REVIEWS

Medical Science of Ukraine
Медична наука України

ISSN: 2664-472X
e-ISSN: 2664-4738
Vol. 22, No. 1, 2026

UDC 004.94:616.36
DOI: https://doi.org/10.32345/2664-4738.1.2026.22

Journal homepage: https://nmuofficial.com/zhurnal-medychna-nauka-ukrayiny/
Received: 08.10.2025 Revised: 17.02.2026 Accepted: 18.03.2026 Published: 30.03.2026

Mykola Bortnyi*

PhD in Medical Sciences, Associate Professor
Kharkiv National Medical University
61022, 4 Nauky Ave., Kharkiv, Ukraine
https://orcid.org/0000-0001-7811-3580

Computer vision as a tool for optimising the early diagnosis of abdominal organ diseases

Abstract. This study aimed to evaluate the effectiveness of computer vision methods in the diagnosis of abdominal pathologies. The methodology included a search of databases covering publications from 2019 to September 2025. It described the application of convolutional neural network architectures, segmentation methods, and multimodal approaches to medical image analysis. The reviewed data showed that DenseNet-201 and EfficientNetV2M achieved accuracies of 97.1-99.4% in classifying endoscopic lesions, outperforming AlexNet and ResNet-101. Convolutional neural network-based systems demonstrated diagnostic performance comparable to that of experts (diagnostic odds ratio 1.03), with sensitivity of 84.8-94.3% and an area under the curve of 0.95-0.98 in the detection of colorectal polyps. Multimodal approaches, including a hierarchical fully convolutional neural network for positron emission tomography-computed tomography and the ENDOANGEL-MM system for endoscopy, achieved a 21% increase in accuracy relative to unimodal models. Automated segmentation systems reduced analysis time by 92.8% while maintaining Dice similarity coefficients of 0.936-0.965 and reducing inter-expert variability from 7.78% to 4.10%. In critical diagnostics, models for detecting parenchymal organ injuries showed variability in performance across organs (receiver operating characteristic area under the curve – 0.841-0.945; sensitivity – 0.736-0.862). The combined model achieved the highest indicators – accuracy of 0.954 and specificity of 0.966, confirming its suitability for primary screening. Data augmentation and multitask learning methods increased accuracy by 15-17.6% for rare pathologies, while domain randomisation ensured an increase in the Dice similarity coefficient of up to 15.8% on independent data. The practical significance of the results lies in their potential application by healthcare institutions to implement automated diagnostic systems, optimise radiological workflows, and standardise clinical conclusions, thereby reducing time to treatment initiation and decreasing diagnostic errors

Keywords: artificial intelligence; intelligent algorithms; medical imaging; automated segmentation; multimodal approaches; diagnostic accuracy; abdominal pathology

INTRODUCTION

In modern medicine, early diagnosis of abdominal diseases is characterised by highly variable clinical manifestations, limited sensitivity of traditional imaging methods, and discrepancies in interpretation among specialists. The introduction of computer vision and deep learning methods creates opportunities to automate the diagnostic process, improve the accuracy of pathology detection, and standardise clinical conclusions. At the same time, there is a need for systematic analysis of the effectiveness of different neural network architectures, training strategies, and multimodal approaches in clinical settings. Epidemiological trends in oncological and urgent pathologies of abdominal organs, as well as the limitations of traditional diagnostic methods, underscore the need to evaluate the potential of artificial intelligence technologies for early detection. The epidemiological context of abdominal cancer determines the relevance of developing diagnostic tools. Yu.O. Protas & O.V. Makarenko [1] conducted an epidemiological analysis of trends in stomach cancer

Suggested Citation:
Bortnyi M. Computer vision as a tool for optimising the early diagnosis of abdominal organ diseases. Med Sci Ukr. 2026;22(1):194–206. DOI: 10.32345/2664-4738.1.2026.22

*Corresponding author

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

incidence in Ukraine from 2014 to 2022, finding a decrease from 8,350 cases in 2014 to 5,401 in 2022. The study notes that this process is slower in Ukraine due to limited screening, with patients aged 45 and older accounting for over 94% of cases. Problems of late diagnosis and unequal access to specialised medical care were identified. Conventional imaging methods vary in their effectiveness in diagnosing complications of abdominal pathologies. V.V. Skyba *et al.* [2] conducted a comparative assessment of X-ray, ultrasound, and computed tomography in the diagnosis of intra-abdominal complications in 191 patients, establishing the informative value of ultrasound in the range from 48.1% to 73.5% depending on the pathology. The study reports that radiography did not detect free air in 17 of 22 patients with perforated ulcers. The authors emphasised the need for a comprehensive diagnostic approach. The integration of deep learning into the segmentation and classification of liver tumours is considered a means of improving diagnostic accuracy. S. Stephe *et al.* [3] presented systems based on temporal graph networks for segmentation and visual transformer blocks controlled by recurrent blocks for the classification of malignant liver tumours. The classifier achieved an accuracy of 98.79%, a sensitivity of 95.21%, and a specificity of 98.33%. The study also compared these results with the accuracy of 2D computed tomography (CT) data (91.02%) and highlighted the potential of transformer architectures in this context. Artificial intelligence (AI) methods are used to diagnose various types of cancer. G. Bulusu *et al.* [4] performed a review of machine and deep learning applications for the diagnosis of cancer of various localisations based on medical images. The review reports that for brain cancer, some architectures achieved 99.98% accuracy on BraTS 2021. It also reports an area under the receiver operating characteristic (ROC) curve of 0.91 for breast cancer and 97% accuracy for kidney cancer.

Deep learning is used for object segmentation in ultrasound images. X. Xiao *et al.* [5] conducted a systematic overview of segmentation methods in ultrasound imaging. They identify the U-Net as the most widely used architecture. Studies have reported that the use of dual U-Net increased the Dice score by 1.8% and by 14.5% for images of increased complexity. The review also highlighted new approaches, including diffusion models and segment anything model (SAM). It works with minimal annotation. The application of computer vision for the differential diagnosis of appendicitis was described by E.G. de Queiroz *et al.* [6], who analysed the literature on the use of computer vision algorithms for the diagnosis of appendicitis and non-specific abdominal pain. The study notes that delays in diagnosis are associated with high morbidity and mortality. The authors described the potential of AI to support differential diagnosis and its possible impact on the speed of therapeutic intervention. V. Ghodke *et al.* [7] organised the technical and ethical aspects of using convolutional neural networks in radiology. It included modified DenseNet models. The study compared the effectiveness of modified convolutional neural networks (CNNs) with traditional diagnostic methods and highlighted issues related to decision-making transparency and the ethical framework for processing medical data. P. Hu *et al.* [8] analysed deep learning methods for organ segmentation in CT and magnetic resonance imaging. The study reports that controlled methods achieved Dice scores of 0.94 for the liver and 0.92 for the kidneys. The authors identified the lack of annotated data as a key problem and considered hybrid CNN-Transformer models as a methodological approach to improving accuracy.

Existing studies primarily focus on individual neural network architectures or specific pathologies, whereas comprehensive comparisons of CNNs, transformers, and multimodal approaches for abdominal disease diagnosis remain limited. The effect of increasing data volume, semi-supervised learning, and domain adaptation on model performance in clinical settings with heterogeneous data remains understudied. The long-term implications of integrating automated systems into radiology workflows and clinical outcomes are poorly described. Most studies are based on retrospective datasets from individual medical centres, which limits the ability to generalise the results to different clinical populations. The aim of this study was to evaluate the effectiveness of deep learning systems in detecting abdominal diseases and to outline their potential for clinical application. The objectives included systematising data on the effectiveness of deep learning architectures (CNNs, transformer models, and multimodal approaches) in the diagnosis of abdominal diseases; assessing the accuracy of automated computer vision systems in comparison with the diagnostic effectiveness of medical experts; and determining the prospects for clinical implementation to optimise radiological workflows and improve the standardisation of diagnostic conclusions.

MATERIALS AND METHODS

This review study conducted a systematic review, selection and summary of scientific publications on the application of deep learning and modern computer vision models for the diagnosis of abdominal organ pathologies. The methodological approach included analyses of CNN architectures, segmentation, classification, and detection models, as well as evaluations of CT, magnetic resonance imaging (MRI), positron emission tomography, endoscopic methods, and multimodal approaches. A systematic search was conducted in the following databases: PubMed Central, the Institute of Electrical and Electronics Engineers, Multidisciplinary Digital Publishing Institute, ResearchGate, Scientific Reports, SpringerLink, the Radiological Society of North America, and ScienceDirect. The search

covered publications from 2019 to September 2025 and focused on articles in English and Ukrainian. Search queries were formed using combinations of keywords and their synonyms: "deep learning", "CNN", "computer vision", "abdominal CT", "abdominal MRI", "lesion detection", "polyps classification", "hepatic lesion segmentation", "colon polyps CNN", "TotalSegmentator", "nnU-Net", "U-Net", "explainable AI", "medical image segmentation", "early diagnosis", and "abdominal pathology". Logical operators (AND, OR), document type filters (peer-reviewed articles, reviews, meta-analyses), publication year, and relevance to medical imaging were applied across platforms. The initial search produced 142 publications. After removing duplicates and selecting relevant titles and abstracts, 103 sources were chosen. Full-text analysis provided a final corpus of 54 publications. It includes studies comparing CNN architectures for classifying endoscopic, CT, and MRI images; evaluating the accuracy of models for segmenting organs and pathological structures; investigating multimodal systems (CT + MRI, CT + clinical data); modelling algorithms for automatic CT processing for acute conditions; and analysing the application of explainable AI in clinical medical imaging. The process of selecting and identifying publications is shown in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart in Figure 1.

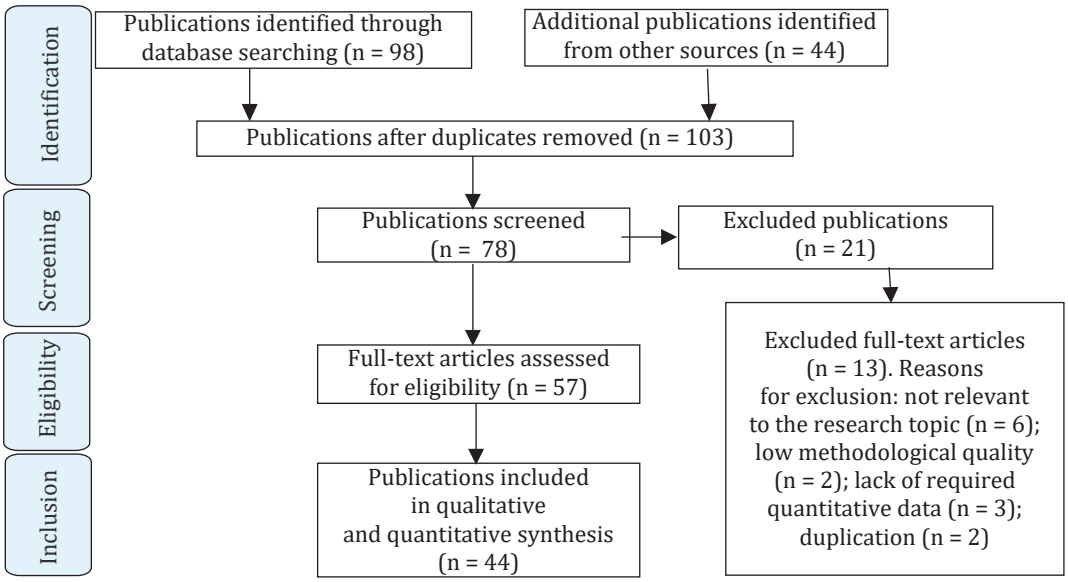

Figure 1. Flow diagram of identification, screening, and inclusion of publications in accordance with PRISMA
Source: compiled by the author based on M.J. Page *et al*. [9]

Required inclusion criteria: availability of performance metrics, including AUC, area under the sensitivity-specificity curve (AUROC), Dice similarity coefficient (DSC), intersection over union (IoU), sensitivity (Recall), positive predictive value (Precision) and overall accuracy (Accuracy); availability of test and validation sets; use of CNN architectures, including residual networks (ResNet, Residual Network), dense networks (DenseNet, Densely Connected Convolutional Network), efficient neural networks (EfficientNet), U-Net, next-generation adaptive U-Net (nnU-Net), second-generation segmentation model (SAM2), as well as deep segmentation networks (DeepLab) and Xception architecture (Extreme Inception) relevant to the early or acute diagnosis of abdominal diseases. The excluded sources included materials without full text, descriptive articles without experimental data, editorial notes and unreviewed preprints without specified indicators. To ensure the quality of the selection, a critical evaluation of empirical studies was conducted using the methodology outlined by E. Aromataris *et al*. [10] and adapted deep learning model evaluation criteria: sample adequacy, external validation, class balance, transparency of labelling methods, clarity of architecture description, and interpretability. The scores obtained were used to determine the reliability of the results and structure the synthesis. To summarise the results, an analytical approach that systematically compared architectures, segmentation methods, diagnostic indicators, and visualisation methods was used. Differences among CNN models were examined separately by task type (classification, segmentation, detection), pathology complexity level, CT/MRI modality and clinical application conditions.

RESULTS AND DISCUSSION

Effectiveness and integration of deep learning in early diagnosis. Early abdominal pathologies require automated diagnosis and the effectiveness of deep learning models depends on architecture, data type,

and task complexity. E.O. Cuevas-Rodriguez *et al.* [11] conducted an analytical comparison of the performance of deep convolutional neural networks for automated classification of early abdominal pathologies on the HyperKvasir dataset (6,792 endoscopic images, 10 types of lesions) and showed that DenseNet-201 outperformed AlexNet, Inception-v3, and ResNet-101, achieving 97.1% accuracy, 96.3% sensitivity, 99.7% specificity, and 94.9% area under the curve (AUC). S. Fanijo [12] conducted a comparative analysis of CNN, ResNet, DenseNet, and EfficientNet architectures in the classification of colorectal polyps and determined the high efficiency of DenseNet121 and EfficientNetV2M, which, after hyperparameter optimisation, achieved an accuracy of 99.4% and an F1 score. These results demonstrate the reliability of these architectures for detecting and classifying polyps in the dataset under study. F.I. Mir *et al.* [13] demonstrated that the use of a hybrid fully convolutional neural network (HFCNN) in combination with 3D positron emission tomography and computed tomography (PET-CT) for liver segmentation and tumour classification improves analytical accuracy. The use of multi-species learning and an ensemble strategy combining deep models with mathematical algorithms improved segmentation efficiency and tumour classification accuracy, demonstrating the promise of integrated solutions for early diagnosis and therapeutic decision-making in liver cancer. Thus, HFCNN demonstrates the suitability of a multimodal approach for improving the accuracy of liver pathology analysis. Y. Xu *et al.* [14] conducted a systematic review and meta-analysis comparing the diagnostic performance of CNN systems with that of human endoscopists in the detection and classification of colonic polyps. AI-based approaches demonstrated high performance: for polyp detection, sensitivity was 84.8% and AUC was 0.98; for classification, sensitivity was 94.3% and AUC was 0.95. The effectiveness of CNN in classification was comparable to that of experts (relative diagnostic odds ratio (RDOR) 1.03, p = 0.9654) and higher than that of non-experts (RDOR 0.18, p = 0.0342), confirming their suitability as clinical decision support tools. Thus, CNN systems demonstrate an accuracy sufficient to serve as an additional component alongside clinical experience in endoscopy.

Deep learning analytical models can be adapted to process multistructural and multimodal data. The introduction of integrated platforms that combine imaging, clinical data, and biomarkers can improve diagnostic accuracy and the effectiveness of early detection of abdominal organ pathologies. J.B. Harringa *et al*. [15] conducted a prospective study comparing the accuracy of CT, MRI, and ultrasound in detecting five early abdominal pathologies in women. The sensitivity of CT was 66.7% with a specificity of 98.3%, which exceeded the indicators of ultrasound (64.3% and 97.7%) and MRI (57.1% and 97.2%), although no statistically significant differences were found. N. Tavakolian *et al*. [16] confirmed the role of CT in a review of deep learning applications for the analysis of liver injuries and pathologies: early studies achieved 97% sensitivity and 95% specificity, while modern models provide a Dice Score of 0.973 for liver segmentation (DLA) and 82.0% accuracy with 84.7% specificity (convolutional block attention module neural network). Q.S. Alhammami [17] compared CT and MRI in diagnosing colorectal cancer metastases and showed that CT sensitivity depends on the number of lesions: overall 87.8%; with one metastasis 100%; with three 71.43%; with four or more 78.57%; whereas MRI demonstrated more stable results. Thus, CT demonstrates high diagnostic performance; however, diagnostic performance declines with increasing metastatic burden. In a review by J. Liu *et al*. [18] on the role of artificial intelligence in medical imaging, it is reported that a model trained on 5,754 chest CT images improved positioning accuracy, which affects the ability to detect minor abdominal organ pathologies. The demand for explainable artificial intelligence (XAI) is linked to requirements for algorithmic safety, transparency, and reproducibility. In image analysis, gradient-weighted class activation maps and similar tools enable visualisation of the regions that influence the model's decisions, thereby enhancing interpretability.

A comparative analysis of the effectiveness of a deep convolutional neural network for classifying early abdominal pathologies demonstrates that diagnostic performance varies with the type of input data and the model architecture. A study by E.O. Cuevas-Rodriguez *et al.* [11] on the HyperKvasir dataset, which comprises 10 pathologic classes, reported that DenseNet-201 achieved 97.1% accuracy, 96.3% sensitivity, 99.7% specificity, and 94.9% AUC for multi-class endoscopic classification. At the same time, the study by S. Fanijo [12] for the narrow task of colorectal polyp classification showed that DenseNet121 and EfficientNetV2M achieved 99% accuracy and 99.4% F1 score after hyperparameter optimisation, exceeding the results of multi-class classification by 1.9-2.3 percentage points. A comparative analysis of the diagnostic sensitivity of various imaging methods for detecting early pathologies shows variability in indicators. In a study by J.B. Harringa *et al.* [15] on pelvic pathologies, the sensitivity of CT was 66.7% with a specificity of 98.3%, which exceeds the indicators of ultrasound examination (64.3%) and MRI (57.1%). However, a study by Q.S. Alhammami [17] revealed the instability of CT in colorectal cancer metastases. Overall sensitivity was 87.8%, but 100% for a single metastasis and only 71.43% for three metastases. MRI demonstrated greater reliability for multifocal lesions. The comparison shows that although CT demonstrates high numerical indicators, its diagnostic reliability varies depending on the complexity of the case, while MRI provides more stable

accuracy for complex pathologies, which is relevant for the development of computer vision models. Overall, the analysis showed that deep convolutional networks achieve high diagnostic accuracy, but their effectiveness varies with the clinical task and available data. Architectures with stable results confirm their suitability for integration into screening and early detection of abdominal organ pathologies. The need to combine strong metrics with model explainability and reproducibility determines their potential as tools to support medical decision-making.

Accuracy and reliability of deep learning in critical abdominal diagnosis. Evaluating the performance of deep learning models for diagnosing abdominal organs enables the determination of their clinical potential. C.-T. Cheng *et al.* [19] conducted an analysis showing variability in effectiveness across organs and pathologies: the model for detecting kidney injuries on CT achieved the highest performance, whereas models for the spleen and liver had high accuracy and specificity but lower sensitivity. The overall performance of the combined model indicates its suitability for primary screening, although sensitivity requires improvement. Thus, the results confirm the potential utility of the algorithm for primary screening, despite its limited sensitivity. To illustrate model performance, Table 1 presents the performance of algorithms in detecting parenchymal organ injuries on CT.

Table 1. Performance of parenchymal organ injury detection models on the test sample

Model	Spleen injury model	p-value	Liver injury model	p-value	Kidney injury model	p-value	Combined model	p-value
AUROC (95% CI)	0.841 (0.781-0.895)	–	0.945 (0.888-0.988)	0.008	0.868 (0.811-0.918)	0.499	0.870 (0.721-0.994)	0.734
Accuracy (95% CI)	0.768 (0.706-0.825)	–	0.938 (0.902-0.969)	<0.05	0.820 (0.768-0.876)	0.259	0.954 (0.923-0.979)	<0.05
Sensitivity (95% CI)	0.736 (0.625-0.833)	–	0.862 (0.724-0.966)	0.270	0.727 (0.591-0.864)	1.000	0.833 (0.667-1.000)	0.581
Specificity (95% CI)	0.787 (0.713-0.861)	–	0.952 (0.915-0.982)	<0.05	0.847 (0.793-0.907)	0.264	0.966 (0.938-0.989)	<0.05
Positive Predictive Value (PPV) (95% CI)	0.671 (0.591-0.759)	–	0.758 (0.632-0.893)	0.494	0.583 (0.490-0.696)	0.384	0.714 (0.560-0.889)	0.908
Negative Predictive Value (NPV) (95% CI)	0.835 (0.780-0.891)	–	0.975 (0.951-0.994)	<0.05	0.915 (0.876-0.954)	0.086	0.983 (0.965-1.000)	<0.05

Note: CI – confidence interval. Table 1 is compiled from performance data of deep learning models for detecting parenchymal organ injuries (spleen, liver, kidneys) using computed tomography in patients with blunt abdominal trauma. The models were trained on 1,302 scans (87%) and tested on 194 scans (13%). Metrics were assessed using AUROC, accuracy, sensitivity, specificity, and positive and negative predictive values

Source: compiled by the author based on C.-T. Cheng *et al.* [19]

Table 1 shows that deep learning models can effectively identify parenchymal organ damage on CT scans. The combined model, which integrates information from all organs, achieved the highest accuracy (0.954) and specificity (0.966). Individual models also performed well: for example, the liver damage detection model achieved the highest AUROC (0.945), while the spleen model performed less well, with a sensitivity of 0.736 and an accuracy of 0.768. The high accuracy and specificity of the combined model indicate the significant potential of automated image analysis for rapid detection of pathologies. K.A. Attallah *et al.* [20] showed that the HCC-Check biomarker index demonstrated high diagnostic efficacy for the early detection of hepatocellular carcinoma in patients with different stages of liver fibrosis. When differentiating cirrhosis from carcinoma, the AUC was 0.95, sensitivity was 90.7%, and specificity was 83.9%; to distinguish all stages of fibrosis from carcinoma, the AUC reached 0.97. The index remained effective for tumours smaller than 5 cm (AUC 0.92, sensitivity 89.8%), confirming its suitability for early diagnosis and non-invasive screening of hepatocellular carcinoma. Therefore, the HCC-Check index is a promising tool for the non-invasive diagnosis of HCC. A retrospective study by M.T. Winkelman *et al.* [21] compared the diagnostic effectiveness of dual-energy CT with functional reconstructions. Iodine maps demonstrated an AUC of 1.0 and 100% sensitivity, exceeding the threshold of 0.95. Virtual images without contrast demonstrated an AUC of 0.93 and a sensitivity of 90%. Standard linear-blended images demonstrated an AUC of 0.64 and a sensitivity of 76.7%. In a systematic review of deep learning models for the diagnosis of abdominal pathologies, M. Bellal *et al.* [22] highlighted the balance between accuracy and completeness. The use of the Ensemble-GAN generative adversarial network for semantic organ segmentation achieved F1 scores of 0.93 for the spleen, 0.96 for the liver, 0.90 for the left kidney, and 0.94 for the right kidney, while Bayesian ensemble learning improved segmentation completeness for rare pathological structures. Thus, ensemble approaches improve segmentation reliability under data imbalance. A.A.A. El-Aziz *et al.* [23] studied the

effectiveness of EfficientNet-B3 for multiclass classification of histopathological bladder images and reported balanced performance: Recall 96.85%, Precision 97.95%, F1 97.36%, and accuracy 99.03%. Comparative analysis showed that performance depended on the optimiser: the RMSprop version achieved Recall 97.47% and Precision 96.99%, while the Adam version achieved Precision 97.95% and Recall 96.85%. This indicates differences among optimisers in balancing completeness and accuracy, which is important for model tuning in clinical applications.

G. Quero *et al.* [24] showed that computer vision methods in colorectal cancer surgery provide high performance in recognising surgical phases (Recall 89-89.4%, Precision 90-96%) and in assessing skill quality (sensitivity 94.1%, specificity 96.5%), confirming their ability to detect critical events and support surgeon training. P. Furtado [25] evaluated DeepLabV3 post-processing methods on MRI images using semantic invariants and showed improved segmentation accuracy. Incorporating anatomical constraints, class reassignment, and smoothing increased IoU to 0.88-0.90 (instead of 0.84-0.87) and Dice to 0.936-0.95 (instead of 0.91-0.93), reflecting a substantial effect of semantic constraints on results. Thus, semantic invariants are effective for improving automated segmentation accuracy. A study by E. Chukwujindu *et al*. [26] on the impact of object characteristics on SAM2 accuracy showed improved performance after fine-tuning on medical data: IoU increased from 0.690 to 0.827 and Dice from 0.7771 to 0.8916, with the largest improvements observed for the pancreas (IoU +65.2%, Dice +48.8%). N. Sharma *et al.* [27] conducted a systematic comparison of 25 encoder-decoder combinations for gastrointestinal organ segmentation on MRI and identified ResNet50+DeepLabV3+ as the most effective architecture (Dice 0.9082, IoU 0.8796). Thus, this architecture may be considered an optimal solution for high-accuracy segmentation in clinical practice. A comparative analysis of deep learning applications across different types of medical images reveals differences in diagnostic performance attributable to the specifics of the input data. The study by C.-T. Cheng *et al.* [19] on detecting parenchymal organ injuries on CT showed that the kidney trauma model achieved an AUC of 0.956 and specificity of 98.9%, but sensitivity was only 83.3%, below the critical 95% threshold for mass screening. This contrasts with A.A.A. El-Aziz *et al.* [23], where EfficientNet-B3 for histopathological bladder images achieved Recall 96.85%, Precision 97.95%, and overall accuracy 99.03%, which is 13.55-15.73 percentage points higher than CT-based sensitivity. The comparison highlights that histopathological images yield more predictable metrics than CT data, where anatomical variability of injuries complicates detection. A comparative analysis of automated segmentation accuracy reveals a gradient of complexity with respect to object characteristics. P. Furtado's [25] study for normal parenchymal organs on MRI achieved IoU 0.88-0.90 and Dice 0.936-0.95. C. Yi *et al.* [28] for malignant lung lesions reported Dice of 78.52%, which is 15.1-16.5 percentage points lower than for normal organs. E. Chukwujindu *et al.* [26] for complex structures reported Dice of 0.838 for the pancreas after fine-tuning, which is 5.3 percentage points higher than the gross tumour volume (GTV) for lung tumours, but 9.8-11.2 percentage points lower than for normal organs. The comparison shows a pronounced gradient: the difference between the highest (normal organs, 0.95) and the lowest (lung tumour GTV, 0.7852) is 16.5 percentage points, with malignant lesions showing the lowest accuracy across all anatomical zones and imaging modalities. The overall body of evidence indicates that deep learning models can provide high accuracy in detecting a wide range of abdominal pathologies; however, effectiveness varies depending on the clinical task. Architectures demonstrating the best results meet the requirements of early diagnosis. At the same time, sensitivity and robustness limitations indicate the need for further adaptation to heterogeneous clinical conditions.

Automation and standardisation of diagnosing acute abdominal conditions. Assessment of modern AI applications in diagnosing acute abdominal conditions shows their effectiveness in accelerating image interpretation and optimising clinical processes. In a review of AI applications for appendicitis, pneumoperitoneum, cholecystitis, and bowel obstruction, J. Yao *et al.* [29] highlighted the potential to speed up diagnosis. The main bottleneck is the interval between image acquisition and the first report, which accounts for 32% of the CT workflow, while the workflow itself represents 29% of the patient's average time in the emergency department. Therefore, implementing computer vision systems may have a cascade effect on the diagnostic process. B. Al Qassabi *et al.* [30] evaluated the impact of standardised organisational measures – procedure optimisation, triage, and automated systems – and showed improved radiology operational efficiency: the proportion of studies completed within defined timeframes increased from 88% to 95%, with a monthly productivity gain of 0.6% ($R^2 = 0.88$). This indicates the effectiveness of structured workflow changes with automation. E.R. Ranschaert *et al.* [31] reported that AI-based optimisation includes image analysis and automated prioritisation management, including reassignment of urgent studies and automatic generation of MRI protocols, reducing time to a preliminary report. D. Veiga-Canuto *et al.* [32] compared inter-individual variability of manual segmentation of neuroblastic tumours with automated segmentation based on the self-configuring medical segmentation

system nnU-Net and demonstrated the possibility of reducing time without loss of accuracy. The median Dice coefficient for manual segmentation was 0.969, and for automated segmentation 0.965. Analysis time decreased from 56 minutes to 4.08 minutes, corresponding to a 92.8% reduction.

L. Abel *et al.* [33] studied the reproducibility of segmenting 34 abdominal anatomical structures using TotalSegmentator and demonstrated stability of automated measurements in multiphase CT: for 29 of 34 structures, volume deviation was less than 5%; the overall absolute deviation was 6.50%, lower than nnU-Net (10.03%) and inter-reader variability. For the liver, deviation was 4.10% compared with 7.78% among radiologists ($p = 0.036$). Consensus comparison of manual and automated neuroblastoma segmentation with nnU-Net showed minimal Dice difference, indicating AI's ability to standardise tumour volumes [32]. The difference in Dice coefficients was 0.004, reflecting the capacity of AI to standardise tumour volumes without dependence on individual reader experience. Y.-J. Park *et al.* [34] developed a cascaded-structure model for detecting intra-abdominal haemorrhage on CT, achieving diagnostic standardisation and reduced human variability: IoU assessed spatial overlap, sensitivity was 93.22%, and specificity was 99.60% (increasing from 86.84% due to cascade logic). M. Saban *et al.* [35] conducted a pilot study of integrating the ESR iGuide decision-support system and showed acceptability among specialists: mean agreement with recommendations was 4.17 ± 0.65, and in 75% of cases all experts totally agreed with the proposed procedure. In general, cascaded algorithms and integrated decision-support systems can improve accuracy, standardise diagnosis, and reduce variability in clinical decision-making for intra-abdominal haemorrhage.

In a systematic review of machine learning for the diagnosis and treatment of pancreatic cancer, I. Večurkovská *et al.* [36] noted the presence of certified solutions ready for implementation. For example, qCT-Liver & Pancreas (Qure.ai) has European certification under Medical Device Regulation 2017/745 and automatically analyses contrast-enhanced CT scans with an AUC of 0.84-0.97. Other commercial systems, such as AI Metrics, deep learning (DL Precise), and Quantib Prostate, provide automated segmentation and lesion tracking, reduce reading time (e.g., from 103 to 81 seconds, -21%), and reduce inter-reader variability by 63% [36, 37]. Thus, certified and commercially available computer vision systems have the potential to improve speed, accuracy, and standardisation in the diagnosis of abdominal organ pathology in clinical practice. A comparative analysis of the impact of automation on operational efficiency reveals differences between organisational measures and technological solutions. J. Yao *et al.* [29] reported that the interval between acquiring a CT image and producing a report accounts for 32% of the entire CT cycle, and the cycle itself accounts for 29% of the patient's time in the emergency department. B. Al Qassabi *et al*. [30] found that structured organisational measures without AI diagnostics increased the proportion of on-time examinations from 88% to 95% over nine months (monthly increase 0.6%, $R^2 = 0.88$). This contrasts with D. Veiga-Canuto *et al.* [32], where nnU-Net automated segmentation reduced analysis time from 56 minutes to 4.08 minutes (92.8% reduction) while maintaining Dice 0.965 versus 0.969 for manual segmentation. The comparison shows that organisational measures provide a gradual increase of 7 percentage points over nine months, whereas technological automation achieves a reduction of over 90% in time while preserving accuracy. A comparative analysis of the reproducibility of automated segmentation methods demonstrates the advantages of standardisation. L. Abel *et al.* [33] showed that TotalSegmentator for 34 anatomical structures achieved an overall volume deviation of 6.50%, lower than nnU-Net (10.03%) and inter-reader variability (4.10% vs 7.78% for the liver, $p = 0.036$). D. Veiga-Canuto *et al.* [32] found a Dice difference of only 0.004 between automated (0.965) and manual (0.969) neuroblastoma segmentation. Y.-J. Park *et al.* [34] showed that a cascaded model increased haemorrhage detection specificity from 86.84% to 99.60% (increase of 12.76 percentage points). The comparison indicates that automated systems provide standardisation at the level of inter-reader agreement, and cascaded approaches further reduce false positives by 12.76 percentage points. Analysis indicates that automated computer vision systems can reduce the time required to produce diagnostic reports while ensuring accuracy and reproducibility. Automation of segmentation, prioritisation, and protocol generation demonstrates AI's capacity to reduce subjective human-factor influence. Certified and commercially available solutions indicate readiness for wider implementation.

Challenges of implementing AI in abdominal diagnostics. The implementation of computer vision systems in diagnosing abdominal organs is constrained by technological, legal, and ethical factors, including a shortage of high-quality data for model training. In a retrospective study on automated diagnosis of early gastric cancer, J.-B. Park *et al.* [38] found that expanding the training set to 37,080 images using AutoAugment and cut-and-paste increased model accuracy from 0.7500 to 0.8417 (+9.1%), and adding test-time augmentation (TTA) increased it to 0.9000 (+5.8%). Thus, data augmentation affected the model's sensitivity to the early stages of oncological processes. To illustrate the impact of data augmentation methods, Table 2 presents gastric cancer classification results depending on the methods applied.

Table 2. Gastric cancer classification results depending on data augmentation methods

Metric	Original	AutoAugment	Cut and Paste
Dataset A			
Precision	0.7658	0.7936	0.8320
Sensitivity	0.7655	0.7929	0.8312
FPR	0.2113	0.1780	0.1737
F1-score	0.7656	0.7933	0.8316
Accuracy	0.7649	0.7921	0.8317
Dataset B			
Precision	0.7747	0.8114	0.8715
Sensitivity	0.7500	0.7833	0.8417
FPR	0.3077	0.2821	0.2338
F1-score	0.7622	0.7971	0.8563
Accuracy	0.7500	0.7833	0.8417

Note: Original – refers to the baseline dataset without augmentation, AutoAugment – refers to automated transformations, Cut and Paste – refers to an image insertion method for expanding the dataset. Indicators include precision, sensitivity, FPR, F1-score, and overall classification Accuracy

Source: compiled by the author based on J.-B. Park *et al*. [38]

H. Du *et al.* [39] conducted a systematic analysis of early gastric cancer diagnostic models based on biomarkers and emphasised the need for metrics that are robust to imbalance; WeightedEnsemble and CatBoost achieved Precision-Recall AUCs of 0.92 and 0.91, respectively. In work on generating synthetic images using StyleGAN and SPADE GAN, Y. Skandarani *et al.* [40] reported high visual realism (Fréchet Inception Distance (FID) 1.09; human recognition accuracy 60%); however, a U-Net trained on synthetic data achieved a Dice Score of 87%, which is 2% lower than a model trained on real images. Insufficient model generalisability is further supported by S.S. Dhesi *et al.* [41], who conducted a retrospective validation of a multilayer perceptron (MLP) for predicting abdominal aortic aneurysm growth: the internal mean absolute error (MAE) was 1.35 mm/year, while the external MAE was 1.8 mm/year; Random Forest, XGBoost (Extreme Gradient Boosting), and support vector machine (SVM) achieved external accuracy of 1.9-1.97 mm/year. This indicates the sensitivity of some models to inter-centre variation. In addressing domain variability in abdominal organ segmentation, Y. Shi *et al.* [42] showed that domain randomisation in nnU-Net improves generalisability: DSC on independent data was 0.88, exceeding TotalSegmentator (0.79) and MRSegmentator (0.76) by 11.4% and 15.8%, respectively ($p < 0.001$). A. Kirpalani *et al.* [43] performed external validation of a spleen injury detection model combining a 2D multi-axis visual transformer (MaxVit), a 2.5D convolution-attention network (CoatNet), and long short-term memory (LSTM) on 1,216 scans from another region, obtaining AUC 0.931 for binary classification and 0.950 for severe injuries, with accuracy 0.849. This indicates that domain randomisation and multi-component architectures increase model generalisability. Legal and ethical considerations in the implementation of AI in abdominal imaging include regulatory certification and model transparency. In a review of US Food and Drug Administration (FDA)-approved products, P. Ajmera *et al.* [44] highlighted the role of certification in minimising legal risks and ensuring patient safety. At the same time, the opacity of large models that operate as "black boxes" creates risks of systemic bias and limits decision auditing. A publication by the European Data Protection Supervisor [45] indicates that implementing explainable AI ensures accountability, transparency, and the auditability of diagnostic conclusions. Therefore, regulatory certification and the adoption of explainable AI are key measures for ensuring safety, transparency, and ethical responsibility in the application of AI in abdominal imaging.

A comparative analysis of methods for addressing training data scarcity reveals differences among augmentation strategies, architectural adaptation, and synthetic data generation. J.-B. Park *et al.* [38] showed that expanding the endoscopic image dataset for early gastric cancer from 4,120 to 37,080 via cut-and-paste increased accuracy from 0.7500 to 0.8417 (+9.1%), and with TTA to 0.9000 (overall gain 15%). H. Zhu *et al.* [46], for extremely imbalanced classes (No.4sa metastasis frequency 2.4%), used multitask learning (LMML-net) to increase overall AUC from 0.802 to 0.821 (+1.9 percentage points), but for the rare station from 0.789 to 0.928 (+17.6%), exceeding the overall gain from augmentation. The comparison indicates that augmentation yields a 15% increase, multitask learning yields 17.6% for imbalanced classes, whereas synthetic generation shows a 2 percentage-point shortfall. A comparative analysis of generalisability under external validation reveals differences among strategies. S.S. Dhesi *et al.* [41] showed that an MLP for aortic aneurysm lost 33.3% accuracy (MAE increased from 1.35 to 1.8 mm/year), while Random Forest, XGBoost, and SVM lost 26-32%. Y. Shi *et al.* [42] found that nnU-Net with domain randomisation achieved DSC 0.88, outperforming TotalSegmentator (0.79) by 11.4% and MRSegmentator

(0.76) by 15.8% ($p < 0.001$). A. Kirpalani *et al.* [43] showed that a multi-component architecture (2D MaxVit+2.5D CoatNet+LSTM) maintained AUC 0.931 without specialised adaptation. The comparison indicates: traditional ML models lose 26-33% accuracy, domain randomisation increases DSC by 11.4-15.8 percentage points, and multi-component architectures maintain stability without adaptation. In summary, the cited studies show that implementing computer vision in abdominal organ diagnostics requires addressing technological, methodological, and regulatory challenges. Augmentation methods, multitask architectures, and domain adaptation increase model reliability even under limited or imbalanced data. Robustness and legal acceptability are achievable under certification and the adoption of explainability principles.

Evolution of computer vision systems in abdominal diagnostics. The development of computer vision systems for abdominal diagnostics focuses on creating multimodal platforms that integrate imaging, laboratory, and clinical data to improve early diagnostic accuracy and predict disease course. M.F. Keicher [47] substantiated the need for integrating multimodal data for holistic clinical decision support. Practical application has been demonstrated by the ENDOANGEL-MM system developed by H. Du *et al.* [48], which integrates endoscopic images, reported prospective video tests yielding accuracy of 93.55% versus 70.97% for the best unimodal model and 90.00% versus 76.17% for average-performing endoscopists; among non-experts, accuracy increased from 70.75% to 85.75%. K. Cui *et al.* [49] developed the multimodal ResNet34-Clinical model for early diagnosis of necrotising enterocolitis in 408 neonates, achieving an accuracy of 0.942 and an AUC of 0.92, outperforming unimodal and traditional ML models; external validation yielded an AUC of 0.83, comparable to that of physicians (0.82). Thus, multimodal data integration substantially increases diagnostic accuracy and decision consistency, outperforming unimodal models and average expert performance, and demonstrates potential to improve clinical diagnosis of complex abdominal pathologies. Predictive models forecast recurrence, disease course, or treatment response. C. Prieto-De-La-Lastra *et al.* [50] developed a model to predict recurrence after colorectal cancer resection in a cohort of 278 patients (33.1% recurrence), combining clinicopathological data and radiomic features, achieving a C-index of 0.678 and a mean dynamic AUC of 0.691; the combined risk variable was the strongest predictor (hazard ratio (HR) 11.74, $p = 0.017$). D. Padhi *et al.* [51], in a study of a deep learning-based survival model for gastric cancer in 1,350 cases, reported a C-index of 0.923 and AUROC 0.93, and in an external cohort (388 patients), 0.936 and 0.94, demonstrating stability of key metrics across centres. Thus, deep learning-based predictive models enable reliable estimation of recurrence risk and disease prognosis, showing high portability between cohorts and potential for personalised treatment planning. Dynamic treatment optimisation was demonstrated by H. Li *et al.* [52] in predicting the causative pathogen of pyogenic liver abscess. Of 201 cases, 66.7% were caused by Klebsiella pneumoniae. A Gradient Boosting Machine model using multimodal clinical data and seven key variables achieved an AUC of 0.848, enabling antibiotic therapy to begin before culture results. Implementation of such systems is limited by the absence of standardised validation rules. In a systematic review of AI randomised controlled trials, R. Shahzad *et al.* [53] found that of 43 items in the consensus reporting standards for clinical trials (CONSORT-AI), the median reported was 30 (~70%), and descriptions of the algorithm and training data were missing in 36 of 42 studies. V. Vadlapudi & D.S.V.G.K. Kaladhar [54] emphasise that high detection performance does not compensate for the absence of external validation and failure to meet reporting standards. Adoption of CONSORT-AI and SPIRIT-AI guidance for interventional trials is a critical condition for clinical translation of computer vision systems. Therefore, the effectiveness of AI models in dynamic prediction and treatment optimisation depends not only on high accuracy but also on adherence to validation standards and transparent reporting, which are critical for clinical deployment.

Comparative analysis of multimodal systems demonstrates differences between data-integration strategies. H. Du *et al.* [48] showed that ENDOANGEL-MM, combining white-light and low-magnification endoscopic images, achieved 93.55% accuracy, which is 22.58 percentage points higher than the unimodal model (70.97%, $p < 0.001$) and 17.38 percentage points higher than endoscopists (76.17%, $p = 0.002$). Among non-experts, accuracy increased from 70.75% to 85.75% (+21%). K. Cui *et al.* [49] showed that integrating ResNet34 with biomarkers (C-reactive protein, procalcitonin, interleukin-6) achieved accuracy 0.942 and AUC 0.92, outperforming ML algorithms by 12.26 percentage points and unimodal ResNet34 by 2.21 percentage points; external validation (AUC 0.83) was equivalent to physicians (0.82). The comparison indicates that the multimodality gain in endoscopy (22.58 percentage points) is almost twice the gain in paediatric diagnosis (12.26 percentage points versus ML; 2.21 percentage points versus unimodal DL), indicating a gradient of informational complementarity. Comparative analysis of predictive models demonstrates a trade-off between discrimination and interpretability. D. Padhi *et al.* [51] reported that a purely imaging-based model achieved a C-index of 0.923 and an AUROC of 0.93, 24.5 percentage points higher than the combined approach; however, external validation reduced accuracy from 0.911 to 0.855 (a decrease of 5.6 percentage points). H. Li *et al.* [52] achieved AUC 0.848 using multimodal clinical data. The comparison indicates that purely imaging

approaches achieve the highest discrimination (C-index 0.923-0.936), multimodal clinical models achieve intermediate performance (AUC 0.848), and radiomic-clinical combinations have the lowest absolute values (C-index 0.678) but the strongest risk predictors (HR 11.74). Computer vision systems for abdominal diagnostics are shifting from unimodal solutions to multimodal platforms that provide a more comprehensive representation of the patient. Multimodal and predictive models increase diagnostic accuracy and consistency and enable treatment optimisation by assessing risk and transferring across cohorts. At the same time, successful deployment critically depends on adherence to validation standards, transparent reporting, and regulatory requirements.

CONCLUSIONS

This review study systematised and analysed scientific evidence on the use of deep learning and computer vision technologies for early diagnosis of abdominal organ pathologies. The analysis showed that deep learning architectures demonstrate high diagnostic accuracy in detecting a wide range of abdominal pathologies. DenseNet-201 and EfficientNetV2M achieved accuracy up to 99.4% in colorectal polyp classification, and CNN systems achieved results comparable to experts in endoscopic diagnosis (RDOR 1.03). Multimodal approaches, including HFCNN for PET-CT liver analysis and ENDOANGEL-MM for endoscopy, outperformed unimodal solutions, increasing accuracy by 21% compared with non-experts. In critical diagnosis, parenchymal organ trauma detection models showed variability in performance by organ (AUROC 0.841-0.945), and the combined model achieved accuracy 0.954 and specificity 0.966. Automated segmentation using nnU-Net reduced analysis time by 92.8% (from 56 minutes to 4.08 minutes) while maintaining a Dice Score of 0.965 versus 0.969 for manual segmentation. TotalSegmentator provided standardised measurements with a volume deviation of 6.50%, lower than the inter-reader variability; for the liver, the deviation was 4.10% versus 7.78% for manual assessment ($p = 0.036$). Implementing prioritisation systems and automated protocol generation reduced the time to reporting, which accounts for 32% of the CT workflow cycle. At the same time, substantial deployment barriers were identified. The shortage of representative data is mitigated by augmentation and multitask learning, which increases accuracy by 15-17.6% for rare pathologies. The generalisability problem is addressed through domain randomisation, which increases DSC by up to 15.8% on independent data. Legal and ethical challenges require regulatory certification and the adoption of Explainable AI to ensure transparency and accountability. The findings confirm that effective integration of AI into abdominal diagnostics is possible provided that technological improvements, methodological rigour, and compliance with regulatory standards are combined. For practical application, the implementation of multimodal platforms, automated segmentation systems, and XAI tools is recommended. Critical needs include creating international data repositories, standardising validation protocols in line with CONSORT-AI/SPIRIT-AI, and developing certification mechanisms. Promising areas include the development of federated learning to preserve data confidentiality, the creation of predictive models for assessing the risk of recurrence, research into the effectiveness of multimodal integration of clinical and radiological features, and the evaluation of the impact of AI systems on clinical outcomes through prospective randomised studies.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Protas YO, Makarenko OV. Epidemiology of gastric cancer in Ukraine from 2014-2022: Rate, sex and age. Intermedical J. 2025;1:121–4. DOI: 10.32782/2786-7684/2025-1-21

[2] Skyba VV, Rybalchenko VF, Ivanko OV, Yasin AD. Modern imaging technologies in the diagnosis of primary intra-abdominal complications in patients. Ukr Med J. 2021;2(142):92–6. DOI: 10.32471/umj.1680-3051.142.205391

[3] Stephe S, Kumar SB, Thirumalraj A, Dzhyvak V. Transformer based attention guided network for segmentation and hybrid network for classification of liver tumor from CT scan images. East Ukr Med J. 2024;12(3):692–710. DOI: 10.21272/eumj.2024;12(3):692-710

[4] Bulusu G, Vidyasagar KECh, Mudigonda M, Saikia MJ. Cancer detection using artificial intelligence: A paradigm in early diagnosis. Arch Comput Methods Eng. 2025;32:2365–403. DOI: 10.1007/s11831-024-10209-0

[5] Xiao X, Zhang J, Shao Y, Liu J, Shi K, He C, et al. Deep learning-based medical ultrasound image and video segmentation methods: Overview, frontiers, and challenges. Sensors. 2025;25(8):2361. DOI: 10.3390/s25082361

[6] de Queiroz EG, Moura K, Resende ÍNS, Cavalcante JL, Lira TF, Sousa GA. Computer vision algorithms to differentiate appendicitis from non-specific abdominal pain in the emergency department: a narrative literature review. Int Health Sci Rev. 2025;1(4):164–7. DOI: 10.70164/ihsr.V1I4.71

[7] Ghodke V, Musrif P, Shaikh R, Gavhane S. Enhancing radiology diagnosis through convolutional neural networks for computer vision in healthcare. Nanotechnol Perceptions. 2024;20(11):581–92. DOI: 10.62441/nano-ntp.v20iS11.2009
[8] Hu P, Li J, Wang Y, Kong D. Deep learning in abdominal organ segmentation: A review. Int Oncol. 2025;1(4):277–89. DOI: 10.1016/j.intonc.2025.06.007
[9] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. DOI: 10.1136/bmj.N71
[10] Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z. 9.2.8 Assessment of methodological quality. In: JBI manual for evidence synthesis. North Adelaide: JBI; 2024. P. 378–81. DOI: 10.46658/JBIMES-24-01
[11] Cuevas-Rodriguez EO, Galvan-Tejada CE, Maeda-Gutiérrez V, Moreno-Chávez G, Galván-Tejada JI, Gamboa-Rosales H, et al. Comparative study of convolutional neural network architectures for gastrointestinal lesions classification. PeerJ. 2023;11:e14806. DOI: 10.7717/peerj.14806
[12] Fanijo S. AI4CRC: A deep learning approach towards preventing colorectal cancer. J Fut Artif Intell Tech. 2024;1(2):143–59. DOI: 10.62411/faith.2024-28
[13] Mir FI, Raj G, Lata K, Verma AS. Deep data analysis for liver cancer prediction using machine learning approaches. In: 2024 1st international conference on advances in computing, communication and networking. Greater Noida: IEEE; 2024. P. 864–70. DOI: 10.1109/icac2n63387.2024.10895272
[14] Xu Y, Ding W, Wang Y, Tan Y, Xi C, Ye N, et al. Comparison of diagnostic performance between convolutional neural networks and human endoscopists for diagnosis of colorectal polyp: A systematic review and meta-analysis. PLoS ONE. 2021;16(2):e0246892. DOI: 10.1371/journal.pone.0246892
[15] Harringa JB, Bracken RL, Markhardt BK, Ziemlewicz TJ, Lubner M, Chiu A, et al. Magnetic resonance imaging versus computed tomography and ultrasound for the diagnosis of female pelvic pathology. Emerg Radiol. 2021;28(4):789–96. DOI: 10.1007/s10140-021-01923-4
[16] Tavakolian N, Nazemi A, Suen CY. Review of the AI-based analysis of abdominal organs from routine CT scans. Appl Sci. 2025;15(5):2516. DOI: 10.3390/app15052516
[17] Alhammami QS. Assessment of the diagnostic accuracy of CT as compared to MRI in detecting metastases in patients with colorectal cancer. Cureus. 2024;16(8):e66125. DOI: 10.7759/cureus.66125
[18] Liu J, Xiao H, Fan J, Hu W, Yang Y, Dong P, et al. An overview of artificial intelligence in medical physics and radiation oncology. J Nat Cancer Center. 2023;3(3):211–21. DOI: 10.1016/j.jncc.2023.08.002
[19] Cheng CT, Lin HH, Hsu CP, Chen HW, Huang JF, Hsieh CH, et al. Deep learning for automated detection and localization of traumatic abdominal solid organ injuries on CT scans. J Imag Inform Med. 2024;37(3):1113–23. DOI: 10.1007/s10278-024-01038-5
[20] Attallah KA, Albannan MS, Farid K, Rizk SM, Fathy N. HCC-check: A novel diagnostic tool for early detection of hepatocellular carcinoma based on cytokeratin-1 and epithelial membrane antigen: A cross-sectional study. Technol Cancer Res Treat. 2024;23. DOI: 10.1177/15330338241234790
[21] Winkelmann MT, Hagen F, Artzner K, Bongers MN, Artzner C. Dual-energy CT for accurate discrimination of intraperitoneal hematoma and intestinal structures. Diagnostics. 2022;12(10):2542. DOI: 10.3390/diagnostics12102542
[22] Bellal M, El Fkihi S, Cengiz K, Ivkovic N. A comprehensive survey on deep learning in abdominal imaging: Datasets, techniques, and performance metrics. IEEE Access. 2025;13:79894–914. DOI: 10.1109/access.2025.3566627
[23] El-Aziz AAA, Mahmood MA, El-Ghany SA. EfficientNet-B3-Based automated deep learning framework for multiclass endoscopic bladder tissue classification. Diagnostics. 2025;15(19):2515. DOI: 10.3390/diagnostics15192515
[24] Quero G, Mascagni P, Kolbinger FR, Fiorillo C, De Sio D, Longo F, et al. Artificial intelligence in colorectal cancer surgery: Present and future perspectives. Cancers. 2022;14(15):3803. DOI: 10.3390/cancers14153803
[25] Furtado P. Improving deep segmentation of abdominal organs MRI by post-processing. BioMedInformatics. 2021;1(3):88–105. DOI: 10.3390/biomedinformatics1030007
[26] Chukwujindu E, Faiz K, De Sequeira A, Chidom S, Faiz H. Improving medical image segmentation with SAM2: Analyzing the impact of object characteristics and finetuning on multi-planar datasets. Eur J Radiol Artif Intell. 2025;3:100034. DOI: 10.1016/j.ejrai.2025.100034
[27] Sharma N, Gupta S, Elkamchouchi DH, Bharany S. Encoder-decoder variant analysis for semantic segmentation of gastrointestinal tract using UW-madison dataset. Bioengineering. 2025;12(3):309. DOI: 10.3390/bioengineering12030309
[28] Yi C, Jiang S, Xiong L, Yang J, Shi H, Xiong Q, et al. D-S-Net: An efficient dual-stage strategy for high-precision segmentation of gross tumor volumes in lung cancer CT images. BMC Cancer. 2025;25(1):1387. DOI: 10.1186/s12885-025-14615-w

[29] Yao J, Chu LC, Patlas M. Applications of artificial intelligence in acute abdominal imaging. Can Assoc Radiol J. 2024;75(4):761–70. DOI: 10.1177/08465371241250197
[30] Al Qassabi B, AlSukaiti R, Alajmi S, Omar AS, Ibrahim R, Banibakr AA, et al. Improving turnaround times and operational efficiency in radiology services: Quality improvement study in Oman. Asian Pac J Cancer Prev. 2025;26(5):1709–18. DOI: 10.31557/apjcp.2025.26.5.1709
[31] Ranschaert ER, Topff L, Pianykh O. Optimization of radiology workflow with artificial intelligence. Radiol Clin North Am. 2021;59(6):955–66. DOI: 10.1016/j.rcl.2021.06.006
[32] Veiga-Canuto D, Cerdà-Alberich L, Sangüesa Nebot C, Martínez de las Heras B, Pötschger U, Gabelloni M, et al. Comparative multicentric evaluation of inter-observer variability in manual and automatic segmentation of neuroblastic tumors in magnetic resonance images. Cancers. 2022;14(15):3648. DOI: 10.3390/cancers14153648
[33] Abel L, Wasserthal J, Meyer MT, Vosshenrich J, Yang S, Donners R, et al. Intra-individual reproducibility of automated abdominal organ segmentation – performance of totalsegmentator compared to human readers and an independent nnU-Net model. J Imag Inform Med. 2025;38(3):1617–27. DOI: 10.1007/s10278-024-01265-w
[34] Park YJ, Cho HS, Kim MN. AI model for detection of abdominal hemorrhage lesions in abdominal CT images. Bioengineering. 2023;10(4):502. DOI: 10.3390/bioengineering10040502
[35] Saban M, Sosna J, Singer C, Vaknin S, Myers V, Shaham D, et al. Clinical decision support system recommendations: how often do radiologists and clinicians accept them? Eur Radiol. 2022;32(2):4218–24. DOI: 10.1007/s00330-021-08479-4
[36] Večurkovská I, Roškovičová V, Kaťuchová J. Challenges in the clinical application of machine learning for pancreatic cancer. Bratisl Med J. 2025;126:2437–50. DOI: 10.1007/s44411-025-00312-4
[37] Mervak BM, Fried JG, Neshewat J, Wasnik AP. Practical applications of artificial intelligence in body imaging. Abdom Radiol. 2026;51:458–66. DOI: 10.1007/s00261-025-05088-3
[38] Park JB, Lee HS, Cho HC. Investigating effective data augmentation techniques for accurate gastric classification in the development of a deep learning-based computer-aided diagnosis system. Appl Sci. 2023;13(22):12325. DOI: 10.3390/app132212325
[39] Du H, Yang Q, Ge A, Zhao C, Ma Y, Wang S. Explainable machine learning models for early gastric cancer diagnosis. Sci Rep. 2024;14:17457. DOI: 10.1038/s41598-024-67892-Z
[40] Skandarani Y, Jodoin PM, Lalande A. GANs for medical image synthesis: An empirical study. J Imaging. 2023;9(3):69. DOI: 10.3390/jimaginG9030069
[41] Dhesi SS, Adusumilli P, Ravikumar N, Waduud MA, Frood R, Frangi AF, et al. Development and external validation of [18F] FDG PET-CT-derived radiomic models for prediction of abdominal aortic aneurysm growth rate. Algorithms. 2025;18(2):86. DOI: 10.3390/a18020086
[42] Shi Y, Wang L, Qureshi TA, Deng Z, Xie Y, Li D. Deep learning with domain randomization in image and feature spaces for abdominal multiorgan segmentation on CT and MRI scans. Radiol Artif Intell. 2025;7(4):e240586. DOI: 10.1148/ryai.240586
[43] Kirpalani A, Viel T, Hu Z, Lin HM, Hermans S, Gomez D, et al. External validation of an RSNA 2023 abdominal trauma ai challenge high performing machine learning model in the detection and grading of splenic injuries on CT. Abdom Radiol. 2025;50(11):5581–90. DOI: 10.1007/s00261-025-04910-2
[44] Ajmera P, Dillard R, Kline T, Missert A, Korfiatis P, Khandelwal A. FDA-approved artificial intelligence products in abdominal imaging: A comprehensive review. Curr Probl Diagn Radiol. 2025. DOI: 10.1067/j.cpradiol.2025.04.011
[45] Attoresi M, Bernardo V, Lareo X, Velasco L. TechDispatch – explainable artificial intelligence. #2/2023. Luxembourg: Publications Office of the European Union; 2023. 23 P. DOI: 10.2804/802043
[46] Zhu H, Yang Z, Zheng C, Jiang P, Fang Y, Xu Y, et al. A 3D end-to-end multi-task learning network for predicting lymph node metastasis at multiple nodal stations in gastric cancer. Biomed Signal Process Control. 2025;108:107802. DOI: 10.1016/j.bspc.2025.107802
[47] Keicher MF. Multimodal deep learning for holistic clinical decision and reasoning support [Doctoral dissertation]. Munich: Technical University of Munich; 2024.
[48] Du H, Dong Z, Wu L, Li Y, Liu J, Luo C, et al. A deep-learning-based system using multi-modal data for diagnosing gastric neoplasms in real-time (with Video). Gastric Cancer. 2023;26:275–85. DOI: 10.1007/s10120-022-01358-X
[49] Cui K, Changrong S, Maomin Y, Hui Z, Xiuxiang L. Development of an artificial intelligence-based multimodal model for assisting in the diagnosis of necrotizing enterocolitis in newborns: A retrospective study. Front Pediatr. 2024;12:1388320. DOI: 10.3389/fped.2024.1388320
[50] Prieto-de-la-Lastra C, Carbonell-Asins JA, Bueno A, Gómez-Alderete A, Busto M, Alcolado-Jaramillo AB, et al. A radiomics-based artificial intelligence model to assess the risk of relapse in localized colon cancer. ESMO Open. 2025;10(8):105495. DOI: 10.1016/j.esmoop.2025.105495

[51] Padhi D, Majumdar SKD, Muralidharan A. Survival prediction in stomach cancer with deep learning: Unveiling model decisions with LIME and SHAP. Asian Pac J Cancer Prev. 2025;26(7):2669–77. DOI: 10.31557/apjcp.2025.26.7.2669
[52] Li H, Yu Y, Chen X, Sun Q, Li X, Shang Q, et al. Establishment of a machine learning-based predictive model for *Klebsiella pneumoniae* liver abscess. Infect Drug Resist. 2025;18:5097–108. DOI: 10.2147/idr.s545440
[53] Shahzad R, Ayub B, Siddiqui MAR. Quality of reporting of randomised controlled trials of artificial intelligence in healthcare: A systematic review. BMJ Open. 2022;12(9):e061519. DOI: 10.1136/bmjopen-2022-061519
[54] Vadlapudi V, Kaladhar DSVGK. Artificial intelligence in medical imaging: Bridging innovation, ethics, and clinical impact. Int J Adv Med. 2025;12(6):621–6. DOI: 10.18203/2349-3933.ijam20253360

Микола Бортний
Кандидат медичних наук, доцент
Харківський національний медичний університет
61022, просп. Науки, 4, м. Харків, Україна
https://orcid.org/0000-0001-7811-3580

Комп’ютерне бачення як інструмент оптимізації ранньої діагностики захворювань органів черевної порожнини

Анотація. Метою дослідження було оцінити результативність методів комп’ютерного бачення в діагностиці абдомінальних патологій. Методологія включала пошук матеріалів у базах даних з охопленням публікацій з 2019 по вересень 2025 року, що описували застосування архітектур згорткових нейронних мереж, методів сегментації та мультимодальних підходів до аналізу медичних зображень. Розглянуті дані показали, що архітектури DenseNet-201 та EfficientNetV2M досягали точності 97,1-99,4 % у класифікації ендоскопічних уражень, перевершивши показники AlexNet та ResNet-101. Системи на основі згорткових нейронних мереж продемонстрували діагностичні характеристики, співмірні з експертами (відношення діагностичних шансів 1,03), із чутливістю 84,8-94,3 % та площею під кривою 0,95-0,98 у виявленні колоректальних поліпів. Мультимодальні підходи, зокрема ієрархічна повнозгорткова нейронна мережа для позитронно-емісійної томографії-комп’ютерної томографії та система ENDOANGEL-MM для ендоскопії, забезпечили підвищення точності на 21 % порівняно з одномодальними моделями. Автоматизовані системи сегментації скоротили час аналізу на 92,8 % при збереженні коефіцієнта подібності Дайса 0,936-0,965 та зменшили міжекспертну варіабельність з 7,78 % до 4,10 %. У сфері критичної діагностики моделі виявлення травм паренхіматозних органів показали варіативність ефективності залежно від органу (характеристична крива «чутливість-специфічність» 0,841-0,945, чутливість 0,736-0,862). Комбінована модель досягла найвищих показників – точність 0,954 та специфічність 0,966, підтверджуючи придатність для первинного скринінгу. Методи аугментації даних і багатозадачного навчання підвищили точність на 15-17,6 % для рідкісних патологій, а доменна рандомізація забезпечила приріст коефіцієнта подібності Дайса до 15,8 % на незалежних даних. Практична значимість результатів полягає у можливості застосування їх медичними закладами для впровадження систем автоматизованої діагностики, оптимізації робочих процесів радіології та підвищення стандартизації клінічних висновків, що сприятиме скороченню часу до початку лікування та зменшенню діагностичних помилок

Ключові слова: штучний інтелект; інтелектуальні алгоритми; медична візуалізація; автоматизована сегментація; мультимодальні підходи; діагностична точність; абдомінальна патологія