost and LightGBM achieved 91% accuracy, 92% sensitivity, and 90% specificity in diagnosing AAp.

PPV and NPV are as crucial as sensitivity and specificity in clinical practice. For example, in the ANN model by Yoldaş *et al*[71], NPV was reported as 100%, meaning the model was highly reliable in ruling out AAp cases and preventing unnecessary surgeries. Similarly, Liang *et al*[91] reported that a DL + radiomics model for complicated AAp achieved an NPV of 80%, which was 7 percentage points higher than that of traditional radiology-based assessments (73%). The model also demonstrated a substantially higher sensitivity (70%) compared to radiologists (45%). Shahmoradi *et al*[67] developed a MLP network-based DL model with 80% sensitivity, 97.5% specificity, 92.3% PPV, and 93% NPV for AAp diagnosis. Hsieh *et al*[72] used a RF model that achieved 94% sensitivity, 100% specificity, 100% PPV, and 87% NPV. Hsieh *et al*[72] reported that a SVM-based model reached 91% sensitivity, 100% specificity, 85% PPV, and 73% NPV. Hsieh *et al*[72] applied an ANN model that achieved 94% sensitivity, 85% specificity, 94% PPV, and 85% NPV. Yazici *et al*[37] demonstrated that using a logistic regression model with only three readily available clinical features age, C-reactive protein, and peri-appendicular fluid collection achieved a diagnostic accuracy of approximately 96% in differentiating uncomplicated and complicated AAp. These studies highlight the potential of various AI approaches in ensuring reliable clinical diagnosis.

THE ROLE OF AI IN CLINICAL DECISION-MAKING

AI plays a crucial role in supporting decision-making processes in managing patients with suspected AAp. AI-based models enhance diagnostic accuracy, reduce unnecessary appendectomies, and improve the prediction of complications such as perforation, contributing directly to patient care. The advantages of AI-based clinical decision support systems include: (1) More accurate classification of uncertain cases in AAp diagnosis; (2) Better determination of surgical necessity for complicated AAp cases; (3) Improved selection of patients suitable for non-operative treatment; and (4) Enhanced postoperative complication prediction and patient management.

Cappuccio *et al*[118] conducted a comprehensive literature review on AI applications in AAp diagnosis and management and their findings indicate that AI significantly enhances diagnostic accuracy, speed, and consistency, leading to improved patient outcomes and potentially reducing healthcare costs. For example, Males *et al*[39] concluded that the developed ML model could potentially reduce unnecessary surgeries in AAp cases by up to 17%, while maintaining a high safety margin, with only 0.3% of cases missing the required surgical intervention. Additionally, Hsieh *et al*[72] reported that a RF-based model achieved 94% sensitivity, 100% specificity, 100% PPV, and 87% NPV in AAp diagnosis. Akbulut *et al*[43] demonstrated that CatBoost-based XAI model improved clinical accuracy in the diagnosis of AAp (accuracy 88%, sensitivity 84%, specificity 93%, AUC = 0.94) and in detecting perforated AAp (accuracy 92%, sensitivity 94%, specificity 90%, AUC = 0.97).

The integration of AI into clinical decision-making also requires XAI techniques for transparency. SHAP and LIME interpretation algorithms help increase clinical trust by clarifying why a model predicts high or low risk for a particular patient. Chadaga *et al*[42] demonstrated that SHAP-based analyses identified key factors affecting the likelihood of complicated AAp, aiding physicians in clinical decision-making. In their study, Schipper *et al*[33] introduced ML models utilizing clinical and laboratory information that achieved AUROCs of 0.919 without and 0.923 with laboratory data. Compared with the Alvarado score (AUROC = 0.824), these models showed markedly improved accuracy and performed on par with or better than emergency physicians, whose AUROCs ranged between 0.791 and 0.923.

From a clinical perspective, AI models should be used as decision-support tools rather than a substitute for physicians. AI should be viewed as a second opinion or triage tool rather than a primary decision-maker. The final clinical decision should always be made by an experienced physician.

In conclusion, AI-based approaches have demonstrated superior diagnostic accuracy over traditional methods in many studies. However, the success of these models depends on the type of data used (clinical data alone or with imaging), the structure of the algorithm, and the population on which the model was trained. Therefore, reported figures in the literature should be interpreted cautiously, and each healthcare institution should select the AI model most appropriate for its patient profile.

FUTURE RESEARCH AND RECOMMENDATIONS

The integration of AI into the diagnosis and management of AAp is rapidly advancing; however, there are still areas that require further development. Several key directions and recommendations for future research are as follows.

Larger and more diverse datasets

Many existing studies have been conducted in single centers with limited patient populations. There is a need for large-scale datasets that include patients from different geographic regions, age groups, and risk categories. Special attention should be given to subgroups such as pediatric patients, the elderly, and pregnant women to evaluate model performance separately. Expanding datasets will enhance the generalizability of DL models while reducing the risk of overfitting. Indeed, evidence from broader AI research suggests that training on heterogeneous datasets can minimize diagnostic errors by increasing sensitivity and specificity, thereby improving reliability in real-world clinical settings [119]. This need arises because the current body of evidence often lacks rigorous external validation, a crucial step for assessing real-world performance. Models validated only on internal test sets may show inflated accuracy, and multi-institutional validation remains uncommon in the AAp AI literature.

Multicenter and prospective studies

To assess the real-world clinical impact of AI models, prospective studies conducted across multiple centers are essential. This will enable independent validation of the models on diverse patient populations and provide concrete evidence of their contribution to routine clinical practice, such as reducing negative appendectomy rates and minimizing perforation complications. Kelly *et al*[120] emphasized that for AI to achieve a measurable clinical impact, models must be validated on large, diverse, and multicenter datasets, as reliance on small single-center cohorts risks overestimating performance and limits real-world applicability.

Multimodal data models

Future AAp diagnostic algorithms should not rely solely on a single data type but should incorporate clinical, laboratory, and imaging information in a multimodal framework. However, technical challenges such as data heterogeneity, varying acquisition protocols, missing data across modalities, and computational complexity must be addressed through standardized data preprocessing pipelines, advanced feature alignment techniques, imputation strategies for missing values, and efficient DL architectures designed for multimodal fusion. For example, integrating symptom duration, physical examination findings, blood test results, and imaging data into AI models can lead to more accurate predictions.

A notable example is the study by Liang *et al*[91], where a combined clinical + CT + radiomics model demonstrated promising results. Additional studies have shown that combining clinical and imaging data can increase AUROC values by 15% compared to single-source models. Future advancements may include real-time AI-assisted decision-making systems in emergency departments, integrating clinical data with portable US imaging.

Explainability and model transparency

For clinicians to trust AI, it is crucial to ensure model interpretability. Therefore, future research should not only focus on accuracy but also on elucidating the reasoning behind model predictions. The use of techniques such as SHAP and LIME should be expanded to visualize how AI models reach decisions. In practice, SHAP values can be integrated into electronic health records dashboards to highlight patient-specific feature contributions (*e.g.*, elevated C-reactive protein or leukocytosis) directly within the clinical interface, enabling physicians to rapidly assess the rationale behind an AI-generated risk score; early implementations in sepsis prediction systems have demonstrated that such real-time, interpretable alerts improve diagnostic confidence and reduce decision time by up to 30%[121]. Future AI systems should aim for seamless electronic health records integration, where XAI outputs (*e.g.*, SHAP-based feature importance) are embedded directly into clinician workflows, similar to existing decision support alerts for sepsis or acute kidney injury. A comprehensive review by Han *et al*[122] details the fundamental principles, innovative designs (including passive and active mechanisms), and multidisciplinary applications of micromixers, offering valuable insights into microfluidic system optimization. Beyond diagnostic algorithms, emerging bioinspired microfluidic platforms such as biomimetic leaf-venation groove micromixers for controlled liposome synthesis highlight the potential for scalable, energy-efficient nano-pharmaceutical manufacturing that could complement AI-guided precision medicine approaches in inflammatory and infectious diseases. Emerging bionic micromixer designs, such as fractal baffle systems based on Murray's Law, demonstrate high-efficiency fluid mixing at microscale levels, offering potential for integration into lab-on-a-chip platforms that could supply real-time biochemical data for AI-driven diagnostic models[123,124].

Integration with educational modules

AI systems can also contribute to clinician education. In the future, interactive training platforms for surgical residents and emergency physicians could utilize AI to simulate diagnostic scenarios. This would allow practitioners to learn the key clinical features that AI identifies, improving their diagnostic accuracy in real-world settings. Educational AI modules have already been shown to improve junior physicians' diagnostic accuracy by 18% in pilot studies, underscoring the potential for AI in medical training[125].

Differential diagnosis beyond appendicitis

In patients presenting with acute abdominal pain, accurate diagnosis extends beyond confirming or excluding AAp. A wide range of alternative conditions including mesenteric lymphadenitis, ovarian cyst-related pathology, diverticulitis, and gallbladder disease may mimic AAp both clinically and radiologically, often leading to diagnostic uncertainty and unnecessary interventions. AI has the potential to bridge this gap by advancing from traditional binary classification to multi-diagnostic frameworks that can simultaneously evaluate several plausible conditions. DL applied to CT imaging, particularly when integrated with clinical and laboratory parameters, could enable a more holistic interpretation of abdominal pain presentations. Unlike binary AAp models, multi-class systems would be able to provide differential probabilities across multiple diagnoses, offering clinicians a ranked diagnostic spectrum rather than a single yes/no output. To achieve this, models must be trained on carefully curated, multi-institutional datasets with explicit labels for both AAp and its common mimickers, while addressing class imbalance and overlapping clinical features. Equally important is the issue of transparency: AI models must provide interpretable explanations at both the case and feature levels, highlighting which imaging findings or laboratory markers drive each diagnostic suggestion. Such transparency not only increases clinical trust, but also enhances educational value for trainees by reinforcing key differentiating features. Ultimately, by supporting comprehensive multi-diagnostic classification, AI could transform emergency radiology workflows accelerating triage, reducing negative appendectomies and missed diagnoses, and standardizing reporting across diverse clinical settings.

BEYOND ALGORITHMS: ETHICAL AND CLINICAL OVERSIGHT IN AI RESEARCH

In reviewing the broader literature on the applications of AI in healthcare, it becomes evident that many of the published studies are authored primarily by computer scientists, software developers, and engineers, often without sufficient involvement of clinical experts. As a result, a considerable proportion of these papers contain critical inaccuracies or oversimplifications in their medical content. In fact, when preparing this review, we observed that in more than half of the AAp and AI-related papers, the authors appeared to have limited understanding of what AAp truly entails. This underscores a crucial reality: Healthcare cannot be reduced to a purely mathematical exercise. Therefore, editors and reviewers should be particularly vigilant when evaluating such submissions. They should carefully examine author lists and ensure the inclusion of clinical expertise, and they should scrutinize more rigorously any manuscripts on AI in healthcare that lack clinician involvement. The responsible and ethical application of AI in medicine requires that data be carefully analyzed and interpreted by specialists in the relevant clinical domains before being processed through computational models. Moreover, the outcomes generated by such models must be rigorously evaluated for their clinical applicability and real-world translational value. Going forward, stronger interdisciplinary collaboration between clinicians and computer scientists will be essential to ensure that AI research in healthcare is accurate, clinically relevant,

and ethically responsible.

CONCLUSION

AI is an emerging tool for differentiating normal appendix, AAp, and perforated AAp. Current literature suggests that well-trained algorithms can surpass clinical scoring systems and, in some cases, even outperform experienced clinicians in diagnostic accuracy. ML and DL techniques have achieved sensitivity and specificity values exceeding 90% in AAp diagnosis. These models can expedite decision-making in emergency settings, helping to prevent life-threatening complications and reduce unnecessary surgeries. However, for AI to become part of routine clinical practice, extensive validation studies and increased model transparency are required to gain clinician trust. Further integration of AI into healthcare workflows, coupled with regulatory approvals and physician training programs, will be critical for large-scale adoption. In the future, AI systems integrated with multidisciplinary approaches and clinical workflows may not only revolutionize AAp diagnosis but also become a standard tool in the management of general surgical emergencies.

FOOTNOTES

Author contributions: Akbulut S, Kucukakcali Z, and Colak C contributed equally to the study design, data collection, analysis, manuscript writing, and revision; all authors have read and approved the final version of the manuscript.

Conflict-of-interest statement: The authors declare that they have no conflict of interest.

Open Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

Country of origin: Türkiye

ORCID number: Sami Akbulut 0000-0002-6864-7711; Zeynep Kucukakcali 0000-0001-7956-9272; Cemil Colak 0000-0002-7529-1100.

S-Editor: Fan M
L-Editor: A
P-Editor: Lei YY

REFERENCES

1 **Di Saverio S**, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, Boermeester M, Sartelli M, Coccolini F, Tarasconi A, De' Angelis N, Weber DG, Tolonen M, Birindelli A, Biffl W, Moore EE, Kelly M, Soreide K, Kashuk J, Ten Broek R, Gomes CA, Sugrue M, Davies RJ, Damaskos D, Leppäniemi A, Kirkpatrick A, Peitzman AB, Fraga GP, Maier RV, Coimbra R, Chiarugi M, Sganga G, Pisanu A, De' Angelis GL, Tan E, Van Goor H, Pata F, Di Carlo I, Chiara O, Litvin A, Campanile FC, Sakakushev B, Tomadze G, Demetrashvili Z, Latifi R, Abu-Zidan F, Romeo O, Segovia-Lohse H, Baiocchi G, Costa D, Rizoli S, Balogh ZJ, Bendinelli C, Scalea T, Ivatury R, Velmahos G, Andersson R, Kluger Y, Ansaloni L, Catena F. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. *World J Emerg Surg* 2020; **15**: 27 [*RCA*] [PMID: 32295644 DOI: 10.1186/s13017-020-00306-3] [FullText] [Full Text(PDF)]

2 **Kabir SA**, Kabir SI, Sun R, Jafferbhoy S, Karim A. How to diagnose an acutely inflamed appendix; a systematic review of the latest evidence. *Int J Surg* 2017; **40**: 155-162 [*RCA*] [PMID: 28279749 DOI: 10.1016/j.ijsu.2017.03.013] [FullText]

3 **Chen S**, Xia J, Xu B, Huang Y, Teng M, Pan J. Risk prediction and effect evaluation of complicated appendicitis based on XGBoost modeling. *BMC Gastroenterol* 2025; **25**: 295 [*RCA*] [PMID: 40275206 DOI: 10.1186/s12876-025-03847-6] [FullText]

4 **Singh JP**, Mariadason JG. Role of the faecolith in modern-day appendicitis. *Ann R Coll Surg Engl* 2013; **95**: 48-51 [*RCA*] [PMID: 23317728 DOI: 10.1308/003588413X13511609954851] [FullText] [Full Text(PDF)]

5 **Bhandarkar S**, Tsutsumi A, Schneider EB, Ong CS, Paredes L, Brackett A, Ahuja V. Emergent Applications of Machine Learning for Diagnosing and Managing Appendicitis: A State-of-the-Art Review. *Surg Infect (Larchmt)* 2024; **25**: 7-18 [*RCA*] [PMID: 38150507 DOI: 10.1089/sur.2023.201] [FullText]

6 **Petrauskas V**, Poskus E, Luksaite-Lukste R, Kryzauskas M, Petrulionis M, Strupas K, Poskus T. Suspected and Confirmed Acute Appendicitis During the COVID-19 Pandemic: First and Second Quarantines-a Prospective Study. *Front Surg* 2022; **9**: 896206 [*RCA*] [PMID: 35800113 DOI: 10.3389/fsurg.2022.896206] [FullText] [Full Text(PDF)]

7 **Akbulut S**, Koc C, Kocaaslan H, Gonultas F, Samdanci E, Yologlu S, Yilmaz S. Comparison of clinical and histopathological features of patients who underwent incidental or emergency appendectomy. *World J Gastrointest Surg* 2019; **11**: 19-26 [*RCA*] [PMID: 30705736 DOI: 10.4240/wjgs.v11.i1.19] [FullText] [Full Text(PDF)]

8 **Li L**, Sun Y, Sun Y, Gao Y, Zhang B, Qi R, Sheng F, Yang X, Liu X, Liu L, Lu C, Chen L, Zhang K. Clinical-radiomics models with machine-learning algorithms to distinguish uncomplicated from complicated acute appendicitis in adults: a multiphase multicenter cohort study. *Gastroenterol Rep (Oxf)* 2025; **13**: goaf039 [*RCA*] [PMID: 40462769 DOI: 10.1093/gastro/goaf039] [FullText] [Full Text(PDF)]

9 **Li J**, Ye J, Luo Y, Xu T, Jia Z. Progress in the application of machine learning in CT diagnosis of acute appendicitis. *Abdom Radiol (NY)* 2025; **50**: 4040-4049 [*RCA*] [PMID: 40095017 DOI: 10.1007/s00261-025-04864-5] [FullText]

10 **Dongarwar D**, Taylor J, Ajewole V, Anene N, Omoyele O, Ogba C, Oluwatoba A, Giger D, Thuy A, Argueta E, Naik E, Salemi JL, Spooner K, Olaleye O, Salihu HM. Trends in Appendicitis Among Pregnant Women, the Risk for Cardiac Arrest, and Maternal-Fetal Mortality. *World J Surg* 2020; **44**: 3999-4005 [*RCA*] [PMID: 32737556 DOI: 10.1007/s00268-020-05717-6] [FullText] [Full Text(PDF)]

11 **Jearwattanakanok K**, Yamada S, Suntornlimsiri W, Smuthtai W, Patumanond J. Validation of the diagnostic score for acute lower abdominal pain in women of reproductive age. *Emerg Med Int* 2014; **2014**: 320926 [*RCA*] [PMID: 24971177 DOI: 10.1155/2014/320926] [FullText] [Full Text(PDF)]

12 **Wahhab RASA**, Mohammed LK. Incidental Gynecological Conditions in Patients Presented with Acute Appendicitis. *Med J Babylon* 2024; **21**: 259-262 [DOI: 10.4103/mjbl.mjbl_153_23] [FullText]

13 **Raman SS**, Osuagwu FC, Kadell B, Cryer H, Sayre J, Lu DS. Effect of CT on false positive diagnosis of appendicitis and perforation. *N Engl J Med* 2008; **358**: 972-973 [*RCA*] [PMID: 18305278 DOI: 10.1056/NEJMc0707000] [FullText]

14 **Köse E**, Hasbahçeci M, Aydın MC, Toy C, Saydam T, Özsoy A, Karahan SR. Is it beneficial to use clinical scoring systems for acute appendicitis in adults? *Ulus Travma Acil Cerrahi Derg* 2019; **25**: 12-19 [*RCA*] [PMID: 30742281 DOI: 10.5505/tjtes.2018.22378] [FullText]

15 **Bom WJ**, Scheijmans JCG, Salminen P, Boermeester MA. Diagnosis of Uncomplicated and Complicated Appendicitis in Adults. *Scand J Surg* 2021; **110**: 170-179 [*RCA*] [PMID: 33851877 DOI: 10.1177/14574969211008330] [FullText] [Full Text(PDF)]

16 **Maleš I**, Kumrić M, Huić Maleš A, Cvitković I, Šantić R, Pogorelić Z, Božić J. A Systematic Integration of Artificial Intelligence Models in Appendicitis Management: A Comprehensive Review. *Diagnostics (Basel)* 2025; **15**: 866 [*RCA*] [PMID: 40218216 DOI: 10.3390/diagnostics15070866] [FullText]

17 **Navaei M**, Doogchi Z, Gholami F, Tavakoli MK. Leveraging Machine Learning for Pediatric Appendicitis Diagnosis: A Retrospective Study Integrating Clinical, Laboratory, and Imaging Data. *Health Sci Rep* 2025; **8**: e70756 [*RCA*] [PMID: 40260048 DOI: 10.1002/hsr2.70756] [Full Text]

18 **Shahul Hameed MR**, Shahul Hameed S, Rafi Ahamed R, Thomas FA, George B. WBC Count vs. CRP Level in Laboratory Markers and USG vs. CT Abdomen in Imaging Modalities: A Retrospective Study in the United Arab Emirates to Determine Which Are the Better Diagnostic Tools for Acute Appendicitis. *Cureus* 2023; **15**: e47454 [*RCA*] [PMID: 37873039 DOI: 10.7759/cureus.47454] [FullText]

19 **Benabbas R**, Hanna M, Shah J, Sinert R. Diagnostic Accuracy of History, Physical Examination, Laboratory Tests, and Point-of-care Ultrasound for Pediatric Acute Appendicitis in the Emergency Department: A Systematic Review and Meta-analysis. *Acad Emerg Med* 2017; **24**: 523-551 [*RCA*] [PMID: 28214369 DOI: 10.1111/acem.13181] [FullText]

20 **Balakrishnan P**, Munisamy P, Vijayakumar S, Sinha P. Clinical Scoring Systems to Diagnose Complicated Acute Appendicitis in a Rural Hospital: Are They Good Enough? *Cureus* 2024; **16**: e64927 [*RCA*] [PMID: 39156474 DOI: 10.7759/cureus.64927] [FullText]

21 **Andersson RE**, Stark J. Diagnostic value of the appendicitis inflammatory response (AIR) score. A systematic review and meta-analysis. *World J Emerg Surg* 2025; **20**: 12 [*RCA*] [PMID: 39923099 DOI: 10.1186/s13017-025-00582-x] [FullText]

22 **Mantoglu B**, Gonullu E, Akdeniz Y, Yigit M, Firat N, Akin E, Altintoprak F, Erkorkmaz U. Which appendicitis scoring system is most suitable for pregnant patients? A comparison of nine different systems. *World J Emerg Surg* 2020; **15**: 34 [*RCA*] [PMID: 32423408 DOI: 10.1186/s13017-020-00310-7] [FullText] [Full Text(PDF)]

23 **Gonullu E**, Bayhan Z, Capoglu R, Mantoglu B, Kamburoglu B, Harmantepe T, Altıntoprak F, Erkorkmaz U. Diagnostic Accuracy Rates of Appendicitis Scoring Systems for the Stratified Age Groups. *Emerg Med Int* 2022; **2022**: 2505977 [*RCA*] [PMID: 36353722 DOI: 10.1155/2022/2505977] [FullText]

24 **Pogorelić Z**, Mihanović J, Ninčević S, Lukšić B, Elezović Baloević S, Polašek O. Validity of Appendicitis Inflammatory Response Score in Distinguishing Perforated from Non-Perforated Appendicitis in Children. *Children (Basel)* 2021; **8**: 309 [*RCA*] [PMID: 33921577 DOI: 10.3390/children8040309] [FullText] [Full Text(PDF)]

25 **Karaman K**, Ercan M, Demir H, Yalkın Ö, Uzunoğlu Y, Gündoğdu K, Zengin İ, Aksoy YE, Bostancı EB. The Karaman score: A new diagnostic score for acute appendicitis. *Ulus Travma Acil Cerrahi Derg* 2018; **24**: 545-551 [*RCA*] [PMID: 30516254 DOI: 10.5505/tjtes.2018.62436] [FullText]

26 **Kaya MG**, Acar E. The Role of Inflammatory Parameters and Scoring Systems in Predicting Complicated Acute Appendicitis. *Meand Med Dent J* 2024; **25**: 305-316 [DOI: 10.69601/meandrosmdj.1528717] [FullText]

27 **Lam A**, Squires E, Tan S, Swen NJ, Barilla A, Kovoor J, Gupta A, Bacchi S, Khurana S. Artificial intelligence for predicting acute appendicitis: a systematic review. *ANZ J Surg* 2023; **93**: 2070-2078 [*RCA*] [PMID: 37458222 DOI: 10.1111/ans.18610] [FullText]

28 **Issaiy M**, Zarei D, Saghazadeh A. Artificial Intelligence and Acute Appendicitis: A Systematic Review of Diagnostic and Prognostic Models. *World J Emerg Surg* 2023; **18**: 59 [*RCA*] [PMID: 38114983 DOI: 10.1186/s13017-023-00527-2] [FullText] [Full Text(PDF)]

29 **Kim M**, Park T, Kang J, Kim MJ, Kwon MJ, Oh BY, Kim JW, Ha S, Yang WS, Cho BJ, Son I. Development and validation of automated three-dimensional convolutional neural network model for acute appendicitis diagnosis. *Sci Rep* 2025; **15**: 7711 [*RCA*] [PMID: 40044743 DOI: 10.1038/s41598-024-84348-6] [FullText] [Full Text(PDF)]

30 **Pati A**, Panigrahi A, Nayak DSK, Sahoo G, Singh D. Predicting Pediatric Appendicitis using Ensemble Learning Techniques. *Procedia Comput Sci* 2023; **218**: 1166-1175 [DOI: 10.1016/j.procs.2023.01.095] [FullText]

31 **Dogan K**, Selcuk T. A Novel Deep Learning Approach for the Automatic Diagnosis of Acute Appendicitis. *J Clin Med* 2024; **13**: 4949 [*RCA*] [PMID: 39201090 DOI: 10.3390/jcm13164949] [FullText]

32 **Echevarria S**, Rauf F, Hussain N, Zaka H, Farwa UE, Ahsan N, Broomfield A, Akbar A, Khawaja UA. Typical and Atypical Presentations of Appendicitis and Their Implications for Diagnosis and Treatment: A Literature Review. *Cureus* 2023; **15**: e37024 [*RCA*] [PMID: 37143626 DOI: 10.7759/cureus.37024] [FullText] [Full Text(PDF)]

33 **Schipper A**, Belgers P, O'Connor R, Jie KE, Dooijes R, Bosma JS, Kurstjens S, Kusters R, van Ginneken B, Rutten M. Machine-learning based prediction of appendicitis for patients presenting with acute abdominal pain at the emergency department. *World J Emerg Surg* 2024; **19**: 40 [*RCA*] [PMID: 39716296 DOI: 10.1186/s13017-024-00570-7] [FullText]

34 **Roupakias S**, Kambouri K, Al Nimer A, Bekiaridou K, Blevrakis E, Tsalikidis C, Sinopidis X. Balancing Between Negative Appendectomy and Complicated Appendicitis: A Persisting Reality Under the Rule of the Uncertainty Principle. *Cureus* 2025; **17**: e81516 [*RCA*] [PMID: 40308422 DOI: 10.7759/cureus.81516] [FullText]

35 **Erman A**, Ferreira J, Ashour WA, Guadagno E, St-Louis E, Emil S, Cheung J, Poenaru D. Machine-learning-assisted Preoperative Prediction of Pediatric Appendicitis Severity. *J Pediatr Surg* 2025; **60**: 162151 [*RCA*] [PMID: 39855986 DOI: 10.1016/j.jpedsurg.2024.162151] [Full Text]

36 **Roshanaei G**, Salimi R, Mahjub H, Faradmal J, Yamini A, Tarokhian A. Accurate diagnosis of acute appendicitis in the emergency

department: an artificial intelligence-based approach. *Intern Emerg Med* 2024; **19**: 2347-2357 [*RCA*] [PMID: 39167270 DOI: 10.1007/s11739-024-03738-w] [FullText]

37 **Yazici H**, Ugurlu O, Aygul Y, Ugur MA, Sen YK, Yildirim M. Predicting severity of acute appendicitis with machine learning methods: a simple and promising approach for clinicians. *BMC Emerg Med* 2024; **24**: 101 [*RCA*] [PMID: 38886641 DOI: 10.1186/s12873-024-01023-9] [FullText]

38 **Gollapalli M**, Rahman A, Kudos SA, Foula MS, Alkhalifa AM, Albisher HM, Al-Hariri MT, Mohammad N. Appendicitis Diagnosis: Ensemble Machine Learning and Explainable Artificial Intelligence-Based Comprehensive Approach. *Big Data Cogn Comput* 2024; **8**: 108 [*RCA*] [DOI: 10.3390/bdcc8090108] [FullText]

39 **Males I**, Boban Z, Kumric M, Vrdoljak J, Berkovic K, Pogorelic Z, Bozic J. Applying an explainable machine learning model might reduce the number of negative appendectomies in pediatric patients with a high probability of acute appendicitis. *Sci Rep* 2024; **14**: 12772 [*RCA*] [PMID: 38834671 DOI: 10.1038/s41598-024-63513-x] [FullText]

40 **Wei W**, Tongping S, Jiaming W. Construction of a clinical prediction model for complicated appendicitis based on machine learning techniques. *Sci Rep* 2024; **14**: 16473 [*RCA*] [PMID: 39013966 DOI: 10.1038/s41598-024-67453-4] [FullText]

41 **Abu-Ashour W**, Emil S, Poenaru D. Using Artificial Intelligence to Label Free-Text Operative and Ultrasound Reports for Grading Pediatric Appendicitis. *J Pediatr Surg* 2024; **59**: 783-790 [*RCA*] [PMID: 38383177 DOI: 10.1016/j.jpedsurg.2024.01.033] [FullText]

42 **Chadaga K**, Khanna V, Prabhu S, Sampathila N, Chadaga R, Umakanth S, Bhat D, Swathi KS, Kamath R. An interpretable and transparent machine learning framework for appendicitis detection in pediatric patients. *Sci Rep* 2024; **14**: 24454 [*RCA*] [PMID: 39424647 DOI: 10.1038/s41598-024-75896-y] [FullText]

43 **Akbulut S**, Yagin FH, Cicek IB, Koc C, Colak C, Yilmaz S. Prediction of Perforated and Nonperforated Acute Appendicitis Using Machine Learning-Based Explainable Artificial Intelligence. *Diagnostics (Basel)* 2023; **13**: 1173 [*RCA*] [PMID: 36980481 DOI: 10.3390/diagnostics13061173] [FullText] [Full Text(PDF)]

44 **Harmantepe AT**, Dikicier E, Gönüllü E, Ozdemir K, Kamburoğlu MB, Yigit M. A different way to diagnosis acute appendicitis: machine learning. *Pol Przegl Chir* 2023; **96**: 38-43 [*RCA*] [PMID: 38629278 DOI: 10.5604/01.3001.0053.5994] [FullText]

45 **Park SH**, Kim YJ, Kim KG, Chung JW, Kim HC, Choi IY, You MW, Lee GP, Hwang JH. Comparison between single and serial computed tomography images in classification of acute appendicitis, acute right-sided diverticulitis, and normal appendix using EfficientNet. *PLoS One* 2023; **18**: e0281498 [*RCA*] [PMID: 37224137 DOI: 10.1371/journal.pone.0281498] [FullText]

46 **Phan-Mai TA**, Thai TT, Mai TQ, Vu KA, Mai CC, Nguyen DA. Validity of Machine Learning in Detecting Complicated Appendicitis in a Resource-Limited Setting: Findings from Vietnam. *Biomed Res Int* 2023; **2023**: 5013812 [*RCA*] [PMID: 37090195 DOI: 10.1155/2023/5013812] [FullText]

47 **Mijwil MM**, Aggarwal K. A diagnostic testing for people with appendicitis using machine learning techniques. *Multimed Tools Appl* 2022; **81**: 7011-7023 [*RCA*] [PMID: 35095329 DOI: 10.1007/s11042-022-11939-8] [FullText]

48 **Shikha A**, Kasem A. The Development and Validation of Artificial Intelligence Pediatric Appendicitis Decision-Tree for Children 0 to 12 Years Old. *Eur J Pediatr Surg* 2023; **33**: 395-402 [*RCA*] [PMID: 36113502 DOI: 10.1055/a-1946-0157] [FullText]

49 **Su D**, Li Q, Zhang T, Veliz P, Chen Y, He K, Mahajan P, Zhang X. Prediction of acute appendicitis among patients with undifferentiated abdominal pain at emergency department. *BMC Med Res Methodol* 2022; **22**: 18 [*RCA*] [PMID: 35026994 DOI: 10.1186/s12874-021-01490-9] [FullText] [Full Text(PDF)]

50 **Akgül F**, Er A, Ulusoy E, Çağlar A, Çitlenbik H, Keskinoğlu P, Şişman AR, Karakuş OZ, Özer E, Duman M, Yılmaz D. Integration of Physical Examination, Old and New Biomarkers, and Ultrasonography by Using Neural Networks for Pediatric Appendicitis. *Pediatr Emerg Care* 2021; **37**: e1075-e1081 [*RCA*] [PMID: 31503129 DOI: 10.1097/PEC.0000000000001904] [FullText]

51 **Xia J**, Wang Z, Yang D, Li R, Liang G, Chen H, Heidari AA, Turabieh H, Mafarja M, Pan Z. Performance optimization of support vector machine with oppositional grasshopper optimization for acute appendicitis diagnosis. *Comput Biol Med* 2022; **143**: 105206 [*RCA*] [PMID: 35101730 DOI: 10.1016/j.compbiomed.2021.105206] [FullText]

52 **Marcinkevičs R**, Reis Wolfertstetter P, Klimiene U, Chin-Cheong K, Paschke A, Zerres J, Denzinger M, Niederberger D, Wellmann S, Ozkan E, Knorr C, Vogt JE. Interpretable and intervenable ultrasonography-based machine learning models for pediatric appendicitis. *Med Image Anal* 2024; **91**: 103042 [*RCA*] [PMID: 38000257 DOI: 10.1016/j.media.2023.103042] [FullText]

53 **Marcinkevics R**, Reis Wolfertstetter P, Wellmann S, Knorr C, Vogt JE. Using Machine Learning to Predict the Diagnosis, Management and Severity of Pediatric Appendicitis. *Front Pediatr* 2021; **9**: 662183 [*RCA*] [PMID: 33996697 DOI: 10.3389/fped.2021.662183] [FullText] [Full Text(PDF)]

54 **Ghareeb WM**, Emile SH, Elshobaky A. Artificial Intelligence Compared to Alvarado Scoring System Alone or Combined with Ultrasound Criteria in the Diagnosis of Acute Appendicitis. *J Gastrointest Surg* 2022; **26**: 655-658 [*RCA*] [PMID: 34519000 DOI: 10.1007/s11605-021-05147-2] [FullText]

55 **Hayashi K**, Ishimaru T, Lee J, Hirai S, Ooke T, Hosokawa T, Omata K, Sanmoto Y, Kakihara T, Kawashima H. Identification of Appendicitis Using Ultrasound with the Aid of Machine Learning. *J Laparoendosc Adv Surg Tech A* 2021; **31**: 1412-1419 [*RCA*] [PMID: 34748429 DOI: 10.1089/lap.2021.0318] [FullText]

56 **Reismann J**, Kiss N, Reismann M. The application of artificial intelligence methods to gene expression data for differentiation of uncomplicated and complicated appendicitis in children and adolescents - a proof of concept study. *BMC Pediatr* 2021; **21**: 268 [*RCA*] [PMID: 34103023 DOI: 10.1186/s12887-021-02735-8] [FullText] [Full Text(PDF)]

57 **Stiel C**, Elrod J, Klinke M, Herrmann J, Junge CM, Ghadban T, Reinshagen K, Boettcher M. The Modified Heidelberg and the AI Appendicitis Score Are Superior to Current Scores in Predicting Appendicitis in Children: A Two-Center Cohort Study. *Front Pediatr* 2020; **8**: 592892 [*RCA*] [PMID: 33313029 DOI: 10.3389/fped.2020.592892] [FullText] [Full Text(PDF)]

58 **Akmese OF**, Dogan G, Kor H, Erbay H, Demir E. The Use of Machine Learning Approaches for the Diagnosis of Acute Appendicitis. *Emerg Med Int* 2020; **2020**: 7306435 [*RCA*] [PMID: 32377437 DOI: 10.1155/2020/7306435] [FullText] [Full Text(PDF)]

59 **Aydin E**, Türkmen İU, Namli G, Öztürk Ç, Esen AB, Eray YN, Eroğlu E, Akova F. A novel and simple machine learning algorithm for preoperative diagnosis of acute appendicitis in children. *Pediatr Surg Int* 2020; **36**: 735-742 [*RCA*] [PMID: 32314055 DOI: 10.1007/s00383-020-04655-7] [FullText]

60 **Rajpurkar P**, Park A, Irvin J, Chute C, Bereket M, Mastrodicasa D, Langlotz CP, Lungren MP, Ng AY, Patel BN. AppendiXNet: Deep Learning for Diagnosis of Appendicitis from A Small Dataset of CT Exams Using Video Pretraining. *Sci Rep* 2020; **10**: 3958 [*RCA*] [PMID: 32127625 DOI: 10.1038/s41598-020-61055-6] [FullText] [Full Text(PDF)]

61 **Park JJ**, Kim KA, Nam Y, Choi MH, Choi SY, Rhie J. Convolutional-neural-network-based diagnosis of appendicitis *via* CT scans in patients with acute abdominal pain presenting in the emergency department. *Sci Rep* 2020; **10**: 9556 [*RCA*] [PMID: 32533053 DOI: 10.1038/s41598-020-66674-7] [FullText] [Full Text(PDF)]

62 **Ramirez-Garcialuna JL**, Vera-Bañuelos LR, Guevara-Torres L, Martínez-Jiménez MA, Ortiz-Dosal A, Gonzalez FJ, Kolosovas-Machuca ES. Infrared thermography of abdominal wall in acute appendicitis: Proof of concept study. *Infrared Phys Techn* 2020; **105**: 103165 [DOI: 10.1016/j.infrared.2019.103165] [FullText]

63 **Zhao Y**, Yang L, Sun C, Li Y, He Y, Zhang L, Shi T, Wang G, Men X, Sun W, He F, Qin J. Discovery of Urinary Proteomic Signature for Differential Diagnosis of Acute Appendicitis. *Biomed Res Int* 2020; **2020**: 3896263 [*RCA*] [PMID: 32337245 DOI: 10.1155/2020/3896263] [FullText] [Full Text(PDF)]

64 **Kang HJ**, Kang H, Kim B, Chae MS, Ha YR, Oh SB, Ahn JH. Evaluation of the diagnostic performance of a decision tree model in suspected acute appendicitis with equivocal preoperative computed tomography findings compared with Alvarado, Eskelinen, and adult appendicitis scores: A STARD compliant article. *Medicine (Baltimore)* 2019; **98**: e17368 [*RCA*] [PMID: 31577737 DOI: 10.1097/MD.0000000000017368] [FullText] [Full Text(PDF)]

65 **Reismann J**, Romualdi A, Kiss N, Minderjahn MI, Kallarackal J, Schad M, Reismann M. Diagnosis and classification of pediatric acute appendicitis by artificial intelligence methods: An investigator-independent approach. *PLoS One* 2019; **14**: e0222030 [*RCA*] [PMID: 31553729 DOI: 10.1371/journal.pone.0222030] [FullText] [Full Text(PDF)]

66 **Gudelis M**, Lacasta Garcia JD, Trujillano Cabello JJ. Diagnosis of pain in the right iliac fossa. A new diagnostic score based on Decision-Tree and Artificial Neural Network Methods. *Cir Esp (Engl Ed)* 2019; **97**: 329-335 [*RCA*] [PMID: 31005266 DOI: 10.1016/j.ciresp.2019.02.006] [FullText]

67 **Shahmoradi L**, Safdari R, Mirhosseini MM, Arji G, Jannat B, Abdar M. Predicting Risk of Acute Appendicitis: A Comparison of Artificial Neural Network and Logistic Regression Models. *Acta Med Iran* 2019; **56**: 784-795

68 **Afshari Safavi A**, Zand Karimi E, Rezaei M, Mohebi H, Mehrvarz S, Khorrami MR. Comparing the accuracy of neural network models and conventional tests in diagnosis of suspected acute appendicitis. *J Mazandaran Univ Med Sci* 2015; **25**: 58-65

69 **Jamshidnezhad A**, Azizi A, Zadeh SR, Shirali S, Shoushtari MH, Sabaghan Y, Ziagham V, Attarzadeh M. A Computer Based Model in Comparison with Sonography Imaging to Diagnosis of Acute Appendicitis in Iran. *J Acute Med* 2017; **7**: 10-18 [*RCA*] [PMID: 32995164 DOI: 10.6705/j.jacme.2017.0701.002] [FullText]

70 **Park SY**, Kim SM. Acute appendicitis diagnosis using artificial neural networks. *Technol Health Care* 2015; **23** Suppl 2: S559-S565 [*RCA*] [PMID: 26410524 DOI: 10.3233/THC-150994] [FullText]

71 **Yoldaş Ö**, Tez M, Karaca T. Artificial neural networks in the diagnosis of acute appendicitis. *Am J Emerg Med* 2012; **30**: 1245-1247 [*RCA*] [PMID: 21908136 DOI: 10.1016/j.ajem.2011.06.019] [FullText]

72 **Hsieh CH**, Lu RH, Lee NH, Chiu WT, Hsu MH, Li YC. Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. *Surgery* 2011; **149**: 87-93 [*RCA*] [PMID: 20466403 DOI: 10.1016/j.surg.2010.03.023] [FullText]

73 **Prabhudesai SG**, Gould S, Rekhraj S, Tekkis PP, Glazer G, Ziprin P. Artificial neural networks: useful aid in diagnosing acute appendicitis. *World J Surg* 2008; **32**: 305-9; discussion 310 [*RCA*] [PMID: 18043966 DOI: 10.1007/s00268-007-9298-6] [FullText]

74 **Grigull L**, Lechner WM. Supporting diagnostic decisions using hybrid and complementary data mining applications: a pilot study in the pediatric emergency department. *Pediatr Res* 2012; **71**: 725-731 [*RCA*] [PMID: 22441377 DOI: 10.1038/pr.2012.34] [FullText]

75 **Lee YH**, Hu PJ, Cheng TH, Huang TC, Chuang WY. A preclustering-based ensemble learning technique for acute appendicitis diagnoses. *Artif Intell Med* 2013; **58**: 115-124 [*RCA*] [PMID: 23623208 DOI: 10.1016/j.artmed.2013.03.007] [FullText]

76 **Son CS**, Jang BK, Seo ST, Kim MS, Kim YN. A hybrid decision support model to discover informative knowledge in diagnosing acute appendicitis. *BMC Med Inform Decis Mak* 2012; **12**: 17 [*RCA*] [PMID: 22410346 DOI: 10.1186/1472-6947-12-17] [FullText] [Full Text(PDF)]

77 **Ting HW**, Wu JT, Chan CL, Lin SL, Chen MH. Decision model for acute appendicitis treatment with decision tree technology--a modification of the Alvarado scoring system. *J Chin Med Assoc* 2010; **73**: 401-406 [*RCA*] [PMID: 20728850 DOI: 10.1016/S1726-4901(10)70087-3] [Full Text]

78 **Sakai S**, Kobayashi K, Toyabe S, Mandai N, Kanda T, Akazawa K. Comparison of the levels of accuracy of an artificial neural network model and a logistic regression model for the diagnosis of acute appendicitis. *J Med Syst* 2007; **31**: 357-364 [*RCA*] [PMID: 17918689 DOI: 10.1007/s10916-007-9077-9] [FullText]

79 **Aparicio PR**, Marcinkevics R, Wolfertstetter PR, Wellmann S, Knorr C, Vogt JE. Learning Medical Risk Scores for Pediatric Appendicitis. 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA); 2021 Dec 13-16; Pasadena, CA, United States. IEEE, 2021: 1507-1512

80 **Bianchi V**, Giambusso M, De Iacob A, Chiarello MM, Brisinda G. Artificial intelligence in the diagnosis and treatment of acute appendicitis: a narrative review. *Updates Surg* 2024; **76**: 783-792 [*RCA*] [PMID: 38472633 DOI: 10.1007/s13304-024-01801-x] [FullText] [Full Text(PDF)]

81 **Chekmeyan M**, Liu SH. Artificial intelligence for the diagnosis of pediatric appendicitis: A systematic review. *Am J Emerg Med* 2025; **92**: 18-31 [*RCA*] [PMID: 40048888 DOI: 10.1016/j.ajem.2025.02.023] [FullText]

82 **Rey R**, Gualtieri R, La Scala G, Posfay Barbe K. Artificial Intelligence in the Diagnosis and Management of Appendicitis in Pediatric Departments: A Systematic Review. *Eur J Pediatr Surg* 2024; **34**: 385-391 [*RCA*] [PMID: 38290564 DOI: 10.1055/a-2257-5122] [FullText]

83 **Hua R**, O'Brien MK, Carter M, Pitt JB, Kwon S, Ghomrawi HMK, Jayaraman A, Abdullah F. Improving Early Prediction of Abnormal Recovery after Appendectomy in Children using Real-world Data from Wearables. *Annu Int Conf IEEE Eng Med Biol Soc* 2024; **2024**: 1-4 [*RCA*] [PMID: 40039430 DOI: 10.1109/EMBC53108.2024.10782031] [FullText]

84 **Sibic O**, Somuncu E, Yilmaz S, Avsar E, Bozdag E, Ozcan A, Aydin MO, Ozkan C. Diagnosis of Acute Appendicitis with Machine Learning-Based Computer Tomography: Diagnostic Reliability and Role in Clinical Management. *J Laparoendosc Adv Surg Tech A* 2025; **35**: 313-317 [*RCA*] [PMID: 39967483 DOI: 10.1089/lap.2024.0374] [FullText]

85 **Singh D**, Nagaraj S, Mashouri P, Drysdale E, Fischer J, Goldenberg A, Brudno M. Assessment of Machine Learning-Based Medical Directives to Expedite Care in Pediatric Emergency Medicine. *JAMA Netw Open* 2022; **5**: e222599 [*RCA*] [PMID: 35294539 DOI: 10.1001/jamanetworkopen.2022.2599] [FullText]

86 **Kucukakcali Z**, Akbulut S. Role of immature granulocyte and blood biomarkers in predicting perforated acute appendicitis using machine learning model. *World J Clin Cases* 2025; **13**: 104379 [*RCA*] [PMID: 40771742 DOI: 10.12998/wjcc.v13.i22.104379] [FullText] [Full Text (PDF)]

87 **Kucukakcali Z**, Akbulut S, Colak C. Evaluating Ensemble-Based Machine Learning Models for Diagnosing Pediatric Acute Appendicitis:

Insights from a Retrospective Observational Study. *J Clin Med* 2025; **14**: 4264 [*RCA*] [PMID: 40566009 DOI: 10.3390/jcm14124264] [Full Text] [Full Text(PDF)]

88 **Kendall J**, Gaspar G, Berger D, Levman J. Machine Learning and Feature Selection in Pediatric Appendicitis. *Tomography* 2025; **11**: 90 [*RCA*] [PMID: 40863881 DOI: 10.3390/tomography11080090] [FullText] [Full Text(PDF)]

89 **Aydın E**, Sarnıç TE, Türkmen İU, Khanmammadova N, Ateş U, Öztan MO, Sekmenli T, Aras NF, Öztaş T, Yalçınkaya A, Özbek M, Gökçe D, Yalçın Cömert HS, Uzunlu O, Kandırıcı A, Ertürk N, Süzen A, Akova F, Paşaoğlu M, Eroğlu E, Göllü Bahadır G, Çakmak AM, Bilici S, Karabulut R, İmamoğlu M, Sarıhan H, Karakuş SC. Diagnostic Accuracy of a Machine Learning-Derived Appendicitis Score in Children: A Multicenter Validation Study. *Children (Basel)* 2025; **12**: 937 [PMID: 40723130 DOI: 10.3390/children12070937] [FullText]

90 **Zhao Y**, Wang X, Zhang Y, Liu T, Zuo S, Sun L, Zhang J, Wang K, Liu J. Combination of clinical information and radiomics models for the differentiation of acute simple appendicitis and non simple appendicitis on CT images. *Sci Rep* 2024; **14**: 1854 [*RCA*] [PMID: 38253872 DOI: 10.1038/s41598-024-52390-z] [FullText]

91 **Liang D**, Fan Y, Zeng Y, Zhou H, Zhou H, Li G, Liang Y, Zhong Z, Chen D, Chen A, Li G, Deng J, Huang B, Wei X. Development and Validation of a Deep Learning and Radiomics Combined Model for Differentiating Complicated From Uncomplicated Acute Appendicitis. *Acad Radiol* 2024; **31**: 1344-1354 [*RCA*] [PMID: 37775450 DOI: 10.1016/j.acra.2023.08.018] [FullText]

92 **Li P**, Zhang Z, Weng S, Nie H. Establishment of predictive models for acute complicated appendicitis during pregnancy-A retrospective case-control study. *Int J Gynaecol Obstet* 2023; **162**: 744-751 [*RCA*] [PMID: 36762585 DOI: 10.1002/ijgo.14719] [FullText]

93 **Lin HA**, Lin LT, Lin SF. Application of Artificial Neural Network Models to Differentiate Between Complicated and Uncomplicated Acute Appendicitis. *J Med Syst* 2023; **47**: 38 [*RCA*] [PMID: 36952043 DOI: 10.1007/s10916-023-01932-5] [FullText]

94 **Iliou T**, Anagnostopoulos C, Stephanakis IM, Anastassopoulos G. Combined Classification of Risk Factors for Appendicitis Prediction in Childhood. In: Iliadis L, Papadopoulos H, Jayne C, editors. Engineering Applications of Neural Networks. Berlin: Springer, 2013: 203-211

95 **Deleger L**, Brodzinski H, Zhai H, Li Q, Lingren T, Kirkendall ES, Alessandrini E, Solti I. Developing and evaluating an automated appendicitis risk stratification algorithm for pediatric patients in the emergency department. *J Am Med Inform Assoc* 2013; **20**: e212-e220 [*RCA*] [PMID: 24130231 DOI: 10.1136/amiajnl-2013-001962] [FullText] [Full Text(PDF)]

96 **Malley JD**, Kruppa J, Dasgupta A, Malley KG, Ziegler A. Probability machines: consistent probability estimation using nonparametric learning machines. *Methods Inf Med* 2012; **51**: 74-81 [*RCA*] [PMID: 21915433 DOI: 10.3414/ME00-01-0052] [FullText]

97 **Forsström JJ**, Irjala K, Selén G, Nyström M, Eklund P. Using data preprocessing and single layer perceptron to analyze laboratory data. *Scand J Clin Lab Invest Suppl* 1995; **222**: 75-81 [*RCA*] [PMID: 7569750 DOI: 10.3109/00365519509088453] [FullText]

98 **Pesonen E**, Eskelinen M, Juhola M. Comparison of different neural network algorithms in the diagnosis of acute appendicitis. *Int J Biomed Comput* 1996; **40**: 227-233 [*RCA*] [PMID: 8666475 DOI: 10.1016/0020-7101(95)01147-1] [FullText]

99 **Alowais SA**, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA, Al Yami MS, Al Harbi S, Albekairy AM. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ* 2023; **23**: 689 [*RCA*] [PMID: 37740191 DOI: 10.1186/s12909-023-04698-z] [FullText] [Full Text(PDF)]

100 **Utmal DM**. Machine Learning Its Applications, Challenges & Tools: A Review. *Int J Comput Sci Mob Comput* 2021; **10**: 32-38 [DOI: 10.47760/ijcsmc.2021.v10i03.004] [FullText]

101 **Yadalam PK**, Thirukkumaran PV, Natarajan PM, Ardila CM. Light gradient boost tree classifier predictions on appendicitis with periodontal disease from biochemical and clinical parameters. *Front Oral Health* 2024; **5**: 1462873 [*RCA*] [PMID: 39346113 DOI: 10.3389/froh.2024.1462873] [FullText] [Full Text(PDF)]

102 **Obaido G**, Mienye ID, Egbelowo OF, Emmanuel ID, Ogunleye A, Ogbuokiri B, Mienye P, Aruleba K. Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects. *Mach Learn Appl* 2024; **17**: 100576 [DOI: 10.1016/j.mlwa.2024.100576] [FullText]

103 **Sarker IH**. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput Sci* 2021; **2**: 160 [*RCA*] [PMID: 33778771 DOI: 10.1007/s42979-021-00592-x] [FullText] [Full Text(PDF)]

104 **Sarker IH**. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Comput Sci* 2021; **2**: 420 [*RCA*] [PMID: 34426802 DOI: 10.1007/s42979-021-00815-1] [FullText] [Full Text(PDF)]

105 **Li M**, Jiang Y, Zhang Y, Zhu H. Medical image analysis using deep learning algorithms. *Front Public Health* 2023; **11**: 1273253 [*RCA*] [PMID: 38026291 DOI: 10.3389/fpubh.2023.1273253] [FullText] [Full Text(PDF)]

106 **Jiang F**, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017; **2**: 230-243 [*RCA*] [PMID: 29507784 DOI: 10.1136/svn-2017-000101] [FullText] [Full Text(PDF)]

107 **Wells L**, Bednarz T. Explainable AI and Reinforcement Learning-A Systematic Review of Current Approaches and Trends. *Front Artif Intell* 2021; **4**: 550030 [*RCA*] [PMID: 34095817 DOI: 10.3389/frai.2021.550030] [FullText] [Full Text(PDF)]

108 **Dazeley R**, Vamplew P, Cruz F. Explainable Reinforcement Learning for Broad-XAI: A Conceptual Framework and Survey. Available from: arXiv:2108.09003 [DOI: 10.48550/arXiv.2108.09003] [FullText]

109 **Yurdem B**, Kuzlu M, Gullu MK, Catak FO, Tabassum M. Federated learning: Overview, strategies, applications, tools and future directions. *Heliyon* 2024; **10**: e38137 [*RCA*] [PMID: 39391509 DOI: 10.1016/j.heliyon.2024.e38137] [FullText]

110 **Kyrimi E**, Dube K, Fenton N, Fahmi A, Neves MR, Marsh W, McLachlan S. Bayesian networks in healthcare: What is preventing their adoption? *Artif Intell Med* 2021; **116**: 102079 [*RCA*] [PMID: 34020755 DOI: 10.1016/j.artmed.2021.102079] [FullText]

111 **Denecke K**, May R, Rivera-Romero O. Transformer Models in Healthcare: A Survey and Thematic Analysis of Potentials, Shortcomings and Risks. *J Med Syst* 2024; **48**: 23 [*RCA*] [PMID: 38367119 DOI: 10.1007/s10916-024-02043-5] [FullText] [Full Text(PDF)]

112 **Oss Boll H**, Amirahmadi A, Ghazani MM, Morais WO, Freitas EP, Soliman A, Etminani F, Byttner S, Recamonde-Mendoza M. Graph neural networks for clinical risk prediction based on electronic health records: A survey. *J Biomed Inform* 2024; **151**: 104616 [*RCA*] [PMID: 38423267 DOI: 10.1016/j.jbi.2024.104616] [FullText]

113 **Nagarajah T**, Poravi G. A Review on Automated Machine Learning (AutoML) Systems. 2019 IEEE 5th International Conference for Convergence in Technology (I2CT); 2019 Mar 29-31; Bombay, India. IEEE, 2019: 1-6

114 **Anandalwar SP**, Callahan MJ, Bachur RG, Feng C, Sidhwa F, Karki M, Taylor GA, Rangel SJ. Use of White Blood Cell Count and Polymorphonuclear Leukocyte Differential to Improve the Predictive Value of Ultrasound for Suspected Appendicitis in Children. *J Am Coll Surg* 2015; **220**: 1010-1017 [*RCA*] [PMID: 25708747 DOI: 10.1016/j.jamcollsurg.2015.01.039] [FullText]

115 **Hao TK**, Chung NT, Huy HQ, Linh NTM, Xuan NT. Combining Ultrasound with a Pediatric Appendicitis Score to Distinguish Complicated from Uncomplicated Appendicitis in a Pediatric Population. *Acta Inform Med* 2020; **28**: 114-118 [*RCA*] [PMID: 32742063 DOI:

10.5455/aim.2020.28.114-118] [FullText] [Full Text(PDF)]

116 **Sadeghi Z**, Alizadehsani R, Cifci MA, Kausar S, Rehman R, Mahanta P, Bora PK, Almasri A, Alkhawaldeh RS, Hussain S, Alatas B, Shoeibi A, Moosaei H, Hladík M, Nahavandi S, Pardalos PM. A review of Explainable Artificial Intelligence in healthcare. *Comput Electr Eng* 2024; **118**: 109370 [DOI: 10.1016/j.compeleceng.2024.109370] [FullText]

117 **Alzubaidi L**, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021; **8**: 53 [*RCA*] [PMID: 33816053 DOI: 10.1186/s40537-021-00444-8] [FullText] [Full Text(PDF)]

118 **Cappuccio M**, Bianco P, Rotondo M, Spiezia S, D'Ambrosio M, Menegon Tasselli F, Guerra G, Avella P. Current use of artificial intelligence in the diagnosis and management of acute appendicitis. *Minerva Surg* 2024; **79**: 326-338 [*RCA*] [PMID: 38477067 DOI: 10.23736/S2724-5691.23.10156-0] [FullText]

119 **Arora A**, Alderman JE, Palmer J, Ganapathi S, Laws E, McCradden MD, Oakden-Rayner L, Pfohl SR, Ghassemi M, McKay F, Treanor D, Rostamzadeh N, Mateen B, Gath J, Adebajo AO, Kuku S, Matin R, Heller K, Sapey E, Sebire NJ, Cole-Lewis H, Calvert M, Denniston A, Liu X. The value of standards for health datasets in artificial intelligence-based applications. *Nat Med* 2023; **29**: 2929-2938 [*RCA*] [PMID: 37884627 DOI: 10.1038/s41591-023-02608-w] [FullText] [Full Text(PDF)]

120 **Kelly CJ**, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; **17**: 195 [*RCA*] [PMID: 31665002 DOI: 10.1186/s12916-019-1426-2] [FullText] [Full Text(PDF)]

121 **Adeniran AA**, Onebunne AP, William P. Explainable AI (XAI) in healthcare: Enhancing trust and transparency in critical decision-making. *World J Adv Res Rev* 2024; **23**: 2447-2658 [DOI: 10.30574/wjarr.2024.23.3.2936] [FullText]

122 **Han W**, Li W, Zhang H. A comprehensive review on the fundamental principles, innovative designs, and multidisciplinary applications of micromixers. *Phys Fluids* 2024; **36**: 101306 [DOI: 10.1063/5.0238393] [FullText]

123 **Chen X**, Tang T, Zhai J, Liang A, Li X, Chen X. Bioinspired Leaf-Vein Micromixer for a Rapid and Efficient Synthesis of Monodisperse Ciprofloxacin Lipid Nanoparticles. *Langmuir* 2025; **41**: 19572-19581 [*RCA*] [PMID: 40679291 DOI: 10.1021/acs.langmuir.5c02493] [Full Text]

124 **Han W**, Li W, Zhang H. Insight into mixing performance of bionic fractal baffle micromixers based on Murray's Law. *Int Commun Heat Mass* 2024; **157**: 107843 [DOI: 10.1016/j.icheatmasstransfer.2024.107843] [FullText]

125 **Hamilton A**. Artificial Intelligence and Healthcare Simulation: The Shifting Landscape of Medical Education. *Cureus* 2024; **16**: e59747 [*RCA*] [PMID: 38840993 DOI: 10.7759/cureus.59747] [FullText] [Full Text(PDF)]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: office@baishideng.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com

© 2025 Baishideng Publishing Group Inc. All rights reserved.