

REVIEW ARTICLE

Medicine Science 2025;14(3):948-53

Deep learning architectures in the prediction of acute appendicitis and perforated appendicitis: A narrative review

Sami Akbulut¹, Cemil Colak²¹*İnönü University Faculty of Medicine, Department of Surgery and Liver Transplant Institute, Malatya, Türkiye*²*İnönü University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Türkiye*

Received 19 March 2025; Accepted 12 June 2025

Available online 25 August 2025 with doi: 10.5455/medscience.2025.03.085

Content of this journal is licensed under a Creative Commons Attribution-NonCommercial-NonDerivatives 4.0 International License.



Abstract

Acute appendicitis remains a significant cause of acute abdominal pain requiring emergency intervention. Timely and accurate differentiation between uncomplicated and complicated (perforated) acute appendicitis is critical for optimizing patient management and minimizing morbidity and mortality rates. Conventional diagnostic techniques, such as clinical assessment, biochemical analysis, and imaging procedures, sometimes encounter difficulties in differentiating uncomplicated from complicated appendicitis due to overlapping clinical characteristics and interobserver variability. Recent advancements in deep learning (DL) architectures have transformed medical diagnostics, offering new opportunities for more precise disease classification. Convolutional neural networks (CNNs) and other DL-based models have demonstrated significant potential in analyzing radiological images, improving diagnostic accuracy, and reducing false negatives. These models can extract subtle imaging features that may not be easily identifiable by human evaluation, thus enhancing early detection and guiding timely surgical intervention. This narrative review explores the role of DL in differentiating uncomplicated and complicated appendicitis, assessing current methodologies, their performance metrics, limitations, and clinical implications. The findings highlight the potential for DL to revolutionize appendicitis diagnostics, ultimately contributing to improved patient outcomes and streamlined clinical workflows.

Keywords: Acute appendicitis, perforated acute appendicitis, deep learning, classification

Introduction

Acute appendicitis represents the most prevalent surgical emergency in abdominal medicine, with appendectomy ranking as the most commonly performed urgent abdominal operation across the globe [1,2]. The condition typically arises due to luminal obstruction of the appendix, most frequently caused by lymphoid hyperplasia in pediatric patients and fecalith formation in adults [1]. Published epidemiological studies suggest a lifetime acute appendicitis risk varying between 5% and 20%, with men exhibiting a higher probability (8.6%) compared to women (6.7%) [1,3]. Interestingly, while the biological risk is greater in males, population-based studies have revealed that women face a relatively higher cumulative probability of undergoing appendectomy over their lifetime—estimated at 23% for women and 12% for men [1,3]. Despite advances in surgical techniques, antibiotic therapy, and perioperative care, appendicitis continues

to impose a significant burden on healthcare systems due to its sudden onset and the urgency of its treatment [4].

Delayed or missed treatment of acute appendicitis may result in progression to a perforated state, leading to serious complications such as generalized peritonitis, intra-abdominal abscess formation, intestinal obstruction and sepsis [5]. These outcomes are associated with markedly higher rates of morbidity and mortality. In fact, perforated cases carry a 3.5- to 10-fold increase in mortality compared to uncomplicated appendicitis [6]. One of the central clinical challenges lies in promptly and accurately distinguishing between uncomplicated and perforated appendicitis [7]. This distinction is crucial, as it informs key therapeutic decisions, including whether non-operative management with antibiotics may be appropriate or whether urgent surgical intervention is required. Thus, the primary clinical challenge lies in the accurate differentiation

CITATION

Akbulut S, Colak C. Deep learning architectures in the prediction of acute appendicitis and perforated appendicitis: A narrative review. Med Science. 2025;14(3):948-53.



Corresponding Author: Sami Akbulut, İnönü University Faculty of Medicine, Department of Surgery and Liver Transplant Institute, Malatya, Türkiye
Email: akbulutsami@gmail.com

between uncomplicated acute appendicitis and perforated acute appendicitis since this distinction dictates the type of appropriate therapeutic strategies, including a choice between conservative management and urgent surgical intervention [8].

The classical diagnosis of acute appendicitis relies on a combination of clinical symptoms (right lower quadrant pain, fever, nausea), physical examination signs (McBurney's point tenderness, rebound tenderness), laboratory parameters (leukocytosis, elevated C-reactive protein), diagnostic laparoscopy, and radiological imaging—particularly ultrasonography and computed tomography (CT), and in select cases such as pregnancy, magnetic resonance imaging (MRI) [9-13]. Physical examination and symptom assessment are subjective and vary between clinicians. Laboratory markers lack specificity and may overlap with other inflammatory or infectious conditions. Ultrasonography is highly operator-dependent and has reduced accuracy in obese patients or in cases with obscured anatomical structures [14]. While computed tomography is considered the gold standard due to its high diagnostic sensitivity and specificity, each of these conventional tools carries intrinsic limitations. Computed tomography, although accurate, exposes patients to ionizing radiation and may yield inconclusive or false-negative results, especially in early-stage or atypical cases [14,15].

Because of the overlapping clinical features, the lack of specificity in laboratory results, and the variability inherent in imaging interpretations, there is an increasing need for more objective, reproducible, and intelligent diagnostic aids [12]. In recent years, artificial intelligence (AI), particularly its subdomain machine learning (ML), and more specifically, deep learning (DL) models have shown promise in addressing these gaps [16-19]. DL algorithms—particularly convolutional neural networks (CNNs)—excel in extracting high-dimensional patterns from both imaging and non-imaging datasets [17,20,21]. In the context of appendicitis, DL models have analyzed CT images, integrate clinical and laboratory data, and predict disease severity with high accuracy [22]. In addition, DL approaches have proven instrumental in reducing diagnostic errors, shortening decision-making time, and minimizing unnecessary surgeries. For instance, The use of DL has accelerated improvements in model performance, even in settings with limited medical datasets, thereby broadening its applicability across various clinical contexts [15].

The application of DL in acute appendicitis has been inspired by its success in other diagnostic domains, including pulmonary nodule detection, diabetic retinopathy screening, and classification of gastrointestinal lesions [23]. In appendicitis, DL models have demonstrated the capacity to integrate multimodal inputs—such as radiological scans, clinical history, laboratory values—and predict not only the presence of disease but also its severity. Moreover, the integration of explainable AI (XAI) frameworks, such as SHAP (SHapley Additive exPlanations), allows clinicians to interpret the rationale behind algorithmic predictions, fostering clinical trust and accountability [22].

Given the increasing availability of annotated imaging datasets and the push toward precision diagnostics, the use of DL in appendicitis triage is poised for rapid evolution. However, challenges such as external validation, model generalizability, and ethical considerations around data privacy and algorithmic bias remain to be addressed.

This narrative review synthesizes the current landscape of DL-based approaches in distinguishing acute appendicitis from perforated appendicitis. It evaluates the methodological innovations, diagnostic performance, practical limitations, and future research directions of these emerging technologies, with the ultimate goal of informing their responsible integration into clinical practice.

Methods

A narrative review of the current literature was conducted to examine the use of DL and ML models in the prediction, classification, and diagnostic support of acute appendicitis and perforated appendicitis. The review aimed to evaluate the diagnostic accuracy of DL-based systems, identify integration strategies of clinical, laboratory, and radiological features, and highlight limitations, biases, and opportunities for future research in AI-assisted appendicitis diagnostics.

An extensive literature search was carried out using the PubMed, Scopus, and Web of Science databases to identify relevant articles published between January 2010 and February 2024. The Boolean query applied was: ('acute appendicitis' OR 'perforated appendicitis') AND ('deep learning' OR 'machine learning' OR 'artificial intelligence') AND ('prediction' OR 'classification' OR 'diagnosis'). Reference lists of key papers were also manually reviewed to identify additional eligible studies.

As this is a narrative review, we did not apply a systematic screening protocol or flow diagram. Instead, a conceptual and thematic synthesis approach was adopted. The article selection was performed by assessing titles, abstracts, and full texts for relevance to the review's objectives, and grouped according to their primary focus areas (CNN-based imaging models, multimodal integration, pediatric-specific applications, XAI, and ethical implications etc.).

Inclusion criteria were defined as peer-reviewed original articles that employed DL or ML techniques in the diagnostic evaluation of acute appendicitis or perforated acute appendicitis. Studies were included if they utilized medical imaging data (such as ultrasound, CT, or MRI) and/or clinical parameters (including laboratory results, patient symptoms, or demographic characteristics etc.). Additional eligible studies were those that assessed model performance metrics (accuracy, AUC, sensitivity, specificity etc.) or discussed aspects of algorithm transparency, bias mitigation strategies, or decision-support frameworks related to artificial intelligence in clinical contexts.

Exclusion criteria consisted of case reports, editorials, letters to the editor, narrative commentaries, or abstracts from conferences

without full peer-reviewed publications. We also excluded studies unrelated to acute appendicitis or those that did not apply any artificial intelligence methodology. Research relying solely on conventional diagnostic methods without the integration of ML or DL techniques was also excluded.

This review emphasizes how AI, particularly DL architectures, can transform the diagnostic landscape of appendicitis by improving accuracy, standardization, and timeliness. In addition, it discusses practical limitations such as data scarcity, algorithmic bias, and explainability concerns, aiming to provide a roadmap for responsible and effective clinical translation of these technologies.

Clinical and Research Consequences

Deep/Machine Learning Architectures in Diagnosis of Acute Appendicitis

1. Convolutional Neural Networks in Imaging Analysis

The main technique for DL in medical imaging is CNNs. In a research article, a CNN based model was developed to determine acute appendicitis with 96% rate of accuracy in cases with advanced radiological findings and 83.3% accuracy in radiologically ambiguous cases [8]. Thus, this approach reduces the occurrence of false negative to minimize the incidence of appendicular perforation. In another study, a U-Net was employed to perform fully automated appendix detection in CT scans, achieving a Dice Similarity Coefficient of 85.94%. An article claim that the U-Net model was able to efficiently segment the appendix regions, thus facilitating a rapid and accurate diagnosis of radiologists [17]. There is additional research that supports the use of DL architectures in the improvement of diagnostic accuracy and speed. For instance, a medical article described how CT scoring systems could be used to predict certain appendicular perforation, which complements the approaches of DL with a structure criteria for a diagnosis [24]. In addition, another study constructed predictive models using clinical and laboratory data for the purpose of discrimination between simple and perforated appendicitis to demonstrate the feasibility to merge multiple types of data sources with AI models. Together, these studies help to understand how DL should be optimized in clinical diagnostics [6].

2. Transfer Learning and Small Dataset Strategies

The major challenge in medical DL is the limited amount of data, and therefore the model generalizability and robustness. A 3D DL model, pretrained on natural videos and fine-tuned on a small dataset of 438 CT scans, was developed to address this issue. However, pretraining on large, unrelated datasets improved model performance, increasing the AUC values [15]. Moreover, other works have further studied how the transfer learning and data augmentation techniques affect the diagnostic accuracy in the cases of small datasets. An article showed that the combined model exhibited a superior net benefit in predicting complicated acute appendicitis compared to the other three models [25]. In the

same vein, a medical paper employed XAI frameworks combined with ML algorithms to efficiently handle imbalanced data and attained 92% accuracy in detecting whether the appendicitis was perforated or not [22]. It has been highlighted that the integration of multi-source data, clinical, radiological and lab findings is an important strategy to enhance model robustness. Incorporating peripheral blood biomarkers in logistic regression models along with imaging data increases the predictive accuracy for different types of appendicitis. The significance of DL here is to emphasize on the potential of DL to transform appendicitis diagnosis when able to combat such difficulties as limited or even heterogeneous datasets [26].

3. Explainable Artificial Intelligence

The roles of the XAI in discriminating acute appendicitis and perforated appendicitis are used to increase the diagnostic accuracy as well as the interpretability of ML models. XAI is recently shown to enhance prediction performance and to reveal the reasoning behind the decisions of these models, an especially crucial attribute when dealing with them in clinical scenarios. By utilizing the ensemble methods and stacking models, it is demonstrated that diagnostic metrics cannot be any better than 95%. SHAP and feature importance analysis have also identified some of the key clinical features that affect the model predictions, like white blood cell count and total bilirubin levels. XAI provides counterfactual explanations that enable clinicians to understand how changes in patient data affect outcomes, thereby reducing diagnostic errors. Implementing XAI in appendicitis diagnosis may potentially lower the rate negative appendectomies especially in pediatric populations, as there will be refinement in decision-making processes. XAI integration brings trust in ML systems enabling their use in clinical situations for prompt and accurate diagnosis. A medical article presents an ML with XAI hybrid approach for both acute appendicitis and perforated acute appendicitis. The model is used to originally separate perforated acute appendicitis from non-perforated cases using CatBoost algorithm with 92% accuracy. In this hybrid model, the diagnostic accuracy was also high, and the SHAP values gave interpretability of the model by showing how the variables made particular decisions. Although the developments in XAI for diagnosis of appendicitis are promising, there are still hurdles in validating these models externally and in generalizing them to varying clinical environments. However, these limitations need further research to make the robustness of AI applications to medical diagnostics [22,27-29].

4. Clinical Features and Biomarker Integration

Integrating clinical features and biomarkers into DL models significantly enhances diagnostic precision, particularly in predicting simple, purulent, or perforated appendicitis. To derive logistic regression models using peripheral blood biomarkers, such as CD4+/CD8+ ratios, and procalcitonin (PCT) levels along with clinical symptoms, a medical study was constructed. The model yielded a very good predictive capability to differentiate appendicitis types in preoperative stage [26].

Complementary research on CT based score systems reinforced the utility of combining the imaging biomarker and laboratory data in predicting appendicular perforation effectively [24]. Collectively, these findings highlight the promise of combining heterogeneous data sources within DL frameworks to improve diagnostic accuracy and inform more highly personalized and effective patient management strategies [24,26].

5. Pediatric Applications and Specialized Algorithms

In pediatric populations, distinguishing complicated appendicitis is particularly challenging due to atypical symptom presentation and varying responses to inflammation. A medical work addressed this challenge by developing a decision tree-based diagnostic algorithm that integrates CT imaging features and laboratory data to predict complicated appendicitis in children. Their model achieved high performance metrics in the training set [30]. Another paper introduced a method for the automatic diagnosis of pediatric acute appendicitis and the differentiation between complicated and uncomplicated inflammation, utilizing routinely and objectively obtained parameters for patients with suspected appendicitis. The proposed approach has the potential to transform current therapeutic approaches for appendicitis and illustrates the ability of AI and ML algorithms to enhance diagnostics significantly, even when based on standard diagnostic parameters [31]. A study examined the application of ML classifiers for predicting the diagnosis, therapy, and severity of appendicitis in pediatric patients. This work introduces the inaugural ML model for appendicitis and this has been realized as an accessible web tool. ML algorithms help to address the diagnostic and therapeutic issues in cases of appendicitis in pediatric patients and assist in making a more individual approach in the medical decision-making. A more detailed trial study is needed to finalize a clinical decision support system [32]. In supporting studies, combining clinical and radiological data for pediatric appendicitis diagnosis has been found to be valid. The advances described in these advancements provide the opportunity of ML to contribute towards the management of pediatric appendicitis by making clinicians extremely powerful tools with which to increase diagnostic accuracy and to guide timely surgical intervention.

6. Future Perspectives and Ethical Considerations

6.1. Emerging Trends in Artificial Intelligence for Appendicitis Diagnosis

Since DL is a developing stream, we can expect there are some trends that will affect appendicitis diagnosis in the near future. Federated learning also becomes a popular practice these days that allows training a model collaboratively across institutions without compromising the patient privacy. Emergency departments are being testbed for real-time AI assisted diagnostic tools to streamline the workflow and minimize the diagnostic time. Moreover, AI incorporation to wearable health devices will be able to monitor patients at risk of appendicitis continuously, making early detection of symptoms possible [33].

6.2. Potential for Personalized Medicine

AI and precision medicine have important promise. DL can formulate treatment plan that are patient specific based on Medicare, World Health Organization (WHO) diagnostic criteria, lifestyle factors, genomics and comprehensive medical history. Therefore, personalized models would be able to predict individual risks of acute appendicitis complications, like perforation, and therefore inform the decision of surgery timing and conservative treatment [34].

6.3. Challenges in Data Privacy and Security

Concerns have risen recently about the reliance on AI in health care, which is essentially a matter of privacy and security of the data. In the domain of DL, there are large volumes of data required for training, which makes it susceptible to the problem of a breach and misuses of data. It is essential to be adherent to data protection regulation like General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA). The risks described above are mitigated by techniques such as data anonymization, encryption, and federated learning [35].

6.4. Bias and Fairness in Artificial Intelligence Models

There is still the ethical problem of bias in AI models. If the models are trained on datasets with lack diversity, the models may produce undesirable output that will lead to unequal healthcare delivery. For instance, demographic factors like age, sex, ethnicity may impact on appendicitis presentation that if not taken into consideration may lead to diagnostic errors. Methods to avoid bias involve expanding the training datasets and the usage of algorithms that are developed in a fair sense [36].

6.5. Clinical Implementation and Decision-Making

To be widely adopted in the clinical practice, AI tools need to be explainable and transparent. Building the trust and safe usage of AI models requires that clinicians understand how the AI models reach their conclusions. SHAP (SHapley Additive exPlanations) type of XAI frameworks provide insights into feature importance and decision paths to explain the generative nature of such AI outputs and make them more interpretable to clinicians. It will also be important to integrate AI into existing healthcare workflows using user friendly interface in order to maximize the clinical utility of AI [37].

Discussion

The growing potential of DL as a discriminator of acute appendicitis versus perforated acute appendicitis is displayed in the reviewed studies. CNNs and U-Net architectures have proven to be very powerful in image based diagnosis tasks like appendices detection and segmentation using CT images. Since CNNs are particularly effective in extracting hierarchical features from radiological images, they are highly appropriate for identifying complex patterns that lie in radiology images [8,17]. In many medical imaging applications, transfer learning has significantly helped in improving model performance, as instances of these

problems often lack large annotated datasets. Specifically, by utilizing pretrained networks and fine-tuning them on specific medical datasets, such as the AppendiXNet model which yielded impressive performance gains by video pretraining, Learning's breadth has been increased by researchers and thereby utility and efficacy of diagnosis and care [15].

Explanation of the artificial intelligence (AI) can be further enhanced using XAI frameworks for the clinical application of DL models. SHAP (SHapley Additive exPlanations) type tools can be used for transparent interpretation of outputs generated by the model, allowing clinicians to see which features have the biggest impact on the actual prediction the model makes. In the healthcare setting, it is important that AI- driven diagnoses are transparent so clinician trust and enhanced patient outcomes can be achieved [22]. These advancements collectively suggest that DL- based AI for appendicitis diagnosis could be more accurate, efficient, and interpretable than traditional machine-based methods used in clinical practice. However, there are several challenges still. Model generalizability may be limited by data heterogeneity, lack of standard imaged protocols and small dataset sizes. In addition, DL models must be subjected to very rigorous validation in different clinical settings.

While constructing a diagnosis, future research may be undertaken that uses multimodal methods that combine imaging with clinical, laboratory, and even genetic data. The combination of CCTA with other biomarker tests – including CRP, WBC, and procalcitonin – has been reported to improve diagnostic accuracy substantially. Genomic data in conjunction with background information on the patient may also prove useful in further refining the diagnostic algorithms to increase the chances of accuracy [26]. Shared scientific data and federated learning provide an alternative means to tackle problems regarding insufficient data and data privacy. In the context of federated learning, different institutions while keeping patient data confidential [35], can construct generalizable models. This not only improves the likelihood of minimizing data leaks, but is also compliant with regulations such as GDPR and HIPAA [35].

Moreover, future studies could try synthetic data generation and advanced data augmentation strategies, which would seem like ways to mitigate the limitations that come with small datasets. Techniques such as that of GAN can perhaps offer realistic synthetic medical images thereby backing model training while also boosting generalizability [38]. An XAI model can shed light on the decision-making process and hence gain the confidence of the clinician for adoption of AI tools without losing interpretation and clinical relevance of diagnostic decisions [37,38].

Conclusion

This narrative review demonstrates that DL architectures provide a transformative way to discriminate between acute and perforated appendix. Our synthesis of evidence from multiple studies demonstrates that these technologies transmit an extremely large increase in diagnostic accuracy beyond conventional methods, and CNN based models achieve accuracy rates.

This review makes the following key conclusions:

- Traditional image analysis methods do not compare well to CNN and U-Net architectures in terms of detecting and segmenting appendices.
- With limited amounts of large annotated medical datasets available, transfer learning and data augmentation strategies address the issue of ability to have large datasets, and such strategies make large amounts of data unnecessary in order to perform robust models.
- Multimodal integration of clinical, laboratory, and imaging data achieves the best diagnostic performance, which strengthens the presumption that the future of appendicitis diagnosis is realization of data fusion at the systems level.
- To facilitate clinical adoption of AI, such frameworks are important as they increase healthcare provider trust by contributing the transparency into model decision.
- Algorithms for children are very promising to treat the specificities of diagnosing appendicitis in children.

However, these advancements do not solve all the problems for widespread clinical implementation. Next steps include standardizing the images protocol, having larger multicenter validation studies and also attention toward ethical concerns that we have to deal with vis-à-vis data privacy and also algorithmic bias. Future study should include generalizability across many different patient populations and health care settings.

Conflict of Interests

The authors declare that there is no conflict of interest in the study.

Financial Disclosure

The authors declare that they have received no financial support for the study.

Acknowledgment

The grammar and academic English of this manuscript were revised with the assistance of ChatGPT-based editing tools and QuillBot software.

References

1. Akbulut S, Koc C, Kocaaslan H, et al. Comparison of clinical and histopathological features of patients who underwent incidental or emergency appendectomy. *World J Gastrointest Surg.* 2019;11:19-26.
2. Akbulut S, Koç C, Şahin TT, et al. An investigation into the factors predicting acute appendicitis and perforated appendicitis. *Ulus Travma Acil Cerrahi Derg.* 2021;27:434-42.
3. Téoule P, Laffolie J, Rolle U, Reissfelder C. Acute appendicitis in childhood and adulthood. *Dtsch Arztebl Int.* 2020;117:764-74.
4. Guan L, Liu Z, Pan G, et al. The global, regional, and national burden of appendicitis in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. *BMC Gastroenterol.* 2023;23:44.
5. Gerall CD, DeFazio JR, Kahan AM, et al. Delayed presentation and sub-optimal outcomes of pediatric patients with acute appendicitis during the COVID-19 pandemic. *J Pediatr Surg.* 2021;56:905-10.
6. Bröker ME, van Lieshout EM, van der Elst M, et al. Discriminating between simple and perforated appendicitis. *J Surg Res.* 2012;176:79-83.

7. Atema JJ, van Rossem CC, Leeuwenburgh MM, et al. Scoring system to distinguish uncomplicated from complicated acute appendicitis. *Br J Surg*. 2015;102:979-90.
8. Dogan K, Selcuk T. A novel deep learning approach for the automatic diagnosis of acute appendicitis. *J Clin Med*. 2024;13:4949.
9. Andersson RE. Meta-analysis of the clinical and laboratory diagnosis of appendicitis. *Br J Surg*. 2004;91:28-37.
10. Di Saverio S, Podda M, De Simone B, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. *World J Emerg Surg*. 2020;15:27.
11. 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;24:523-51.
12. Issaiy M, Zarei D, Saghaadeh A. Artificial intelligence and acute appendicitis: a systematic review of diagnostic and prognostic models. *World J Emerg Surg*. 2023;18:59.
13. Burns M, Hague CJ, Vos P, et al. Utility of magnetic resonance imaging for the diagnosis of appendicitis during pregnancy: a canadian experience. *Can Assoc Radiol J*. 2017;68:392-400.
14. Corson-Knowles D, Russell FM. Clinical ultrasound is safe and highly specific for acute appendicitis in moderate to high pre-test probability patients. *West J Emerg Med*. 2018;19:460-4.
15. Rajpurkar P, Park A, Irvin J, et al. AppendiXNet: deep learning for diagnosis of appendicitis from a small dataset of CT exams using video pretraining. *Sci Rep*. 2020;10:3958.
16. 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;50:4040-9.
17. Baştuğ BT, Güneri G, Yıldırım MS, et al. Fully automated detection of the appendix using u-net deep learning architecture in CT scans. *J Clin Med*. 2024;13:5893.
18. Chekmeyan M, Liu SH. Artificial intelligence for the diagnosis of pediatric appendicitis: a systematic review. *Am J Emerg Med*. 2025;92:18-31.
19. Takaishi, T., Kawai, T., Kokubo, Y. et al. Deep learning for appendicitis: development of a three-dimensional localization model on CT. *Jpn J Radiol*. 2025 Jul 16. doi: 10.1007/s11604-025-01834-1. Epub ahead of print.
20. Kim M, Park T, Kang J, et al. Development and validation of automated three-dimensional convolutional neural network model for acute appendicitis diagnosis. *Sci Rep*. 2025;15:7711.
21. Park JJ, Kim KA, Nam Y, et al. 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.
22. Akbulut S, Yagin FH, Cicek IB, et al. Prediction of perforated and nonperforated acute appendicitis using machine learning-based explainable artificial intelligence. *Diagnostics (Basel)*. 2023;13:1173.
23. Egger J, Gsaxner C, Pepe A, et al. Medical deep learning-a systematic meta-review. *Comput Methods Programs Biomed*. 2022;221:106874.
24. Lai B, Chu C, Leung B, et al. Appendicitis computed tomography score: a useful tool for predicting perforation and surgical course of acute appendicitis. *Hong Kong J Radiol*. 2015;18:267-76.
25. Liang D, Fan Y, Zeng Y, et al. Development and validation of a deep learning and radiomics combined model for differentiating complicated from uncomplicated acute appendicitis. *Acad Radiol*. 2024;31:1344-54.
26. Kang CB, Li XW, Hou SY, et al. Preoperatively predicting the pathological types of acute appendicitis using machine learning based on peripheral blood biomarkers and clinical features: a retrospective study. *Ann Transl Med*. 2021;9:835.
27. Gollapalli M, Rahman A, Kudos SA, et al. Appendicitis diagnosis: ensemble machine learning and explainable artificial intelligence-based comprehensive approach. *Big Data Cogn Computing*. 2024;8:108.
28. Chadaga K, Khanna V, Prabhu S, et al. An interpretable and transparent machine learning framework for appendicitis detection in pediatric patients. *Sci Rep*. 2024;14:24454.
29. Males I, Boban Z, Kumric M, et al. 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.
30. Byun J, Park S, Hwang SM. Diagnostic algorithm based on machine learning to predict complicated appendicitis in children using CT, laboratory, and clinical features. *Diagnostics (Basel)*. 2023;13:923.
31. Reismann J, Romualdi A, Kiss N, et al. Diagnosis and classification of pediatric acute appendicitis by artificial intelligence methods: an investigator-independent approach. *PLoS One*. 2019;14:e0222030.
32. Marcinkevics R, Wolfertstetter PR, Wellmann S, et al. Using machine learning to predict the diagnosis, management and severity of pediatric appendicitis. *Front Pediatr*. 2021;9:662183.
33. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. 2018;29:102-27.
34. Topol E. Deep medicine: how artificial intelligence can make healthcare human again: Basic Books, Inc., 2019.
35. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell*. 2020;2:305-11.
36. Chen IY, Pierson E, Rose S, et al. Ethical machine learning in healthcare. *Annu Rev Biomed Data Sci*. 2021;4:123-44.
37. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Van Luxburg U, Guyon I, Bengio S, et al., eds. *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, 4-9 December, 2017; Long Beach, California, USA. Curran Associates Inc., 4768-77.
38. Frid-Adar M, Klang E, Amitai M, et al. Synthetic data augmentation using GAN for improved liver lesion classification. *Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*; 2018. IEEE: 289-93.